

# Package ‘scaeData’

May 16, 2024

**Type** Package

**Title** Data Package for SingleCellAlleleExperiment

**Version** 1.1.1

**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** devel

**git\_last\_commit** 039d55a

**git\_last\_commit\_date** 2024-05-02

**Repository** Bioconductor 3.20

**Date/Publication** 2024-05-16

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

---

## Description

Internal function used by ehub\_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	<i>Build output list per dataset</i>
----------	--------------------------------------

---

**Description**

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

**Usage**

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

**Arguments**

<code>bc_dir</code>	character string containing full path to barcode file
<code>feature_dir</code>	character string containing full path to feature file
<code>mtx_dir</code>	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

---

<code>get_barcodes_10k</code>	<i>Get barcode identifiers for pbmc-10k dataset</i>
-------------------------------	---

---

**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	<i>Get barcode identifiers for pbmc-20k dataset</i>
------------------	---

---

**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	<i>Get barcode identifiers for pbmc-5k dataset</i>
-----------------	--

---

**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	<i>Get quantification matrix for pbmc-10k dataset</i>
----------------	---

---

**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	<i>Get quantification matrix for pbmc-20k dataset</i>
----------------	---

---

**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	<i>Get quantification matrix for pbmc-5k dataset</i>
---------------	--

---

**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	<i>Get feature identifiers for pbmc-10k dataset</i>
------------------	---

---

**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	<i>Get feature identifiers for pbmc-20k dataset</i>
------------------	---

---

**Description**

Internal getter function 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	<i>Get feature identifiers for pbmc-5k dataset</i>
-----------------	--

---

**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