Package 'TENxBUSData'

October 27, 2026

Type Package

Title Single cell dataset from 10x in BUS format

Version 1.26.0

Description Download Barcode, UMI, and Set (BUS) format of 10x datasets from within R. This package accompanies the package BUSpaRse, which can load BUS format into R as a sparse matrix, and which has utility functions related to using the C++ command line package bustools.

License BSD_2_clause + file LICENSE

Encoding UTF-8

Imports ExperimentHub, AnnotationHub, utils, BiocGenerics

biocViews SingleCellData, SequencingData, ExperimentHub, ExperimentData, ExpressionData

BugReports https://github.com/BUStools/TENxBUSData/issues

URL https://github.com/BUStools/TENxBUSData

RoxygenNote 7.1.0

Roxygen list(markdown = TRUE)

Suggests knitr, rmarkdown, BiocStyle

VignetteBuilder knitr

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

git_branch RELEASE_3_23

git_last_commit badd53e

git_last_commit_date 2026-04-28

Repository Bioconductor 3.23

Date/Publication 2026-10-27

Author Lambda Moses [aut, cre] (ORCID: <https://orcid.org/0000-0002-7092-9427>)

Maintainer Lambda Moses <dlu2@caltech.edu>

Contents

TENxBUSData 2

Index **4**

`TENxBUSData` *TENxBUSData: 10x human and mouse cell mixture datasets in the BUS format*

Description

This package provides 5 10x datasets in the BUS format, to be downloaded from within R. The files downloaded from this package are sufficient to generate a sparse matrix with package BUSpaRse to be used for downstream analysis with Seurat. Human and mouse transcriptomes from Ensembl version 99 were used to generate the BUS format from FASTQ files. This package server the following purposes: First, to demonstrate the kallisto bus workflow and downstream analyses. Second, for advanced users to experiment with other ways to collapse UMIs mapped to multiple genes and with other ways of barcode correction. The datasets are on `ExperimentHub`.

This function will download the 10x datasets, already processed and stored in the BUS format, from `ExperimentHub`. This function will decompress the downloaded file and return the directory where the files necessary to construct the sparse matrix with `BUSpaRse` are located.

Usage

```
TENxBUSData(
  file_path,
  dataset = c("hgmm100", "hgmm1k", "pbmc1k", "neuron10k"),
  force = FALSE,
  verbose = TRUE
)
```

Arguments

`file_path`	Character vector of length 1, specifying where to download the data.
`dataset`	Character, must be one of "hgmm100", "hgmm1k", "pbmc1k", "neuron10k".
`force`	Logical, whether to force redownload if the files are already present. Defaults to `FALSE`.
`verbose`	Whether to display progress of download.

Details

The following files will be downloaded:

`matrix.ec` Text file with 2 columns. The first column is the index of equivalence classes used in BUS files. The second column is the equivalence classes themselves, consisted of sets of transcript indices from the kallisto index.

`output.sorted` Binary BUS file sorted by barcode, UMI, and equivalence classes. This can be directly parsedd by `bustools` to get the gene count matrix.

`output.sorted.txt` Sorted BUS file in text form. This is to be used with `BUSpaRse` to get the gene count matrix.

`transcripts.txt` Transcripts included in the BUS data, in the same order as in the kallisto index.

The gzipped file downloaded from `ExperimentHub` will be in a cache directory that can be retrieved by `getExperimentHubOption("CACHE")`. The cache will remain even if the decompressed files in the directory specified when calling this function are deleted. To delete cache, use `removeCache`.

Value

Character, directory to be used in `BUSpaRse`.

Datasets

hgmm100 100 1:1 Mixture of Fresh Frozen Human (HEK293T) and Mouse (NIH3T3) Cells. The raw data can be found here: https://support.10xgenomics.com/single-cell-gene-expression/datasets/2.1.0/hgmm_100

hgmm1k 1k 1:1 Mixture of Fresh Frozen Human (HEK293T) and Mouse (NIH3T3) Cells (v3 chemistry). The raw data can be found here: https://support.10xgenomics.com/single-cell-gene-expression/datasets/3.0.0/hgmm_1k_v3

pbmc1k 1k PBMCs from a Healthy Donor (v3 chemistry). The raw data can be found here: https://support.10xgenomics.com/single-cell-gene-expression/datasets/3.0.0/pbmc_1k_v3

neuron10k 10k Brain Cells from an E18 Mouse (v3 chemistry). The raw data can be found here: https://support.10xgenomics.com/single-cell-gene-expression/datasets/3.0.0/neuron_10k_v3

Examples

```
TENxBUSData(".", dataset = "hgmm100")
```

Index

removeCache, *3*

TENxBUSData, 2