Package 'scRNAseq'

June 18, 2019

Title Collection of Public Single-Cell RNA-Seq Datasets

Version 1.11.3

Date 2019-06-16

Description Gene-level counts for a collection of public scRNA-seq datasets, provided as SingleCellExperiment objects with cell- and gene-level metadata.

License CC0

NeedsCompilation no

Depends SingleCellExperiment

Imports SummarizedExperiment, S4Vectors, ExperimentHub

Suggests BiocStyle, knitr, rmarkdown

VignetteBuilder knitr

Encoding UTF-8

biocViews ExperimentHub, ExperimentData, ExpressionData, SequencingData, RNASeqData

BuildResaveData no

RoxygenNote 6.1.1

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

git_branch master

git_last_commit bc6c82c

git_last_commit_date 2019-06-16

Date/Publication 2019-06-18

Author Davide Risso [aut, cre, cph], Michael Cole [aut], Aaron Lun [ctb]

Maintainer Davide Risso <risso.davide@gmail.com>

R topics documented:

scRNAseq-package *Collection of Public Single-Cell RNA-Seq Datasets*

Description

Gene-level counts for a collection of public scRNA-seq datasets, provided as SingleCellExperiment objects with cell- and gene-level metadata.

Details

This package contains a collection of three publicly available single-cell RNA-seq datasets.

The dataset `fluidigm` contains 65 cells from Pollen et al. (2014), each sequenced at high and low coverage.

The dataset `th2` contains 96 T helper cells from Mahata et al. (2014).

The dataset `allen` contains 379 cells from the mouse visual cortex. This is a subset of the data published in Tasic et al. (2016).

See the package vignette for details on the pre-processing of the data.

Author(s)

NA

Maintainer: NA

References

Pollen, Nowakowski, Shuga, Wang, Leyrat, Lui, Li, Szpankowski, Fowler, Chen, Ramalingam, Sun, Thu, Norris, Lebofsky, Toppani, Kemp II, Wong, Clerkson, Jones, Wu, Knutsson, Alvarado, Wang, Weaver, May, Jones, Unger, Kriegstein, West. Low-coverage single-cell mRNA sequencing reveals cellular heterogeneity and activated signaling pathways in developing cerebral cortex. Nature Biotechnology, 32, 1053-1058 (2014).

Mahata, Zhang, Kolodziejczyk, Proserpio, Haim-Vilmovsky, Taylor, Hebenstreit, Dingler, Moignard, Gottgens, Arlt, McKenzie, Teichmann. Single-Cell RNA Sequencing Reveals T Helper Cells Synthesizing Steroids De Novo to Contribute to Immune Homeostasis. Cell Reports, 7(4): 1130–1142 (2014).

Tasic, Menon, Nguyen, Kim, Jarsky, Yao, Levi, Gray, Sorensen, Dolbeare, Bertagnolli, Goldy, Shapovalova, Parry, Lee, Smith, Bernard, Madisen, Sunkin, Hawrylycz, Koch, Zeng. Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. Nature Neuroscience, 19, 335–346 (2016).

`ReprocessedAllenData` *Reprocessed single-cell data sets*

Description

Obtain the legacy count matrices for three publicly available single-cell RNA-seq datasets. Raw sequencing data were downloaded from NCBI's SRA or from EBI's ArrayExpress, aligned to the relevant genome build and used to quantify gene expression.

Usage

```
ReprocessedAllenData()

ReprocessedTh2Data()

ReprocessedFluidigmData()
```

Details

`ReprocessedFluidigmData` returns a dataset of 65 cells from Pollen et al. (2014), each sequenced at high and low coverage (SRA accession SRP041736).

`ReprocessedTh2Data` returns a dataset of 96 T helper cells from Mahata et al. (2014), obtained from ArrayExpress accession E-MTAB-2512. This will contain spike-in information labelled with `isSpike`.

`ReprocessedAllenData` return a dataset of 379 cells from Tasic et al. (2016). This is a reprocessed subset of the data from `TasicBrainData`, It also contains spike-in information labelled with `isSpike`.

In each dataset, the first columns of the `colData` are sample quality metrics from FastQC and Picard. The remaining fields were obtained from the original study in their GEO/SRA submission and/or as Supplementary files in the associated publication. These two categories of `colData` are distinguished by a `which_qc` element in the `metadata`, which contains the names of the quality-related columns in each object.

Value

A SingleCellExperiment object containing one or more expression matrices, column metadata and (possibly) spike-in information.

Pre-processing details

FASTQ files were either obtained directly from ArrayExpress, or converted from SRA files (downloaded from the Sequence Read Archive) using the SRA Toolkit.

Reads were aligned with TopHat (v. 2.0.11) to the appropriate reference genome (GRCh38 for human samples, GRCm38 for mouse). RefSeq mouse gene annotation (GCF_000001635.23_GRCm38.p3) was downloaded from NCBI on Dec. 28, 2014. RefSeq human gene annotation (GCF_000001405.28) was downloaded from NCBI on Jun. 22, 2015.

featureCounts (v. 1.4.6-p3) was used to compute gene-level read counts. Cufflinks (v. 2.2.0) was used to compute gene-leve FPKMs. Reads were also mapped to the transcriptome using RSEM (v. 1.2.19) to compute read counts and TPM's.

FastQC (v. 0.10.1) and Picard (v. 1.128) were used to compute sample quality control (QC) metrics. However, no filtering on the QC metrics has been performed for any dataset.

References

Pollen AA et al. (2014). Low-coverage single-cell mRNA sequencing reveals cellular heterogeneity and activated signaling pathways in developing cerebral cortex. *Nat. Biotechnol.* 32(10), 1053-8.

Mahata B et al. (2014). Single-cell RNA sequencing reveals T helper cells synthesizing steroids de novo to contribute to immune homeostasis. *Cell Rep*, 7(4), 1130-42.

Tasic A et al. (2016). Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. *Nat. Neurosci.* 19(2), 335-46.

Examples

```
sce <- ReprocessedAllenData()
```

`SegerstolpePancreasData` *Obtain the Segerstolpe pancreas data*

Description

Download and cache the Segerstolpe pancreas single-cell RNA-seq (scRNA-seq) dataset from ExperimentHub, returning a SingleCellExperiment object for further use.

Usage

```
SegerstolpePancreasData()
```

Details

This function provides the pancreas scRNA-seq data from Segerstolpe et al. (2015) in the form of a SingleCellExperiment object with a single matrix of read counts.

Row data contains fields for the gene symbol and RefSeq transcript IDs corresponding to each gene. Spike-ins are specially labelled with the `isSpike` function.

Column metadata were extracted from the `Characteristcs` fields of the SDRF file for ArrayExpress E-MTAB-5061. This contains information such as the cell type labels and patient status.

Value

A SingleCellExperiment object.

Author(s)

Aaron Lun

References

Segerstolpe A et al. (2016). Single-cell transcriptome profiling of human pancreatic islets in health and type 2 diabetes. *Cell Metab.* 24(4), 593-607.

Examples

```
sce <- SegerstolpePancreasData()
```

`TasicBrainData` *Obtain the Tasic brain data*

Description

Download and cache the Tasic brain single-cell RNA-seq (scRNA-seq) dataset from ExperimentHub, returning a SingleCellExperiment object for further use.

Usage

```
TasicBrainData()
```

Details

This function provides the brain scRNA-seq data from Tasic et al. (2015) in the form of a SingleCellExperiment object with a single matrix of read counts.

Column metadata is provided in the same form as supplied in GSE71585. This contains information such as the reporter gene expressed in each cell, the mouse line, dissection type and so on.

Rows corresponding to spike-in transcripts are labelled with the `isSpike` function. Note that some of the spike-in rows have `NA` observations for some (but not all) cells.

The last 9 columns (containing `_CTX_` in their names) correspond to no-cell control libraries.

Value

A SingleCellExperiment object.

Author(s)

Aaron Lun

References

Tasic A et al. (2016). Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. *Nat. Neurosci.* 19(2), 335-46.

Examples

```
sce <- TasicBrainData()
```

`ZeiselBrainData` *Obtain the Zeisel brain data*

Description

Download and cache the Zeisel brain single-cell RNA-seq (scRNA-seq) dataset from ExperimentHub, returning a SingleCellExperiment object for further use.

Usage

```
ZeiselBrainData()
```

Details

This function provides the brain scRNA-seq data from Zeisel et al. (2015) in the form of a SingleCellExperiment object with a single matrix of UMI counts.

Row data contains a single `"featureType"` field describing the type of each feature (endogenous genes, mitochondrial genes, spike-in transcripts and repeats). Spike-ins are also specially labelled with the `isSpike` function.

Column metadata is provided in the same form as supplied in `http://linnarssonlab.org/cortex/`. This contains information such as the cell diameter and the published cell type annotations.

Value

A SingleCellExperiment object.

Author(s)

Aaron Lun

References

Zeisel A et al. (2015). Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science* 347(6226), 1138-42.

Examples

```
sce <- ZeiselBrainData()
```

Index