Package ‘TENxPBMCData’

January 25, 2024

Title  PBMC data from 10X Genomics
Version  1.20.0
Description  Single-cell RNA-seq data for on PBMC cells, generated by 10X Genomics.
License  CC BY 4.0
Depends  SingleCellExperiment, HDF5Array
Imports  AnnotationHub, ExperimentHub
Suggests  rmarkdown, knitr, BiocStyle, snow, BiocFileCache, BiocParallel
VignetteBuilder  knitr
biocViews  SequencingData, RNASeqData, ExpressionData, ExperimentHub, ExperimentData, SingleCellData
NeedsCompilation  no

git_url  https://git.bioconductor.org/packages/TENxPBMCData
git_branch  RELEASE_3_18
git_last_commit  0039287
git_last_commit_date  2023-10-24
Repository  Bioconductor 3.18
Date/Publication  2024-01-25

Author  Kasper D. Hansen [aut], Davide Risso [aut], Milan Malfait [ctb], Jeroen Gilis [ctb], Theodore Killian [ctb], Murat Cem Kose [ctb], Chong Tang [ctb], Teun van den Brand [ctb], Dania Machlab [ctb], Stephanie Hicks [aut, cre]

Maintainer  Stephanie Hicks <shicks19@jhu.edu>
Description

Various single-cell RNA-seq data on PBMC, generated by 10X Genomics.

Usage

```r
TENxPBMCData(dataset = c("pbmc4k", "pbmc68k",
                       "frozen_pbmc_donor_a", "frozen_pbmc_donor_b",
                       "frozen_pbmc_donor_c", "pbmc33k", "pbmc3k",
                       "pbmc6k", "pbmc8k", "pbmc5k-CITEseq"),
          as.sparse = TRUE)
```

Arguments

- **dataset**: Which PBMC dataset from 10X Genomics should be retrieved?
- **as.sparse**: Logical, specifies whether the underlying HDF5 dataset should be treated as sparse or not - will be passed to the call to `HDF5Array()`. Defaults to `TRUE`, i.e. by using the `DelayedArray` infrastructure.

Details

Single-cell RNA-seq and CITE-seq data were generated by 10X Genomics at various times, using different versions of CellRanger, different chemistries and different genome builds. For details, see the 10X website.

We obtained ‘filtered’ data and generated `SingleCellExperiment` containers with data stored as an HDF5 Assay.

As `rowData` we include ENSEMBL and Symbol_TENx which are ENSEMBL gene ID and gene symbol provided by TENx genomics and a remapping of the Ensembl identifier to a Hugo gene symbol as columns Symbol using the `org.Hs.eg.db` package. The difference between Symbol and Symbol_TENx is that the former has many missing values (for non-protein coding genes) whereas the later is technically not a Hugo gene symbol. In case of CITE-seq data, the `rowData` has an additional column `Type` specifying if the counts are "Gene Expression" or "Antibody Capture". Note that there is a separate `rowData` for the `altExp`. 
TENxPBMCData

Value
A SingleCellExperiment object with a HDF5Matrix in the counts assay, which contains UMI counts for each gene in each cell. Row- and column-level metadata are also provided. In case of CITE-seq data, the "Antibody Capture" counts are stored in the altExp of the SingleCellExperiment.

Author(s)
Kasper D. Hansen, Stephanie Hicks, Davide Risso

References
10X Genomics (2017). 1.3 Million PBMC Cells from E18 Mice. https://support.10xgenomics.com/single-cell-gene-expression/datasets/1.3.0/1M_neurons

See Also
SingleCellExperiment

Examples
sce <- TENxPBMCData()
sce
lib.size <- colSums(assay(sce))
hist(log10(lib.size))

# CITE-seq data
sce <- TENxPBMCData(dataset = "pbmc5k-CITEseq")
sce
altExp(sce)
Index

SingleCellExperiment, 3
TENxPBMCData, 2