

Package ‘SCArray’

April 7, 2023

Type Package

Title Large-scale single-cell RNA-seq data manipulation with GDS files

Version 1.6.0

Date 2022-04-20

Depends R (>= 3.5.0), gdsfmt (>= 1.27.4), methods, DelayedArray (>= 0.16.0)

Imports BiocGenerics, S4Vectors, IRanges, utils, SummarizedExperiment, SingleCellExperiment, DelayedMatrixStats

Suggests Matrix, scater, uwot, RUnit, knitr, markdown, rmarkdown, rhdf5, HDF5Array

Description Provides large-scale single-cell RNA-seq data manipulation using Genomic Data Structure (GDS) files. It combines dense and sparse matrices stored in GDS files and the Bioconductor infrastructure framework (SingleCellExperiment and DelayedArray) to provide out-of-memory data storage and large-scale manipulation using the R programming language.

License GPL-3

VignetteBuilder knitr

ByteCompile TRUE

URL <https://github.com/AbbVie-ComputationalGenomics/SCArray>

biocViews Infrastructure, DataRepresentation, DataImport, SingleCell, RNASeq

git_url <https://git.bioconductor.org/packages/SCArray>

git_branch RELEASE_3_16

git_last_commit 56c7925

git_last_commit_date 2022-11-01

Date/Publication 2023-04-07

Author Xiuwen Zheng [aut, cre] (<<https://orcid.org/0000-0002-1390-0708>>)

Maintainer Xiuwen Zheng <xiuwen.zheng@abbvie.com>

R topics documented:

SCArray-package	2
scArray	3
scClose	3
scConvGDS	4
scExperiment	5
scHDF2GDS	6
scMEX2GDS	7
scObj	8
scOpen	9
Index	10

SCArray-package	<i>Large-scale single-cell RNA-seq data manipulation with GDS files</i>
-----------------	---

Description

The package combines dense/sparse matrices stored in GDS files and the Bioconductor infrastructure framework to provide out-of-memory data storage and manipulation using the R programming language.

Details

Package: SCArray
Type: Package
License: GPL version 3

Author(s)

Xiuwen Zheng <xiuwen.zheng@abbvie.com>

Examples

```
# a GDS file for SingleCellExperiment
fn <- system.file("extdata", "LaMannoBrainData.gds", package="SCArray")

sce <- scExperiment(fn)
sce
```

scArray	<i>Get an DelayedArray instance</i>
---------	-------------------------------------

Description

Gets an DelayedArray instance from a single-cell GDS file.

Usage

```
scArray(gdsfile, varname)
```

Arguments

gdsfile	character for a file name, or a single-cell GDS object with class SCArrayFileClass
varname	character for the node name in the GDS file

Value

Return an object of class [DelayedArray](#).

Author(s)

Xiuwen Zheng

See Also

[scOpen](#), [scExperiment](#)

Examples

```
# a GDS file for SingleCellExperiment
fn <- system.file("extdata", "LaMannoBrainData.gds", package="SCArray")

cnt <- scArray(fn, "counts")
cnt
```

scClose	<i>Close the Single-cell GDS File</i>
---------	---------------------------------------

Description

Closes a single-cell GDS file which is open.

Usage

```
scClose(gdsfile)
```

Arguments

gdsfile a single-cell GDS object with class SCArraryFileClass

Value

None.

Author(s)

Xiuwen Zheng

See Also

[scOpen](#)

Examples

```
# a GDS file for SingleCellExperiment
fn <- system.file("extdata", "LaMannoBrainData.gds", package="SCArray")

# open the GDS file
(f <- scOpen(fn))

# read
cell.id <- read.gdsn(index.gdsn(f, "feature.id"))
samp.id <- read.gdsn(index.gdsn(f, "sample.id"))

scClose(f)
```

scConvGDS

Create a GDS file

Description

Creates a single-cell GDS file from an R object.

Usage

```
scConvGDS(obj, outfn, save.sp=TRUE, type=c("float32", "float64", "int32"),
  compress="LZMA_RA", clean=TRUE, verbose=TRUE)
```

Arguments

obj a dense/sparse matrix, DelayedMatrix, SummarizedExperiment or SingleCell-Experiment

outfn the output file name in GDS format

save.sp if TRUE, save it to a sparse matrix in GDS; otherwise, store dense matrix

type	numeric data type in the output file
compress	the compression method, see add.gdsn
clean	TRUE
verbose	if TRUE, show information

Value

Return the path of the output file.

Author(s)

Xiuwen Zheng

See Also

[scOpen](#), [scClose](#), [scMEX2GDS](#), [scHDF2GDS](#)

Examples

```
# load a SingleCellExperiment object
fn <- system.file("extdata", "LaMannoBrainSub.rds", package="SCArray")
sce <- readRDS(fn)

scConvGDS(sce, "test.gds")

# remove the temporary output file
unlink("test.gds")
```

scExperiment	<i>Get a SummarizedExperiment</i>
--------------	-----------------------------------

Description

Gets an instance of SingleCellExperiment or SummarizedExperiment.

Usage

```
scExperiment(gdsfile, sce=TRUE, use.names=TRUE, load.row=TRUE, load.col=TRUE)
```

Arguments

gdsfile	character for a file name, or a single-cell GDS object with class SCArrayFileClass
sce	if TRUE, return an instance of SingleCellExperiment, otherwise an instance of SummarizedExperiment
use.names	if TRUE, load dimnames from 'feature.id' and 'sample.id'
load.row	TRUE for loading rowData from the gds node "feature.data" in gdsfile
load.col	TRUE for loading colData from the gds node "sample.data" in gdsfile

Value

Return an instance of [SingleCellExperiment](#) or [SummarizedExperiment](#).

Author(s)

Xiuwen Zheng

See Also

[scOpen](#), [scClose](#)

Examples

```
# a GDS file for SingleCellExperiment
fn <- system.file("extdata", "LaMannoBrainData.gds", package="SCArray")

sce <- scExperiment(fn)
sce
```

scHDF2GDS

Convert HDF5 files to GDS

Description

Creates a single-cell GDS file from Cell Ranger HDF5 files.

Usage

```
scHDF2GDS(h5_fn, outfn, group=c("matrix", "mm10"), feature_path=character(),
  type=c("float32", "float64", "int32"), compress="LZMA_RA", clean=TRUE,
  verbose=TRUE)
```

Arguments

h5_fn	the input HDF5 file name
outfn	the output file name in GDS format
group	the name of the group in the HDF5 file where the sparse matrix is stored; if there are more than one group names, the first existing group in the HDF5 file is used; "mm10" is usually used for 10x Genomics datasets
feature_path	a character vector for feature variables, otherwise detecting automatically
type	numeric data type in the output file
compress	the compression method, see add.gdsn
clean	TRUE
verbose	if TRUE, show information

Value

Return the path of the output file.

Author(s)

Xiuwen Zheng

See Also

[scConvGDS](#), [scMEX2GDS](#)

scMEX2GDS	<i>Convert MEX files to GDS</i>
-----------	---------------------------------

Description

Creates a single-cell GDS file from Cell Ranger MEX files.

Usage

```
scMEX2GDS(feature_fn, barcode_fn, mtx_fn, outfn,  
           feature_colnm=c("id", "gene", "feature_type"),  
           type=c("float32", "float64", "int32"), compress="LZMA_RA", clean=TRUE,  
           verbose=TRUE)
```

Arguments

feature_fn	the input file name for features
barcode_fn	the input file name for barcodes
mtx_fn	the input count matrix in MEX format
outfn	the output file name in GDS format
feature_colnm	the column names used in feature_fn
type	numeric data type in the output file
compress	the compression method, see add.gdsn
clean	TRUE
verbose	if TRUE, show information

Value

Return the path of the output file.

Author(s)

Xiuwen Zheng

See Also

[scConvGDS](#), [schDF2GDS](#)

`scObj`*Seed-aware DelayedArray Object*

Description

Convert to SC_GDSArray for utilizing seed-aware DelayedArray functions.

Usage

```
scObj(obj, verbose=TRUE)
```

Arguments

<code>obj</code>	a SummarizedExperiment, SingleCellExperiment or DelayedArray object
<code>verbose</code>	if TRUE, show information

Value

Return an object with SC_GDSMatrix/SC_GDSArray.

Author(s)

Xiuwen Zheng

See Also

[scArray](#), [scExperiment](#)

Examples

```
# a GDS file for SingleCellExperiment
fn <- system.file("extdata", "LaMannoBrainData.gds", package="SCArray")

cnt <- scArray(fn, "counts")
cnt # it is SC_GDSMatrix

logcnt <- log2(cnt + 1)
logcnt # it is a general DelayedMatrix

# convert to SC_GDSMatrix
lcnt <- scObj(logcnt)
lcnt
```

`scOpen`*Open a Single-cell GDS File*

Description

Opens a single-cell GDS file.

Usage

```
scOpen(gdsfn, readonly=TRUE, allow.duplicate=TRUE)
```

Arguments

<code>gdsfn</code>	the input file name
<code>readonly</code>	whether read-only or not
<code>allow.duplicate</code>	if TRUE, it is allowed to open a GDS file with read-only mode when it has been opened in the same R session

Value

Return an object of class `SCArrayFileClass` inherited from `gds.class`.

Author(s)

Xiuwen Zheng

See Also

`scClose`, `scArray`

Examples

```
# a GDS file for SingleCellExperiment
fn <- system.file("extdata", "LaMannoBrainData.gds", package="SCArray")

# open the GDS file
(f <- scOpen(fn))

# read a GDS file
cell.id <- read.gdsn(index.gdsn(f, "feature.id"))
samp.id <- read.gdsn(index.gdsn(f, "sample.id"))

# get a DelayedArray object
(cnt <- scArray(f, "counts"))

scClose(f)
```

Index

* **CellRanger**

scHDF2GDS, [6](#)

scMEX2GDS, [7](#)

* **GDS**

scArray, [3](#)

SCArray-package, [2](#)

scClose, [3](#)

scConvGDS, [4](#)

scExperiment, [5](#)

scHDF2GDS, [6](#)

scMEX2GDS, [7](#)

scObj, [8](#)

scOpen, [9](#)

* **SingleCell**

scArray, [3](#)

SCArray-package, [2](#)

scClose, [3](#)

scConvGDS, [4](#)

scExperiment, [5](#)

scHDF2GDS, [6](#)

scMEX2GDS, [7](#)

scObj, [8](#)

scOpen, [9](#)

add.gdsn, [5–7](#)

DelayedArray, [3](#)

gds.class, [9](#)

scArray, [3](#), [8](#), [9](#)

SCArray-package, [2](#)

scClose, [3](#), [5](#), [6](#), [9](#)

scConvGDS, [4](#), [7](#)

scExperiment, [3](#), [5](#), [8](#)

scHDF2GDS, [5](#), [6](#), [7](#)

scMEX2GDS, [5](#), [7](#), [7](#)

scObj, [8](#)

scOpen, [3–6](#), [9](#)

SingleCellExperiment, [6](#)

SummarizedExperiment, [6](#)