Package 'scaeData'

November 10, 2026

Type Package

Title Data Package for SingleCellAlleleExperiment

Version 1.8.0

Date 2024-03-27

Description Contains default datasets used by the Bioconductor package SingleCellAlleleExperiment. The raw FASTQ files were sourced from publicly accessible datasets provided by 10x Genomics. Subsequently, our scIGD snakemake workflow was employed to process these FASTQ files. The resulting output from scIGD constitutes to the contents of this data package.

Depends R (>= 4.4.0)

Imports ExperimentHub

Suggests knitr, rmarkdown, markdown, SingleCellAlleleExperiment, Matrix, BiocStyle

biocViews ExperimentHub, ExperimentData, Homo_sapiens_Data, SingleCellData

License MIT + file LICENSE

VignetteBuilder knitr

Encoding UTF-8

Roxygen list(markdown = TRUE)

RoxygenNote 7.3.1

URL https://github.com/AGImkeller/scaeData

BugReports https://github.com/AGImkeller/scIGD/issues

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

git_branch RELEASE_3_23

git_last_commit 91e094e

git_last_commit_date 2026-04-28

Repository Bioconductor 3.23

Date/Publication 2026-11-10

Author Ahmad Al Ajami [aut, cre] (ORCID:
<https://orcid.org/0009-0006-5615-7447>),
Jonas Schuck [aut] (ORCID: <https://orcid.org/0009-0003-5705-4579>),
Federico Marini [aut] (ORCID: <https://orcid.org/0000-0003-3252-7758>),
Katharina Imkeller [aut] (ORCID:
<https://orcid.org/0000-0002-5177-0852>)

Maintainer Ahmad Al Ajami <alajami@med.uni-frankfurt.de>

Contents

demo_dir_file 2
ehub_out 3
get_barcodes_10k 3
get_barcodes_20k 4
get_barcodes_5k 4
get_counts_10k 5
get_counts_20k 5
get_counts_5k 6
get_features_10k 6
get_features_20k 7
get_features_5k 7
scaeData-pkg 8
scaeDataGet 8

Index **10**

`demo_dir_file` *Split getter output*

Description

Internal function used by `ehoub_out()` that splits the output from the getter into a file path and the corresponding file name. This is necessary as the read-in function `read_allele_counts()` from the `SingleCellAlleleExperiment` package expects a directory path as well as the names of each expected file.

Usage

```
demo_dir_file(ehub_dir, dir = TRUE)
```

Arguments

`ehub_dir`	character string that is retrieved by the getters
`dir`	binary if the output should contain file path and file name or only file name

Value

list containing (file path) and file name

ehub_out *Build output list per dataset*

Description

Internal function used by `scaeDataGet()` to build a list containing the file path

Usage

```
ehub_out(bc_dir, feature_dir, mtx_dir)
```

Arguments

`bc_dir`	character string containing full path to barcode file
`feature_dir`	character string containing full path to feature file
`mtx_dir`	character string containing full path to matrix file

Value

list with four elements containing file path to the directory containing all files and each file name for barcodes, features and matrix

get_barcodes_10k *Get barcode identifiers for* `pbmc-10k` *dataset*

Description

Internal getter function that retrieves the `pbmc_10k_barcodes.txt` from `ExperimentHub` (`eh[["EH9456"]]`). This is one out of 3 files that are necessary to load for using the data in a *SingleCellAlleleExperiment* data structure. File is generated with the *scIGD* workflow.

Usage

```
get_barcodes_10k()
```

Value

character string containing the file path of the barcode file

---

`get_barcodes_20k` *Get barcode identifiers for* `pbmc-20k` *dataset*

Description

Internal getter function that retrieves the `pbmc_20k_barcodes.txt` from `ExperimentHub` (`eh[["EH9459"]]`). This is one out of 3 files that are necessary to load for using the data in a *SingleCellAlleleExperiment* data structure. File is generated with the *scIGD* workflow.

Usage

```
get_barcodes_20k()
```

Value

character string containing the file path of the barcode file

---

`get_barcodes_5k` *Get barcode identifiers for* `pbmc-5k` *dataset*

Description

Internal getter function that retrieves the `pbmc_5k_barcodes.txt` from `ExperimentHub` (`eh[["EH9453"]]`). This is one out of 3 files that are necessary to load for using the data in a *SingleCellAlleleExperiment* data structure. File is generated with the *scIGD* workflow.

Usage

```
get_barcodes_5k()
```

Value

character string containing the file path of the barcode file

get_counts_10k *Get quantification matrix for* pbmc-10k *dataset*

Description

Internal getter function that retrieves the pbmc_10k_count_mtx.mtx from ExperimentHub (eh[["EH9458"]]). This is one out of 3 files that are necessary to load for using the data in a *SingleCellAlleleExperiment* data structure. File is generated with the *scIGD* workflow.

Usage

```
get_counts_10k()
```

Value

character string containing the file path of the matrix file

get_counts_20k *Get quantification matrix for* pbmc-20k *dataset*

Description

Internal getter function that retrieves the pbmc_20k_count_mtx.mtx from ExperimentHub (eh[["EH9461"]]). This is one out of 3 files that are necessary to load for using the data in a *SingleCellAlleleExperiment* data structure. File is generated with the *scIGD* workflow.

Usage

```
get_counts_20k()
```

Value

character string containing the file path of the matrix file

get_counts_5k *Get quantification matrix for* pbmc-5k *dataset*

Description

Internal getter function that retrieves the pbmc_5k_count_mtx.mtx from ExperimentHub (eh[["EH9455"]]). This is one out of 3 files that are necessary to load for using the data in a *SingleCellAlleleExperiment* data structure. File is generated with the *scIGD* workflow.

Usage

```
get_counts_5k()
```

Value

character string containing the file path of the matrix file

get_features_10k *Get feature identifiers for* pbmc-10k *dataset*

Description

Internal getter function that retrieves the pbmc_10k_features.txt from ExperimentHub (eh[["EH9457"]]). This is one out of 3 files that are necessary to load for using the data in a *SingleCellAlleleExperiment* data structure. File is generated with the *scIGD* workflow.

Usage

```
get_features_10k()
```

Value

character string containing the file path of the feature file

get_features_20k *Get feature identifiers for* pbmc-20k *dataset*

Description

Internal getter unction that retrieves the `pbmc_20k_features.txt` from `ExperimentHub` (`eh[["EH9460"]]`). This is one out of 3 files that are necessary to load for using the data in a *SingleCellAlleleExperiment* data structure. File is generated with the *scIGD* workflow.

Usage

```
get_features_20k()
```

Value

character string containing the file path of the feature file

get_features_5k *Get feature identifiers for* pbmc-5k *dataset*

Description

Internal getter function that retrieves the `pbmc_5k_features.txt` from `ExperimentHub` (`eh[["EH9454"]]`). This is one out of 3 files that are necessary to load for using the data in a *SingleCellAlleleExperiment* data structure. File is generated with the *scIGD* workflow.

Usage

```
get_features_5k()
```

Value

character string containing the file path of the features file

scaeData-pkg *scaeData*

Description

scaeData: Data Package for SingleCellAlleleExperiment

Details

This package contains some exemplary datasets used by the Bioconductor package `SingleCellAlleleExperiment`. The raw FASTQ files were sourced from publicly accessible datasets provided by 10x Genomics. Subsequently, our `scIGD` snakemake workflow was employed to process these FASTQ files in order to obtain allele-level quantifications. The resulting output from `scIGD` constitutes to the contents of this data package.

Author(s)

Maintainer: Ahmad Al Ajami <alajami@med.uni-frankfurt.de> (ORCID)

Authors:

- Jonas Schuck <schuck@med.uni-frankfurt.de> (ORCID)
- Federico Marini <marinif@uni-mainz.de> (ORCID)
- Katharina Imkeller <imkeller@med.uni-frankfurt.de> (ORCID)

See Also

https://github.com/AGImkeller/scIGD/ for the definition of the quantification workflow.

scaeDataGet *Download and process demo dataset*

Description

Wrapper function used to retrieve a list containing the file path and file names for the chosen dataset. The corresponding dataset is retrieved from `ExperimentHub`. The following demo datasets are available:

- **pbmc_5k**: using the following ExperimentHub entries: **barcodes**(EH9453); **features**(EH9454); **counts matrix**(EH9455)
- **pbmc_10k**: using the following ExperimentHub entries: **barcodes**(EH9456); **features**(EH9457); **counts matrix**(EH9458)
- **pbmc_20k**: using the following ExperimentHub entries: **barcodes**(EH9459); **features**(EH9460); **counts matrix**(EH9461)

Usage

```
scaeDataGet(dataset = c("pbmc_5k", "pbmc_10k", "pbmc_20k"))
```

Arguments

dataset	character vector describing which dataset should be retrieved. Choose one as input.

Value

list with four elements containing file path and file names for the chosen dataset

Examples

```
if (interactive()) {
    scae_data_5k <- scaeDataGet(dataset = "pbmc_5k")
    scae_data_5k
}
```

Index

∗ **internal**
scaeData-pkg, 8

demo_dir_file, 2

ehub_out, 3

get_barcodes_10k, 3
get_barcodes_20k, 4
get_barcodes_5k, 4
get_counts_10k, 5
get_counts_20k, 5
get_counts_5k, 6
get_features_10k, 6
get_features_20k, 7
get_features_5k, 7

scaeData *(*scaeData-pkg*)*, 8
scaeData-package *(*scaeData-pkg*)*, 8
scaeData-pkg, 8
scaeDataGet, 8