

Package: splatter (via r-universe)

September 3, 2024

Type Package

Title Simple Simulation of Single-cell RNA Sequencing Data

Version 1.29.0

Date 2024-05-01

Description Splatter is a package for the simulation of single-cell RNA sequencing count data. It provides a simple interface for creating complex simulations that are reproducible and well-documented. Parameters can be estimated from real data and functions are provided for comparing real and simulated datasets.

License GPL-3 + file LICENSE

LazyData FALSE

Depends R (>= 4.0), SingleCellExperiment

Imports BiocGenerics, BiocParallel, checkmate (>= 2.0.0), crayon, edgeR, fitdistrplus, grDevices, locfit, matrixStats, methods, rlang, S4Vectors, scuttle, stats, SummarizedExperiment, utils, withr

Suggests BASiCS (>= 1.7.10), BiocManager, BiocSingular, BiocStyle, Biostrings, covr, cowplot, GenomeInfoDb, GenomicRanges, ggplot2 (>= 3.4.0), IRanges, igraph, knitr, limSolve, lme4, magick, mfa, phenopath, progress, preprocessCore, pscl, rmarkdown, scales, scater (>= 1.15.16), scDD, scran, SparseDC, spelling, testthat, VariantAnnotation, zinbwave,

biocViews SingleCell, RNASeq, Transcriptomics, GeneExpression, Sequencing, Software, ImmunoOncology

URL <https://bioconductor.org/packages/splatter/>,
<https://github.com/Oshlack/splatter>

BugReports <https://github.com/Oshlack/splatter/issues>

RoxygenNote 7.3.1

Encoding UTF-8

VignetteBuilder knitr

Language en-GB

Repository <https://lazappi.r-universe.dev>

RemoteUrl <https://github.com/Oshlack/splatter>

RemoteRef HEAD

RemoteSha 4018ddc450d91a244c5ba10cd497697aa820f87e

Contents

addGeneLengths	3
BASiCSEstimate	4
BASiCSParams	7
BASiCSSimulate	7
compareSCEs	8
diffSCEs	10
getParam	11
getParams	12
kersplatEstimate	12
KersplatParams	13
kersplatSample	15
kersplatSetup	16
kersplatSimulate	18
listSims	19
lun2Estimate	19
Lun2Params	21
lun2Simulate	21
lunEstimate	23
LunParams	24
lunSimulate	24
makeCompPanel	25
makeDiffPanel	26
makeOverallPanel	27
mfaEstimate	28
MFAParams	29
mfaSimulate	29
minimiseSCE	30
mockBulkeQTL	31
mockBulkMatrix	32
mockEmpiricalSet	32
mockGFF	33
mockVCF	34
newParams	35
Params	36
phenoEstimate	36
PhenoParams	37
phenoSimulate	38

scDDEstimate	39
SCDDParams	40
scDDSimulate	41
setParam	42
setParams	44
simpleEstimate	45
SimpleParams	46
simpleSimulate	46
sparseDCEstimate	47
SparseDCParams	48
sparseDCSimulate	49
splatEstimate	50
SplatParams	51
splatPopEstimate	53
SplatPopParams	54
splatPopParseEmpirical	55
splatPopQuantNorm	56
splatPopSimulate	57
splatPopSimulateMeans	58
splatPopSimulateSC	60
splatSimulate	61
summariseDiff	64
zinbEstimate	64
ZINBParams	66
zinbSimulate	67
Index	68

addGeneLengths	<i>Add gene lengths</i>
----------------	-------------------------

Description

Add gene lengths to an SingleCellExperiment object

Usage

```
addGeneLengths(
  sce,
  method = c("generate", "sample"),
  loc = 7.9,
  scale = 0.7,
  lengths = NULL
)
```

Arguments

sce	SingleCellExperiment to add gene lengths to.
method	Method to use for creating lengths.
loc	Location parameter for the generate method.
scale	Scale parameter for the generate method.
lengths	Vector of lengths for the sample method.

Details

This function adds simulated gene lengths to the `rowData` slot of a `SingleCellExperiment` object that can be used for calculating length normalised expression values such as TPM or FPKM. The `generate` method simulates lengths using a (rounded) log-normal distribution, with the default `loc` and `scale` parameters based on human protein-coding genes. Alternatively the `sample` method can be used which randomly samples lengths (with replacement) from a supplied vector.

Value

SingleCellExperiment with added gene lengths

Examples

```
# Default generate method
sce <- simpleSimulate()
sce <- addGeneLengths(sce)
head(rowData(sce))
# Sample method (human coding genes)
## Not run:
library(TxDb.Hsapiens.UCSC.hg19.knownGene)
library(GenomicFeatures)
txdb <- TxDb.Hsapiens.UCSC.hg19.knownGene
tx.lens <- transcriptLengths(txdb, with.cds_len = TRUE)
tx.lens <- tx.lens[tx.lens$cds_len > 0, ]
gene.lens <- max(splitAsList(tx.lens$tx_len, tx.lens$gene_id))
sce <- addGeneLengths(sce, method = "sample", lengths = gene.lens)

## End(Not run)
```

Description

Estimate simulation parameters for the BASiCS simulation from a real dataset.

Usage

```
BASiCSEstimate(  
  counts,  
  spike.info = NULL,  
  batch = NULL,  
  n = 20000,  
  thin = 10,  
  burn = 5000,  
  regression = TRUE,  
  params = newBASiCSParams(),  
  verbose = TRUE,  
  progress = TRUE,  
  ...  
)  
  
## S3 method for class 'SingleCellExperiment'  
BASiCSEstimate(  
  counts,  
  spike.info = NULL,  
  batch = NULL,  
  n = 20000,  
  thin = 10,  
  burn = 5000,  
  regression = TRUE,  
  params = newBASiCSParams(),  
  verbose = TRUE,  
  progress = TRUE,  
  ...  
)  
  
## S3 method for class 'matrix'  
BASiCSEstimate(  
  counts,  
  spike.info = NULL,  
  batch = NULL,  
  n = 20000,  
  thin = 10,  
  burn = 5000,  
  regression = TRUE,  
  params = newBASiCSParams(),  
  verbose = TRUE,  
  progress = TRUE,  
  ...  
)
```

Arguments

counts	either a counts matrix or a SingleCellExperiment object containing count data to estimate parameters from.
spike.info	data.frame describing spike-ins with two columns: "Name" giving the names of the spike-in features (must match rownames(counts)) and "Input" giving the number of input molecules.
batch	vector giving the batch that each cell belongs to.
n	total number of MCMC iterations. Must be $\geq \max(4, \text{thin})$ and a multiple of thin.
thin	thinning period for the MCMC sampler. Must be ≥ 2 .
burn	burn-in period for the MCMC sampler. Must be in the range $1 \leq \text{burn} < n$ and a multiple of thin.
regression	logical. Whether to use regression to identify over-dispersion. See BASiCS_MCMC for details.
params	BASiCSParams object to store estimated values in.
verbose	logical. Whether to print progress messages.
progress	logical. Whether to print additional BASiCS progress messages.
...	Optional parameters passed to BASiCS_MCMC .

Details

This function is just a wrapper around [BASiCS_MCMC](#) that takes the output and converts it to a BASiCSParams object. Either a set of spike-ins or batch information (or both) must be supplied. If only batch information is provided there must be at least two batches. See [BASiCS_MCMC](#) for details.

Value

BASiCSParams object containing the estimated parameters.

Examples

```
# Load example data
library(scuttle)
set.seed(1)
sce <- mockSCE()

spike.info <- data.frame(
  Name = rownames(sce)[1:10],
  Input = rnorm(10, 500, 200),
  stringsAsFactors = FALSE
)
params <- BASiCSEstimate(sce[1:100, 1:30], spike.info)
params
```

BASiCSParams

The BASiCSParams class

Description

S4 class that holds parameters for the BASiCS simulation.

Parameters

The BASiCS simulation uses the following parameters:

`nGenes` The number of genes to simulate.

`nCells` The number of cells to simulate.

[`seed`] Seed to use for generating random numbers.

Batch parameters `nBatches` Number of batches to simulate.

`batchCells` Number of cells in each batch.

Gene parameters `gene.params` A data.frame containing gene parameters with two columns: Mean (mean expression for each biological gene) and Delta (cell-to-cell heterogeneity for each biological gene).

Spike-in parameters `nSpikes` The number of spike-ins to simulate.

`spike.means` Input molecules for each spike-in.

Cell parameters `cell.params` A data.frame containing gene parameters with two columns: Phi (mRNA content factor for each cell, scaled to sum to the number of cells in each batch) and S (capture efficient for each cell).

Variability parameters `theta` Technical variability parameter for each batch.

The parameters not shown in brackets can be estimated from real data using [BASiCSestimate](#). For details of the BASiCS simulation see [BASiCSSimulate](#).

BASiCSSimulate

BASiCS simulation

Description

Simulate counts using the BASiCS method.

Usage

```
BASiCSSimulate(
  params = newBASiCSParams(),
  sparsify = TRUE,
  verbose = TRUE,
  ...
)
```

Arguments

params	BASiCSParams object containing simulation parameters.
sparsify	logical. Whether to automatically convert assays to sparse matrices if there will be a size reduction.
verbose	logical. Whether to print progress messages
...	any additional parameter settings to override what is provided in params.

Details

This function is just a wrapper around `BASiCS_Sim` that takes a `BASiCSParams`, runs the simulation then converts the output to a `SingleCellExperiment` object. See `BASiCS_Sim` for more details of how the simulation works.

Value

`SingleCellExperiment` containing simulated counts

References

Vallejos CA, Marioni JC, Richardson S. BASiCS: Bayesian Analysis of Single-Cell Sequencing data. *PLoS Computational Biology* (2015).

Paper: [10.1371/journal.pcbi.1004333](https://doi.org/10.1371/journal.pcbi.1004333)

Code: <https://github.com/catavallejos/BASiCS>

Examples

```
if (requireNamespace("BASiCS", quietly = TRUE)) {
  sim <- BASiCSSimulate()
}
```

compareSCEs

Compare SingleCellExperiment objects

Description

Combine the data from several `SingleCellExperiment` objects and produce some basic plots comparing them.

Usage

```
compareSCEs(
  sces,
  point.size = 0.1,
  point.alpha = 0.1,
  fits = TRUE,
  colours = NULL
)
```

Arguments

<code>sces</code>	named list of <code>SingleCellExperiment</code> objects to combine and compare.
<code>point.size</code>	size of points in scatter plots.
<code>point.alpha</code>	opacity of points in scatter plots.
<code>fits</code>	whether to include fits in scatter plots.
<code>colours</code>	vector of colours to use for each dataset.

Details

The returned list has three items:

`RowData` Combined row data from the provided `SingleCellExperiments`.

`ColData` Combined column data from the provided `SingleCellExperiments`.

`Plots` Comparison plots

`Means` Boxplot of mean distribution.

`Variances` Boxplot of variance distribution.

`MeanVar` Scatter plot with fitted lines showing the mean-variance relationship.

`LibrarySizes` Boxplot of the library size distribution.

`ZerosGene` Boxplot of the percentage of each gene that is zero.

`ZerosCell` Boxplot of the percentage of each cell that is zero.

`MeanZeros` Scatter plot with fitted lines showing the mean-zeros relationship.

`VarGeneCor` Heatmap of correlation of the 100 most variable genes.

The plots returned by this function are created using [ggplot](#) and are only a sample of the kind of plots you might like to consider. The data used to create these plots is also returned and should be in the correct format to allow you to create further plots using [ggplot](#).

Value

List containing the combined datasets and plots.

Examples

```
sim1 <- splatSimulate(nGenes = 1000, batchCells = 20)
sim2 <- simpleSimulate(nGenes = 1000, nCells = 20)
comparison <- compareSCEs(list(Splat = sim1, Simple = sim2))
names(comparison)
names(comparison$Plots)
```

diffSCEs

*Diff SingleCellExperiment objects***Description**

Combine the data from several SingleCellExperiment objects and produce some basic plots comparing them to a reference.

Usage

```
diffSCEs(
  sces,
  ref,
  point.size = 0.1,
  point.alpha = 0.1,
  fits = TRUE,
  colours = NULL
)
```

Arguments

sces	named list of SingleCellExperiment objects to combine and compare.
ref	string giving the name of the SingleCellExperiment to use as the reference
point.size	size of points in scatter plots.
point.alpha	opacity of points in scatter plots.
fits	whether to include fits in scatter plots.
colours	vector of colours to use for each dataset.

Details

This function aims to look at the differences between a reference SingleCellExperiment and one or more others. It requires each SingleCellExperiment to have the same dimensions. Properties are compared by ranks, for example when comparing the means the values are ordered and the differences between the reference and another dataset plotted. A series of Q-Q plots are also returned.

The returned list has five items:

Reference The SingleCellExperiment used as the reference.

RowData Combined feature data from the provided SingleCellExperiments.

ColData Combined column data from the provided SingleCellExperiments.

Plots Difference plots

Means Boxplot of mean differences.

Variances Boxplot of variance differences.

MeanVar Scatter plot showing the difference from the reference variance across expression ranks.

LibrarySizes Boxplot of the library size differences.
 ZerosGene Boxplot of the differences in the percentage of each gene that is zero.
 ZerosCell Boxplot of the differences in the percentage of each cell that is zero.
 MeanZeros Scatter plot showing the difference from the reference percentage of zeros across expression ranks.
 QQPlots Quantile-Quantile plots
 Means Q-Q plot of the means.
 Variances Q-Q plot of the variances.
 LibrarySizes Q-Q plot of the library sizes.
 ZerosGene Q-Q plot of the percentage of zeros per gene.
 ZerosCell Q-Q plot of the percentage of zeros per cell.

The plots returned by this function are created using [ggplot](#) and are only a sample of the kind of plots you might like to consider. The data used to create these plots is also returned and should be in the correct format to allow you to create further plots using [ggplot](#).

Value

List containing the combined datasets and plots.

Examples

```

sim1 <- splatSimulate(nGenes = 1000, batchCells = 20)
sim2 <- simpleSimulate(nGenes = 1000, nCells = 20)
difference <- diffSCEs(list(Splat = sim1, Simple = sim2), ref = "Simple")
names(difference)
names(difference$Plots)

```

getParam

Get a parameter

Description

Accessor function for getting parameter values.

Usage

```

getParam(object, name)

## S4 method for signature 'Params'
getParam(object, name)

```

Arguments

object object to get parameter from.
 name name of the parameter to get.

Value

The extracted parameter value

Examples

```
params <- newSimpleParams()
getParam(params, "nGenes")
```

getParam

Get parameters

Description

Get multiple parameter values from a Params object.

Usage

```
getParam(params, names)
```

Arguments

params	Params object to get values from.
names	vector of names of the parameters to get.

Value

List with the values of the selected parameters.

Examples

```
params <- newSimpleParams()
getParam(params, c("nGenes", "nCells", "mean.rate"))
```

kersplatEstimate

Estimate Kersplat simulation parameters

Description

Estimate simulation parameters for the Kersplat simulation from a real dataset. See the individual estimation functions for more details on how this is done.

Usage

```
kersplatEstimate(counts, params = newKersplatParams(), verbose = TRUE)

## S3 method for class 'SingleCellExperiment'
kersplatEstimate(counts, params = newKersplatParams(), verbose = TRUE)

## S3 method for class 'matrix'
kersplatEstimate(counts, params = newKersplatParams(), verbose = TRUE)
```

Arguments

counts	either a counts matrix or a SingleCellExperiment object containing count data to estimate parameters from.
params	KersplatParams object to store estimated values in.
verbose	logical. Whether to print progress messages.

Value

KersplatParams object containing the estimated parameters.

See Also

[kersplatEstMean](#), [kersplatEstBCV](#), [kersplatEstLib](#)

Examples

```
if (requireNamespace("igraph", quietly = TRUE)) {
  # Load example data
  library(scuttle)
  set.seed(1)
  sce <- mockSCE()

  params <- kersplatEstimate(sce)
  params
}
```

KersplatParams

The KersplatParams class

Description

S4 class that holds parameters for the Kersplat simulation.

Parameters

The Kersplat simulation uses the following parameters:

`nGenes` The number of genes to simulate.

`nCells` The number of cells to simulate.

[`seed`] Seed to use for generating random numbers.

Mean parameters `mean.shape` Shape parameter for the mean gamma distribution.

`mean.rate` Rate parameter for the mean gamma distribution.

`mean.outProb` Probability that a gene is an expression outlier.

`mean.outFacLoc` Location (meanlog) parameter for the expression outlier factor log-normal distribution.

`mean.outFacScale` Scale (sdlog) parameter for the expression outlier factor log-normal distribution.

`mean.dens` [density](#) object describing the log gene mean density.

[`mean.method`] Method to use for simulating gene means. Either "fit" to sample from a gamma distribution (with expression outliers) or "density" to sample from the provided density object.

[`mean.values`] Vector of means for each gene.

Biological Coefficient of Variation parameters `bcv.common` Underlying common dispersion across all genes.

[`bcv.df`] Degrees of Freedom for the BCV inverse chi-squared distribution.

Network parameters [`network.graph`] Graph containing the gene network.

[`network.nRegs`] Number of regulators in the network.

Paths parameters [`paths.programs`] Number of expression programs.

[`paths.design`] `data.frame` describing path structure. See [kersplatSimPaths](#) for details.

Library size parameters `lib.loc` Location (meanlog) parameter for the library size log-normal distribution, or mean parameter if a normal distribution is used.

`lib.scale` Scale (sdlog) parameter for the library size log-normal distribution, or sd parameter if a normal distribution is used.

`lib.dens` [density](#) object describing the library size density.

[`lib.method`] Method to use for simulating library sizes. Either "fit" to sample from a log-normal distribution or "density" to sample from the provided density object.

Design parameters [`cells.design`] `data.frame` describing cell structure. See [kersplatSimCellMeans](#) for details.

Doublet parameters [`doublet.prop`] Proportion of cells that are doublets.

Ambient parameters [`ambient.scale`] Scaling factor for the library size log-normal distribution when generating ambient library sizes.

[`ambient.nEmpty`] Number of empty cells to simulate.

The parameters not shown in brackets can be estimated from real data using [kersplatEstimate](#). For details of the Kersplat simulation see [kersplatSimulate](#).

kersplatSample	<i>Kersplat sample</i>
----------------	------------------------

Description

Sample cells for the Kersplat simulation

Usage

```
kersplatSample(params, sparsify = TRUE, verbose = TRUE)
```

Arguments

params	KersplatParams object containing simulation parameters.
sparsify	logical. Whether to automatically convert assays to sparse matrices if there will be a size reduction.
verbose	logical. Whether to print progress messages

Details

The second stage is a two-step Kersplat simulation is to generate cells based on a complete [KersplatParams](#) object. intermediate parameters.

The sampling process involves the following steps:

1. Simulate library sizes for each cell
2. Simulate means for each cell
3. Simulate endogenous counts for each cell
4. Simulate ambient counts for each cell
5. Simulate final counts for each cell

The final output is a [SingleCellExperiment](#) object that contains the simulated counts but also the values for various intermediate steps. These are stored in the [colData](#) (for cell specific information), [rowData](#) (for gene specific information) or [assays](#) (for gene by cell matrices) slots. This additional information includes:

`colData` **Cell** Unique cell identifier.

Type Whether the cell is a Cell, Doublet or Empty.

CellLibSize The expected number of endogenous counts for that cell.

AmbientLibSize The expected number of ambient counts for that cell.

Path The path the cell belongs to.

Step How far along the path each cell is.

Path1 For doublets the path of the first partner in the doublet (otherwise NA).

Step1 For doublets the step of the first partner in the doublet (otherwise NA).

Path2 For doublets the path of the second partner in the doublet (otherwise NA).

Step2 For doublets the step of the second partner in the doublet (otherwise NA).

rowData **Gene** Unique gene identifier.

BaseMean The base expression level for that gene.

AmbientMean The ambient expression level for that gene.

assays **CellMeans** The mean expression of genes in each cell after any differential expression and adjusted for expected library size.

CellCounts Endogenous count matrix.

AmbientCounts Ambient count matrix.

counts Final count matrix.

Values that have been added by Splatter are named using UpperCamelCase in order to differentiate them from the values added by analysis packages which typically use underscore_naming.

Value

SingleCellExperiment object containing the simulated counts and intermediate values.

See Also

[kersplatSimLibSizes](#), [kersplatSimCellMeans](#), [kersplatSimCellCounts](#), [kersplatSimAmbientCounts](#), [kersplatSimCounts](#)

Examples

```
if (requireNamespace("igraph", quietly = TRUE)) {
  params <- kersplatSetup()
  sim <- kersplatSample(params)
}
```

kersplatSetup

Kersplat setup

Description

Setup the parameters required for the Kersplat simulation

Usage

```
kersplatSetup(params = newKersplatParams(), verbose = TRUE, ...)
```

Arguments

params KersplatParams object containing simulation parameters.

verbose logical. Whether to print progress messages

... any additional parameter settings to override what is provided in params.

Details

The first stage is a two-step Kersplat simulation is to generate some of the intermediate parameters. The resulting parameters allow multiple simulated datasets to be generated from the same biological structure (using [kersplatSample](#)). As with all the other parameters these values can be manually overwritten if desired.

The setup involves the following steps:

1. Generate a gene network (if not already present)
2. Select regulator genes (if not already present)
3. Simulate gene means (if not already present)
4. Simulate cell paths

The resulting [KersplatParams](#) object will have the following parameters set (if they weren't already).

- `mean.values`
- `network.graph`
- `network.regsSet`
- `paths.means`

See [KersplatParams](#) for more details about these parameters and the functions for the individual steps for more details about the process.

Value

A complete [KersplatParams](#) object

See Also

[kersplatGenNetwork](#), [kersplatSelectRegs](#), [kersplatSimGeneMeans](#), [kersplatSimPaths](#), [KersplatParams](#)

Examples

```
if (requireNamespace("igraph", quietly = TRUE)) {  
  params <- kersplatSetup()  
}
```

kersplatSimulate *Kersplat simulation*

Description

Simulate scRNA-seq count data using the Kersplat model

Usage

```
kersplatSimulate(  
  params = newKersplatParams(),  
  sparsify = TRUE,  
  verbose = TRUE,  
  ...  
)
```

Arguments

params	KersplatParams object containing simulation parameters.
sparsify	logical. Whether to automatically convert assays to sparse matrices if there will be a size reduction.
verbose	logical. Whether to print progress messages
...	any additional parameter settings to override what is provided in params.

Details

This functions is for simulating data in a single step. It consists of a call to [kersplatSetup](#) followed by a call to [kersplatSample](#). Please see the documentation for those functions for more details of the individual steps.

Value

SingleCellExperiment containing simulated counts and intermediate values

See Also

[kersplatSetup](#), [kersplatSample](#)

Examples

```
if (requireNamespace("igraph", quietly = TRUE)) {  
  sim <- kersplatSimulate  
}
```

listSims	<i>List simulations</i>
----------	-------------------------

Description

List all the simulations that are currently available in Splatter with a brief description.

Usage

```
listSims(print = TRUE)
```

Arguments

print logical. Whether to print to the console.

Value

Invisibly returns a data.frame containing the information that is displayed.

Examples

```
listSims()
```

lun2Estimate	<i>Estimate Lun2 simulation parameters</i>
--------------	--

Description

Estimate simulation parameters for the Lun2 simulation from a real dataset.

Usage

```
lun2Estimate(  
  counts,  
  plates,  
  params = newLun2Params(),  
  min.size = 200,  
  verbose = TRUE,  
  BPPARAM = SerialParam()  
)  
  
## S3 method for class 'SingleCellExperiment'  
lun2Estimate(  
  counts,  
  plates,
```

```

    params = newLun2Params(),
    min.size = 200,
    verbose = TRUE,
    BPPARAM = SerialParam()
  )

## S3 method for class 'matrix'
lun2Estimate(
  counts,
  plates,
  params = newLun2Params(),
  min.size = 200,
  verbose = TRUE,
  BPPARAM = SerialParam()
)

```

Arguments

counts	either a counts matrix or a SingleCellExperiment object containing count data to estimate parameters from.
plates	integer vector giving the plate that each cell originated from.
params	Lun2Params object to store estimated values in.
min.size	minimum size of clusters when identifying group of cells in the data.
verbose	logical. Whether to show progress messages.
BPPARAM	A BiocParallelParam instance giving the parallel back-end to be used. Default is SerialParam which uses a single core.

Details

See [Lun2Params](#) for more details on the parameters.

Value

LunParams object containing the estimated parameters.

Examples

```

# Load example data
library(scuttle)
set.seed(1)
sce <- mockSCE()

plates <- as.numeric(factor(colData(sce)$Mutation_Status))
params <- lun2Estimate(sce, plates, min.size = 20)
params

```

Lun2Params

*The Lun2Params class***Description**

S4 class that holds parameters for the Lun2 simulation.

Parameters

The Lun2 simulation uses the following parameters:

`nGenes` The number of genes to simulate.

`nCells` The number of cells to simulate.

[`seed`] Seed to use for generating random numbers.

Gene parameters `gene.params` A data.frame containing gene parameters with two columns: Mean (mean expression for each gene) and Disp (dispersion for each gene).

`zi.params` A data.frame containing zero-inflated gene parameters with three columns: Mean (mean expression for each gene), Disp (dispersion for each, gene), and Prop (zero proportion for each gene).

[`nPlates`] The number of plates to simulate.

Plate parameters `plate.ingroup` Character vector giving the plates considered to be part of the "ingroup".

`plate.mod` Plate effect modifier factor. The plate effect variance is divided by this value.

`plate.var` Plate effect variance.

Cell parameters `cell.plates` Factor giving the plate that each cell comes from.

`cell.libSizes` Library size for each cell.

`cell.libMod` Modifier factor for library sizes. The library sizes are multiplied by this value.

Differential expression parameters `de.nGenes` Number of differentially expressed genes.

`de.fc` Fold change for differentially expressed genes.

The parameters not shown in brackets can be estimated from real data using [lun2Estimate](#). For details of the Lun2 simulation see [lun2Simulate](#).

lun2Simulate

*Lun2 simulation***Description**

Simulate single-cell RNA-seq count data using the method described in Lun and Marioni "Overcoming confounding plate effects in differential expression analyses of single-cell RNA-seq data".

Usage

```
lun2Simulate(  
  params = newLun2Params(),  
  zinb = FALSE,  
  sparsify = TRUE,  
  verbose = TRUE,  
  ...  
)
```

Arguments

params	Lun2Params object containing simulation parameters.
zinb	logical. Whether to use a zero-inflated model.
sparsify	logical. Whether to automatically convert assays to sparse matrices if there will be a size reduction.
verbose	logical. Whether to print progress messages.
...	any additional parameter settings to override what is provided in params.

Details

The Lun2 simulation uses a negative-binomial distribution where the means and dispersions have been sampled from a real dataset (using `lun2Estimate`). The other core feature of the Lun2 simulation is the addition of plate effects. Differential expression can be added between two groups of plates (an "ingroup" and all other plates). Library size factors are also applied and optionally a zero-inflated negative-binomial can be used.

If the number of genes to simulate differs from the number of provided gene parameters or the number of cells to simulate differs from the number of library sizes the relevant parameters will be sampled with a warning. This allows any number of genes or cells to be simulated regardless of the number in the dataset used in the estimation step but has the downside that some genes or cells may be simulated multiple times.

Value

SingleCellExperiment containing simulated counts.

References

Lun ATL, Marioni JC. Overcoming confounding plate effects in differential expression analyses of single-cell RNA-seq data. *Biostatistics* (2017).

Paper: [dx.doi.org/10.1093/biostatistics/kxw055](https://doi.org/10.1093/biostatistics/kxw055)

Code: <https://github.com/MarioniLab/PlateEffects2016>

Examples

```
sim <- lun2Simulate()
```

lunEstimate	<i>Estimate Lun simulation parameters</i>
-------------	---

Description

Estimate simulation parameters for the Lun simulation from a real dataset.

Usage

```
lunEstimate(counts, params = newLunParams())

## S3 method for class 'SingleCellExperiment'
lunEstimate(counts, params = newLunParams())

## S3 method for class 'matrix'
lunEstimate(counts, params = newLunParams())
```

Arguments

counts	either a counts matrix or an SingleCellExperiment object containing count data to estimate parameters from.
params	LunParams object to store estimated values in.

Details

The nGenes and nCells parameters are taken from the size of the input data. No other parameters are estimated. See [LunParams](#) for more details on the parameters.

Value

LunParams object containing the estimated parameters.

Examples

```
# Load example data
library(scuttle)
set.seed(1)
sce <- mockSCE()

params <- lunEstimate(sce)
params
```

LunParams

The LunParams class

Description

S4 class that holds parameters for the Lun simulation.

Parameters

The Lun simulation uses the following parameters:

nGenes The number of genes to simulate.

nCells The number of cells to simulate.

[nGroups] The number of groups to simulate.

[groupCells] Vector giving the number of cells in each simulation group/path.

[seed] Seed to use for generating random numbers.

Mean parameters [mean.shape] Shape parameter for the mean gamma distribution.

[mean.rate] Rate parameter for the mean gamma distribution.

Counts parameters [count.disp] The dispersion parameter for the counts negative binomial distribution.

Differential expression parameters [de.nGenes] The number of genes that are differentially expressed in each group

[de.upProp] The proportion of differentially expressed genes that are up-regulated in each group

[de.upFC] The fold change for up-regulated genes

[de.downFC] The fold change for down-regulated genes

The parameters not shown in brackets can be estimated from real data using [lunEstimate](#). For details of the Lun simulation see [lunSimulate](#).

lunSimulate

Lun simulation

Description

Simulate single-cell RNA-seq count data using the method described in Lun, Bach and Marioni "Pooling across cells to normalize single-cell RNA sequencing data with many zero counts".

Usage

```
lunSimulate(params = newLunParams(), sparsify = TRUE, verbose = TRUE, ...)
```

Arguments

params	LunParams object containing Lun simulation parameters.
sparsify	logical. Whether to automatically convert assays to sparse matrices if there will be a size reduction.
verbose	logical. Whether to print progress messages.
...	any additional parameter settings to override what is provided in params.

Details

The Lun simulation generates gene mean expression levels from a gamma distribution with shape = mean.shape and rate = mean.rate. Counts are then simulated from a negative binomial distribution with mu = means and size = 1 / bcv.common. In addition each cell is given a size factor ($2^{\text{rnorm}(n\text{Cells}, \text{mean} = 0, \text{sd} = 0.5)}$) and differential expression can be simulated with fixed fold changes.

See [LunParams](#) for details of the parameters.

Value

SingleCellExperiment object containing the simulated counts and intermediate values.

References

Lun ATL, Bach K, Marioni JC. Pooling across cells to normalize single-cell RNA sequencing data with many zero counts. *Genome Biology* (2016).

Paper: [dx.doi.org/10.1186/s13059-016-0947-7](https://doi.org/10.1186/s13059-016-0947-7)

Code: <https://github.com/MarioniLab/Deconvolution2016>

Examples

```
sim <- lunSimulate()
```

makeCompPanel	<i>Make comparison panel</i>
---------------	------------------------------

Description

Combine the plots from compareSCEs into a single panel.

Usage

```
makeCompPanel(
  comp,
  title = "Comparison",
  labels = c("Means", "Variance", "Mean-variance relationship", "Library size",
            "Zeros per gene", "Zeros per cell", "Mean-zeros relationship")
)
```

Arguments

comp list returned by [compareSCEs](#).
title title for the panel.
labels vector of labels for each of the seven plots.

Value

Combined panel plot

Examples

```
sim1 <- splatSimulate(nGenes = 1000, batchCells = 20)
sim2 <- simpleSimulate(nGenes = 1000, nCells = 20)
comparison <- compareSCEs(list(Splat = sim1, Simple = sim2))
panel <- makeCompPanel(comparison)
```

makeDiffPanel

Make difference panel

Description

Combine the plots from diffSCEs into a single panel.

Usage

```
makeDiffPanel(  
  diff,  
  title = "Difference comparison",  
  labels = c("Means", "Variance", "Library size", "Zeros per cell", "Zeros per gene",  
            "Mean-variance relationship", "Mean-zeros relationship")  
)
```

Arguments

diff list returned by [diffSCEs](#).
title title for the panel.
labels vector of labels for each of the seven sections.

Value

Combined panel plot

Examples

```
sim1 <- splatSimulate(nGenes = 1000, batchCells = 20)
sim2 <- simpleSimulate(nGenes = 1000, nCells = 20)
difference <- diffSCEs(list(Splat = sim1, Simple = sim2), ref = "Simple")
panel <- makeDiffPanel(difference)
```

makeOverallPanel	<i>Make overall panel</i>
------------------	---------------------------

Description

Combine the plots from compSCEs and diffSCEs into a single panel.

Usage

```
makeOverallPanel(
  comp,
  diff,
  title = "Overall comparison",
  row.labels = c("Means", "Variance", "Mean-variance relationship", "Library size",
    "Zeros per cell", "Zeros per gene", "Mean-zeros relationship")
)
```

Arguments

comp	list returned by compareSCEs .
diff	list returned by diffSCEs .
title	title for the panel.
row.labels	vector of labels for each of the seven rows.

Value

Combined panel plot

Examples

```
sim1 <- splatSimulate(nGenes = 1000, batchCells = 20)
sim2 <- simpleSimulate(nGenes = 1000, nCells = 20)
comparison <- compareSCEs(list(Splat = sim1, Simple = sim2))
difference <- diffSCEs(list(Splat = sim1, Simple = sim2), ref = "Simple")
panel <- makeOverallPanel(comparison, difference)
```

`mfaEstimate`*Estimate mfa simulation parameters*

Description

Estimate simulation parameters for the mfa simulation from a real dataset.

Usage

```
mfaEstimate(counts, params = newMFAParams())

## S3 method for class 'SingleCellExperiment'
mfaEstimate(counts, params = newMFAParams())

## S3 method for class 'matrix'
mfaEstimate(counts, params = newMFAParams())
```

Arguments

<code>counts</code>	either a counts matrix or a <code>SingleCellExperiment</code> object containing count data to estimate parameters from.
<code>params</code>	<code>MFAParams</code> object to store estimated values in.

Details

The `nGenes` and `nCells` parameters are taken from the size of the input data. The dropout lambda parameter is estimate using [empirical_lambda](#). See [MFAParams](#) for more details on the parameters.

Value

`MFAParams` object containing the estimated parameters.

Examples

```
# Load example data
if (requireNamespace("mfa", quietly = TRUE)) {
  library(mfa)
  synth <- create_synthetic(
    C = 20, G = 5, zero_negative = TRUE,
    model_dropout = TRUE
  )

  params <- mfaEstimate(synth$X)
  params
}
```

MFAParams

*The MFAParams class***Description**

S4 class that holds parameters for the mfa simulation.

Parameters

The mfa simulation uses the following parameters:

nGenes The number of genes to simulate.

nCells The number of cells to simulate.

[seed] Seed to use for generating random numbers.

[trans.prop] Proportion of genes that show transient expression. These genes are briefly up or down-regulated before returning to their initial state

[zero.neg] Logical. Whether to set negative expression values to zero. This will zero-inflate the data.

[dropout.present] Logical. Whether to simulate dropout.

dropout.lambda Lambda parameter for the exponential dropout function.

The parameters not shown in brackets can be estimated from real data using [mfaEstimate](#). See [create_synthetic](#) for more details about the parameters. For details of the Splatter implementation of the mfa simulation see [mfaSimulate](#).

mfaSimulate

*MFA simulation***Description**

Simulate a bifurcating pseudotime path using the mfa method.

Usage

```
mfaSimulate(params = newMFAParams(), sparsify = TRUE, verbose = TRUE, ...)
```

Arguments

params MFAParams object containing simulation parameters.

sparsify logical. Whether to automatically convert assays to sparse matrices if there will be a size reduction.

verbose Logical. Whether to print progress messages.

... any additional parameter settings to override what is provided in params.

Details

This function is just a wrapper around `create_synthetic` that takes a `MFAParams`, runs the simulation then converts the output from log-expression to counts and returns a `SingleCellExperiment` object. See `create_synthetic` and the mfa paper for more details about how the simulation works.

Value

`SingleCellExperiment` containing simulated counts

References

Campbell KR, Yau C. Probabilistic modeling of bifurcations in single-cell gene expression data using a Bayesian mixture of factor analyzers. Wellcome Open Research (2017).

Paper: [10.12688/wellcomeopenres.11087.1](https://doi.org/10.12688/wellcomeopenres.11087.1)

Code: <https://github.com/kieranrcampbell/mfa>

Examples

```
if (requireNamespace("mfa", quietly = TRUE)) {  
  sim <- mfaSimulate()  
}
```

minimiseSCE

Minimise SCE

Description

Reduce the size of a `SingleCellExperiment` object by unneeded information.

Usage

```
minimiseSCE(  
  sce,  
  rowData.keep = FALSE,  
  colData.keep = FALSE,  
  metadata.keep = FALSE,  
  assays.keep = "counts",  
  sparsify = c("auto", "all", "none"),  
  verbose = TRUE  
)
```

Arguments

sce	SingleCellExperiment object
rowData.keep	Either TRUE (keep all rowData columns), FALSE (remove all rowData columns) or a character vector with the names of the rowData columns to keep
colData.keep	Either TRUE (keep all colData columns), FALSE (remove all colData columns) or a character vector with the names of the colData columns to keep
metadata.keep	Either TRUE (keep all metadata), FALSE (remove all metadata) or a character vector with the names of the metadata items to keep
assays.keep	Either TRUE (keep all assays), FALSE (remove all assays) or a character vector with the names of the assays to keep
sparsify	Whether to convert assay matrices to sparse format. Either "all", "none" or "auto" (default) to only convert those matrices that will result in a size reduction
verbose	Whether to print status messages

Value

SingleCellExperiment object

Examples

```
sce <- splatSimulate(verbose = FALSE)
sce.min <- minimiseSCE(sce, verbose = FALSE)
object.size(sce)
object.size(sce.min)
```

mockBulkeQTL

Generate mock eQTL mapping results

Description

Quick function to generate mock eQTL mapping results, with parameters estimated using real eQTL mapping results from GTEx using thyroid tissue.

Usage

```
mockBulkeQTL(n.genes = 500, seed = NULL)
```

Arguments

n.genes	Number of genes in mock eQTL data.
seed	Optional: seed for random seed

Value

data.frame containing mock bulk eQTL mapping results.

Examples

```
eqtl <- mockBulkeQTL()
```

mockBulkMatrix	<i>Generate mock bulk population scale expression data</i>
----------------	--

Description

Quick function to generate mock bulk expression data for a population, with parameters estimated using real thyroid tissue data from GTEx.

Usage

```
mockBulkMatrix(n.genes = 100, n.samples = 50, seed = NULL)
```

Arguments

n.genes	Number of genes in mock bulk data.
n.samples	Number of samples in mock bulk data.
seed	Optional: seed for random seed

Value

matrix containing mock bulk expression data.

Examples

```
bulk <- mockBulkMatrix
```

mockEmpiricalSet	<i>Generate set of "empirical" mock data</i>
------------------	--

Description

Quick function to generate matching mock VCF, bulk expression, and eQTL data, useful for running `splatPopEmpiricalMeans`

Usage

```
mockEmpiricalSet(
  n.genes = 20,
  n.snps = 1000,
  n.samples = 10,
  chromosome = 1,
  chr.length = 2e+06,
  seed = NULL
)
```

Arguments

n.genes	Number of genes in mock eQTL data.
n.snps	Number of SNPs in mock vcf file.
n.samples	Number of samples in mock bulk data.
chromosome	Chromosome name
chr.length	Length of mock chromosome
seed	Optional: seed for random seed

Value

list(gff=mockGFF, vcf=mockVCF, means=mockMEANS, eqtl=mockEQTL)

Examples

```
empirical <- mockEmpiricalSet()
```

mockGFF

Generate mock gff

Description

Quick function to generate a mock gff.

Usage

```
mockGFF(n.genes = 50, chromosome = 1, chr.length = 2e+06, seed = NULL)
```

Arguments

n.genes	Number of genes in mock gff file
chromosome	Chromosome name
chr.length	Length of mock chromosome
seed	Optional: seed for random seed

Value

data.frame containing mock gff data.

Examples

```
gff <- mockGFF()
```

mockVCF

Generate mock vcf

Description

Quick function to generate mock vcf file. Note this data has unrealistic population structure.

Usage

```
mockVCF(  
  n.snps = 200,  
  n.samples = 5,  
  chromosome = 1,  
  chr.length = 2e+06,  
  seed = NULL  
)
```

Arguments

n.snps	Number of SNPs in mock vcf file.
n.samples	Number of samples in mock bulk data.
chromosome	Chromosome name
chr.length	Length of mock chromosome
seed	Optional: seed for random seed

Value

data.frame containing mock vcf data.

Examples

```
vcf <- mockVCF()
```

`newParams`*New Params*

Description

Create a new Params object. Functions exist for each of the different Params subtypes.

Usage`newBASiCParams(...)``newKersplatParams(...)``newLun2Params(...)``newLunParams(...)``newMFAParams(...)``newPhenoParams(...)``newSCDDParams(...)``newSimpleParams(...)``newSparseDCParams(...)``newSplatParams(...)``newSplatPopParams(...)``newZINBParams(...)`**Arguments**

... additional parameters passed to [setParams](#).

Value

New Params object.

Examples

```
params <- newSimpleParams()  
params <- newSimpleParams(nGenes = 200, nCells = 10)
```

Params	<i>The Params virtual class</i>
--------	---------------------------------

Description

Virtual S4 class that all other Params classes inherit from.

Parameters

The Params class defines the following parameters:

nGenes The number of genes to simulate.

nCells The number of cells to simulate.

[seed] Seed to use for generating random numbers.

The parameters not shown in brackets can be estimated from real data.

phenoEstimate	<i>Estimate PhenoPath simulation parameters</i>
---------------	---

Description

Estimate simulation parameters for the PhenoPath simulation from a real dataset.

Usage

```
phenoEstimate(counts, params = newPhenoParams())
```

```
## S3 method for class 'SingleCellExperiment'  
phenoEstimate(counts, params = newPhenoParams())
```

```
## S3 method for class 'matrix'  
phenoEstimate(counts, params = newPhenoParams())
```

Arguments

counts either a counts matrix or an SingleCellExperiment object containing count data to estimate parameters from.

params PhenoParams object to store estimated values in.

Details

The nGenes and nCells parameters are taken from the size of the input data. The total number of genes is evenly divided into the four types. See [PhenoParams](#) for more details on the parameters.

Value

PhenoParams object containing the estimated parameters.

Examples

```
if (requireNamespace("phenopath", quietly = TRUE)) {  
  # Load example data  
  library(scuttle)  
  set.seed(1)  
  sce <- mockSCE()  
  
  params <- phenoEstimate(sce)  
  params  
}
```

PhenoParams

The PhenoParams class

Description

S4 class that holds parameters for the PhenoPath simulation.

Parameters

The PhenoPath simulation uses the following parameters:

`nGenes` The number of genes to simulate.

`nCells` The number of cells to simulate.

`[seed]` Seed to use for generating random numbers.

`[n.de]` Number of genes to simulate from the differential expression regime

`[n.pst]` Number of genes to simulate from the pseudotime regime

`[n.pst.beta]` Number of genes to simulate from the pseudotime + beta interactions regime

`[n.de.pst.beta]` Number of genes to simulate from the differential expression + pseudotime + interactions regime

The parameters not shown in brackets can be estimated from real data using [phenoEstimate](#). For details of the PhenoPath simulation see [phenoSimulate](#).

phenoSimulate	<i>PhenoPath simulation</i>
---------------	-----------------------------

Description

Simulate counts from a pseudotime trajectory using the PhenoPath method.

Usage

```
phenoSimulate(params = newPhenoParams(), sparsify = TRUE, verbose = TRUE, ...)
```

Arguments

params	PhenoParams object containing simulation parameters.
sparsify	logical. Whether to automatically convert assays to sparse matrices if there will be a size reduction.
verbose	logical. Whether to print progress messages
...	any additional parameter settings to override what is provided in params.

Details

This function is just a wrapper around [simulate_phenopath](#) that takes a [PhenoParams](#), runs the simulation then converts the output from log-expression to counts and returns a [SingleCellExperiment](#) object. The original simulated log-expression values are returned in the `LogExprs` assay. See [simulate_phenopath](#) and the PhenoPath paper for more details about how the simulation works.

Value

SingleCellExperiment containing simulated counts

References

Campbell K, Yau C. Uncovering genomic trajectories with heterogeneous genetic and environmental backgrounds across single-cells and populations. *bioRxiv* (2017).

Paper: [10.1101/159913](https://doi.org/10.1101/159913)

Code: <https://github.com/kieranrcampbell/phenopath>

Examples

```
if (requireNamespace("phenopath", quietly = TRUE)) {  
  sim <- phenoSimulate()  
}
```

`scDDEstimate`*Estimate scDD simulation parameters*

Description

Estimate simulation parameters for the scDD simulation from a real dataset.

Usage

```
scDDEstimate(  
  counts,  
  params = newSCDDParams(),  
  verbose = TRUE,  
  BPPARAM = SerialParam(),  
  ...  
)  
  
## S3 method for class 'matrix'  
scDDEstimate(  
  counts,  
  params = newSCDDParams(),  
  verbose = TRUE,  
  BPPARAM = SerialParam(),  
  conditions,  
  ...  
)  
  
## S3 method for class 'SingleCellExperiment'  
scDDEstimate(  
  counts,  
  params = newSCDDParams(),  
  verbose = TRUE,  
  BPPARAM = SerialParam(),  
  condition = "condition",  
  ...  
)  
  
## Default S3 method:  
scDDEstimate(  
  counts,  
  params = newSCDDParams(),  
  verbose = TRUE,  
  BPPARAM = SerialParam(),  
  condition,  
  ...  
)
```

Arguments

counts	either a counts matrix or a SingleCellExperiment object containing count data to estimate parameters from.
params	SCDDParams object to store estimated values in.
verbose	logical. Whether to show progress messages.
BPPARAM	A BiocParallelParam instance giving the parallel back-end to be used. Default is SerialParam which uses a single core.
...	further arguments passed to or from other methods.
conditions	Vector giving the condition that each cell belongs to. Conditions can be 1 or 2.
condition	String giving the column that represents biological group of interest.

Details

This function applies [preprocess](#) to the counts then uses [scDD](#) to estimate the numbers of each gene type to simulate. The output is then converted to a SCDDParams object. See [preprocess](#) and [scDD](#) for details.

Value

SCDDParams object containing the estimated parameters.

Examples

```
if (requireNamespace("scDD", quietly = TRUE)) {
  library(scuttle)
  set.seed(1)
  sce <- mockSCE(ncells = 20, ngenes = 100)

  colData(sce)$condition <- sample(1:2, ncol(sce), replace = TRUE)
  params <- scDDEstimate(sce, condition = "condition")
  params
}
```

 SCDDParams

The SCDDParams class

Description

S4 class that holds parameters for the scDD simulation.

Parameters

The SCDD simulation uses the following parameters:

nGenes The number of genes to simulate (not used).

nCells The number of cells to simulate in each condition.

[seed] Seed to use for generating random numbers.

SCdat [SingleCellExperiment](#) containing real data.

nDE Number of DE genes to simulate.

nDP Number of DP genes to simulate.

nDM Number of DM genes to simulate.

nDB Number of DB genes to simulate.

nEE Number of EE genes to simulate.

nEP Number of EP genes to simulate.

[sd.range] Interval for fold change standard deviations.

[modeFC] Values for DP, DM and DB mode fold changes.

[varInflation] Variance inflation factors for each condition. If all equal to 1 will be set to NULL (default).

[condition] String giving the column that represents biological group of interest.

The parameters not shown in brackets can be estimated from real data using [scDDEstimate](#). See [simulateSet](#) for more details about the parameters. For details of the Splatter implementation of the scDD simulation see [scDDSimulate](#).

scDDSimulate

scDD simulation

Description

Simulate counts using the scDD method.

Usage

```
scDDSimulate(
  params = newSCDDParams(),
  plots = FALSE,
  plot.file = NULL,
  sparsify = TRUE,
  verbose = TRUE,
  BPPARAM = SerialParam(),
  ...
)
```

Arguments

params	SCDDParams object containing simulation parameters.
plots	logical. whether to generate scDD fold change and validation plots.
plot.file	File path to save plots as PDF.
sparsify	logical. Whether to automatically convert assays to sparse matrices if there will be a size reduction.
verbose	logical. Whether to print progress messages
BPPARAM	A BiocParallelParam instance giving the parallel back-end to be used. Default is SerialParam which uses a single core.
...	any additional parameter settings to override what is provided in params.

Details

This function is just a wrapper around [simulateSet](#) that takes a [SCDDParams](#), runs the simulation then converts the output to a [SingleCellExperiment](#) object. See [simulateSet](#) for more details about how the simulation works.

Value

SingleCellExperiment containing simulated counts

References

Korthauer KD, Chu L-F, Newton MA, Li Y, Thomson J, Stewart R, et al. A statistical approach for identifying differential distributions in single-cell RNA-seq experiments. *Genome Biology* (2016).

Paper: [10.1186/s13059-016-1077-y](https://doi.org/10.1186/s13059-016-1077-y)

Code: <https://github.com/kdkorthauer/scDD>

Examples

```
sim <- scDDSimulate()
```

setParam

Set a parameter

Description

Function for setting parameter values.

Usage

```
setParam(object, name, value)

## S4 method for signature 'BASiCParams'
setParam(object, name, value)

## S4 method for signature 'KersplatParams'
setParam(object, name, value)

## S4 method for signature 'Lun2Params'
setParam(object, name, value)

## S4 method for signature 'LunParams'
setParam(object, name, value)

## S4 method for signature 'Params'
setParam(object, name, value)

## S4 method for signature 'PhenoParams'
setParam(object, name, value)

## S4 method for signature 'SCDDParams'
setParam(object, name, value)

## S4 method for signature 'SplatParams'
setParam(object, name, value)

## S4 method for signature 'SplatPopParams'
setParam(object, name, value)

## S4 method for signature 'ZINBParams'
setParam(object, name, value)
```

Arguments

object	object to set parameter in.
name	name of the parameter to set.
value	value to set the parameter to.

Value

Object with new parameter value.

Examples

```
params <- newSimpleParams()
setParam(params, "nGenes", 100)
```

setParams	<i>Set parameters</i>
-----------	-----------------------

Description

Set multiple parameters in a Params object.

Usage

```
setParams(object, update = NULL, ...)  
  
## S4 method for signature 'KersplatParams'  
setParams(object, update = NULL, ...)  
  
## S4 method for signature 'Params'  
setParams(object, update = NULL, ...)  
  
## S4 method for signature 'SplatParams'  
setParams(object, update = NULL, ...)
```

Arguments

object	Params object to set parameters in.
update	list of parameters to set where names(update) are the names of the parameters to set and the items in the list are values.
...	additional parameters to set. These are combined with any parameters specified in update.

Details

Each parameter is set by a call to [setParameter](#). If the same parameter is specified multiple times it will be set multiple times. Parameters can be specified using a list via update (useful when collecting parameter values in some way) or individually (useful when setting them manually), see examples.

Value

Params object with updated values.

Examples

```
params <- newSimpleParams()  
params  
# Set individually  
params <- setParams(params, nGenes = 1000, nCells = 50)  
params  
# Set via update list  
params <- setParams(params, list(mean.rate = 0.2, mean.shape = 0.8))  
params
```

simpleEstimate	<i>Estimate simple simulation parameters</i>
----------------	--

Description

Estimate simulation parameters for the simple simulation from a real dataset.

Usage

```
simpleEstimate(counts, params = newSimpleParams())

## S3 method for class 'SingleCellExperiment'
simpleEstimate(counts, params = newSimpleParams())

## S3 method for class 'matrix'
simpleEstimate(counts, params = newSimpleParams())
```

Arguments

counts	either a counts matrix or a SingleCellExperiment object containing count data to estimate parameters from.
params	SimpleParams object to store estimated values in.

Details

The nGenes and nCells parameters are taken from the size of the input data. The mean parameters are estimated by fitting a gamma distribution to the library size normalised mean expression level using `fitdist`. See [SimpleParams](#) for more details on the parameters.

Value

SimpleParams object containing the estimated parameters.

Examples

```
# Load example data
library(scuttle)
set.seed(1)
sce <- mockSCE()

params <- simpleEstimate(sce)
params
```

SimpleParams	<i>The SimpleParams class</i>
--------------	-------------------------------

Description

S4 class that holds parameters for the simple simulation.

Parameters

The simple simulation uses the following parameters:

nGenes The number of genes to simulate.

nCells The number of cells to simulate.

[seed] Seed to use for generating random numbers.

mean.shape The shape parameter for the mean gamma distribution.

mean.rate The rate parameter for the mean gamma distribution.

[count.disp] The dispersion parameter for the counts negative binomial distribution.

The parameters not shown in brackets can be estimated from real data using [simpleEstimate](#). For details of the simple simulation see [simpleSimulate](#).

simpleSimulate	<i>Simple simulation</i>
----------------	--------------------------

Description

Simulate counts from a simple negative binomial distribution without simulated library sizes, differential expression etc.

Usage

```
simpleSimulate(
  params = newSimpleParams(),
  sparsify = TRUE,
  verbose = TRUE,
  ...
)
```

Arguments

params	SimpleParams object containing simulation parameters.
sparsify	logical. Whether to automatically convert assays to sparse matrices if there will be a size reduction.
verbose	logical. Whether to print progress messages
...	any additional parameter settings to override what is provided in params.

Details

Gene means are simulated from a gamma distribution with `shape = mean.shape` and `rate = mean.rate`. Counts are then simulated from a negative binomial distribution with `mu = means` and `size = 1 / counts.disp`. See [SimpleParams](#) for more details of the parameters.

Value

SingleCellExperiment containing simulated counts

Examples

```
sim <- simpleSimulate()
# Override default parameters
sim <- simpleSimulate(nGenes = 1000, nCells = 50)
```

<code>sparseDCEstimate</code>	<i>Estimate SparseDC simulation parameters</i>
-------------------------------	--

Description

Estimate simulation parameters for the SparseDC simulation from a real dataset.

Usage

```
sparseDCEstimate(
  counts,
  conditions,
  nclusters,
  norm = TRUE,
  params = newSparseDCParams()
)

## S3 method for class 'SingleCellExperiment'
sparseDCEstimate(
  counts,
  conditions,
  nclusters,
  norm = TRUE,
  params = newSparseDCParams()
)

## S3 method for class 'matrix'
sparseDCEstimate(
  counts,
  conditions,
  nclusters,
  norm = TRUE,
```

```

  params = newSparseDCParams()
)

```

Arguments

counts	either a counts matrix or an SingleCellExperiment object containing count data to estimate parameters from.
conditions	numeric vector giving the condition each cell belongs to.
nclusters	number of cluster present in the dataset.
norm	logical, whether to library size normalise counts before estimation. Set this to FALSE if counts is already normalised.
params	PhenoParams object to store estimated values in.

Details

The nGenes and nCells parameters are taken from the size of the input data. The counts are preprocessed using [pre_proc_data](#) and then parameters are estimated using [sparsedc_cluster](#) using lambda values calculated using [lambda1_calculator](#) and [lambda2_calculator](#).

See [SparseDCParams](#) for more details on the parameters.

Value

SparseParams object containing the estimated parameters.

Examples

```

if (requireNamespace("SparseDC", quietly = TRUE)) {
  # Load example data
  library(scuttle)
  set.seed(1)
  sce <- mockSCE(ncells = 20, ngenes = 100)

  conditions <- sample(1:2, ncol(sce), replace = TRUE)

  params <- sparseDCEstimate(sce, conditions, nclusters = 3)
  params
}

```

SparseDCParams

The SparseDCParams class

Description

S4 class that holds parameters for the SparseDC simulation.

Parameters

The SparseDC simulation uses the following parameters:

nGenes The number of genes to simulate in each condition.

nCells The number of cells to simulate.

[seed] Seed to use for generating random numbers.

markers.n Number of marker genes to simulate for each cluster.

markers.shared Number of marker genes for each cluster shared between conditions. Must be less than or equal to markers.n.

[markers.same] Logical. Whether each cluster should have the same set of marker genes.

clusts.c1 Numeric vector of clusters present in condition 1. The number of times a cluster is repeated controls the proportion of cells from that cluster.

clusts.c2 Numeric vector of clusters present in condition 2. The number of times a cluster is repeated controls the proportion of cells from that cluster.

[mean.lower] Lower bound for cluster gene means.

[mean.upper] Upper bound for cluster gene means.

The parameters not shown in brackets can be estimated from real data using [sparseDCEstimate](#). For details of the SparseDC simulation see [sparseDCSimulate](#).

sparseDCSimulate	<i>SparseDC simulation</i>
------------------	----------------------------

Description

Simulate counts from cluster in two conditions using the SparseDC method.

Usage

```
sparseDCSimulate(
  params = newSparseDCParams(),
  sparsify = TRUE,
  verbose = TRUE,
  ...
)
```

Arguments

params	SparseDCParams object containing simulation parameters.
sparsify	logical. Whether to automatically convert assays to sparse matrices if there will be a size reduction.
verbose	logical. Whether to print progress messages
...	any additional parameter settings to override what is provided in params.

Details

This function is just a wrapper around `sim_data` that takes a `SparseDCParams`, runs the simulation then converts the output from log-expression to counts and returns a `SingleCellExperiment` object. The original simulated log-expression values are returned in the `LogExprs` assay. See `sim_data` and the SparseDC paper for more details about how the simulation works.

Value

SingleCellExperiment containing simulated counts

References

Campbell K, Yau C. Uncovering genomic trajectories with heterogeneous genetic and environmental backgrounds across single-cells and populations. bioRxiv (2017).

Barron M, Zhang S, Li J. A sparse differential clustering algorithm for tracing cell type changes via single-cell RNA-sequencing data. Nucleic Acids Research (2017).

Paper: [10.1093/nar/gkx1113](https://doi.org/10.1093/nar/gkx1113)

Examples

```
if (requireNamespace("SparseDC", quietly = TRUE)) {
  sim <- sparseDCSimulate()
}
```

splatEstimate	<i>Estimate Splat simulation parameters</i>
---------------	---

Description

Estimate simulation parameters for the Splat simulation from a real dataset. See the individual estimation functions for more details on how this is done.

Usage

```
splatEstimate(counts, params = newSplatParams())

## S3 method for class 'SingleCellExperiment'
splatEstimate(counts, params = newSplatParams())

## S3 method for class 'matrix'
splatEstimate(counts, params = newSplatParams())
```

Arguments

counts	either a counts matrix or a SingleCellExperiment object containing count data to estimate parameters from.
params	SplatParams object to store estimated values in.

Value

SplatParams object with estimated values.

See Also

[splatEstMean](#), [splatEstLib](#), [splatEstOutlier](#), [splatEstBCV](#), [splatEstDropout](#)

Examples

```
# Load example data
library(scuttle)
set.seed(1)
sce <- mockSCE()

params <- splatEstimate(sce)
params
```

SplatParams

The SplatParams class

Description

S4 class that holds parameters for the Splat simulation.

Parameters

The Splat simulation requires the following parameters:

`nGenes` The number of genes to simulate.

`nCells` The number of cells to simulate.

`[seed]` Seed to use for generating random numbers.

Batch parameters `[nBatches]` The number of batches to simulate.

`[batchCells]` Vector giving the number of cells in each batch.

`[batch.facLoc]` Location (meanlog) parameter for the batch effect factor log-normal distribution. Can be a vector.

`[batch.facScale]` Scale (sdlog) parameter for the batch effect factor log-normal distribution. Can be a vector.

`[batch.rmEffect]` Logical, removes the batch effect and continues with the simulation when TRUE. This allows the user to test batch removal algorithms without having to calculate the new expected cell means with batch removed.

Mean parameters `mean.shape` Shape parameter for the mean gamma distribution.

`mean.rate` Rate parameter for the mean gamma distribution.

Library size parameters `lib.loc` Location (meanlog) parameter for the library size log-normal distribution, or mean parameter if a normal distribution is used.

`lib.scale` Scale (sdlog) parameter for the library size log-normal distribution, or sd parameter if a normal distribution is used.

`lib.norm` Logical. Whether to use a normal distribution for library sizes instead of a log-normal.

Expression outlier parameters `out.prob` Probability that a gene is an expression outlier.

`out.facLoc` Location (meanlog) parameter for the expression outlier factor log-normal distribution.

`out.facScale` Scale (sdlog) parameter for the expression outlier factor log-normal distribution.

Group parameters `[nGroups]` The number of groups or paths to simulate.

`[group.prob]` Probability that a cell comes from a group.

Differential expression parameters `[de.prob]` Probability that a gene is differentially expressed in a group. Can be a vector.

`[de.downProb]` Probability that a differentially expressed gene is down-regulated. Can be a vector.

`[de.facLoc]` Location (meanlog) parameter for the differential expression factor log-normal distribution. Can be a vector.

`[de.facScale]` Scale (sdlog) parameter for the differential expression factor log-normal distribution. Can be a vector.

Biological Coefficient of Variation parameters `bcv.common` Underlying common dispersion across all genes.

`bcv.df` Degrees of Freedom for the BCV inverse chi-squared distribution.

Dropout parameters `dropout.type` The type of dropout to simulate. "none" indicates no dropout, "experiment" is global dropout using the same parameters for every cell, "batch" uses the same parameters for every cell in each batch, "group" uses the same parameters for every cell in each groups and "cell" uses a different set of parameters for each cell.

`dropout.mid` Midpoint parameter for the dropout logistic function.

`dropout.shape` Shape parameter for the dropout logistic function.

Differentiation path parameters `[path.from]` Vector giving the originating point of each path. This allows path structure such as a cell type which differentiates into an intermediate cell type that then differentiates into two mature cell types. A path structure of this form would have a "from" parameter of $c(0, 1, 1)$ (where 0 is the origin). If no vector is given all paths will start at the origin.

`[path.nSteps]` Vector giving the number of steps to simulate along each path. If a single value is given it will be applied to all paths. This parameter was previously called `path.length`.

`[path.skew]` Vector giving the skew of each path. Values closer to 1 will give more cells towards the starting population, values closer to 0 will give more cells towards the final population. If a single value is given it will be applied to all paths.

`[path.nonlinearProb]` Probability that a gene follows a non-linear path along the differentiation path. This allows more complex gene patterns such as a gene being equally expressed at the beginning and end of a path but lowly expressed in the middle.

`[path.sigmaFac]` Sigma factor for non-linear gene paths. A higher value will result in more extreme non-linear variations along a path.

The parameters not shown in brackets can be estimated from real data using [splatEstimate](#). For details of the Splat simulation see [splatSimulate](#).

splatPopEstimate *Estimate population/eQTL simulation parameters*

Description

Estimate simulation parameters for the eQTL population simulation from real data. See the individual estimation functions for more details on how this is done.

Usage

```
splatPopEstimate(
  counts = NULL,
  means = NULL,
  eqtl = NULL,
  params = newSplatPopParams()
)
```

Arguments

counts	either a counts matrix or a SingleCellExperiment object containing count data to estimate parameters from.
means	Matrix of real gene means across a population, where each row is a gene and each column is an individual in the population.
eqtl	data.frame with all or top eQTL pairs from a real eQTL analysis. Must include columns: 'gene_id', 'pval_nominal', and 'slope'.
params	SplatPopParams object containing parameters for the simulation of the mean expression levels for the population. See SplatPopParams for details.

Value

SplatPopParams object containing the estimated parameters.

See Also

[splatPopEstimateEffectSize](#), [splatPopEstimateMeanCV](#)

Examples

```
if (requireNamespace("VariantAnnotation", quietly = TRUE) &&
    requireNamespace("preprocessCore", quietly = TRUE)) {
  # Load example data
  library(scuttle)

  sce <- mockSCE()
```

```

    params <- splatPopEstimate(sce)
  }

```

SplatPopParams

The SplatPopParams class

Description

S4 class that holds parameters for the splatPop simulation.

Parameters

In addition to the [SplatParams](#) parameters, splatPop simulation requires the following parameters:

[similarity.scale] Scaling factor for pop.cv.param.rate, where values larger than 1 increase the similarity between individuals in the population and values less than one make the individuals less similar.

[eqtl.n] The number (>1) or percent (<=1) of genes to assign eQTL effects.

[eqtl.dist] Maximum distance between eSNP and eGene

[eqtl.maf.min] Minimum Minor Allele Frequency of eSNPs.

[eqtl.maf.max] Maximum Minor Allele Frequency of eSNPs.

[eqtl.coreg] Proportion of eGenes to have a shared eSNP (i.e., co-regulated genes)

[eqtl.group.specific] Percent of eQTL effects to simulate as group specific.

[eqtl.condition.specific] Percent of eQTL effects to simulate as condition specific.

eQTL Effect size distribution parameters. Defaults estimated from GTEx eQTL mapping results, see vignette for more information.

Shape parameter for the effect size gamma distribution.

eqtl.ES.rate Rate parameter for the effect size gamma distribution.

Bulk Mean Expression distribution parameters. Defaults estimated from GTEx data, see vignette for more information.

Shape parameter for the mean (i.e. bulk) expression gamma distribution

pop.mean.rate Rate parameter for the mean (i.e. bulk) expression gamma distribution

Bulk Expression Coefficient of Variation distribution parameters binned. Defaults estimated from GTEx data, see vignette for more information.

Dataframe containing gene mean bin range, and the CV shape, and CV rate parameters for each of those bins.

Specify number of samples per batch. Note that splatPop will randomly assign donors to be present in multiple batches to each batch.

The number of donors in each pool/batch.

Specify shape and rate of gamma distribution to sample number of cells per batch per donor. Will only be used if nCells per batch is set.

True/False if nCells should be set as nCells or sampled from a gamma distribution for each batch/donor.

nCells.shape Shape parameter for the nCells per batch per donor distribution.

nCells.rate Rate parameter for the nCells per batch per donor distribution.

Condition/treatment differential expression parameters [nConditions] The number of conditions/treatments to divide samples into.

- [condition.prob] Probability that a sample belongs to each condition/treatment group. Can be a vector.
- [cde.prob] Probability that a gene is differentially expressed in a condition group. Can be a vector.
- [cde.downProb] Probability that a conditionally differentially expressed gene is down-regulated. Can be a vector.
- [cde.facLoc] Location (meanlog) parameter for the conditional differential expression factor log-normal distribution. Can be a vector.
- [cde.facScale] Scale (sdlog) parameter for the conditional differential expression factor log-normal distribution. Can be a vector.

The parameters not shown in brackets can be estimated from real data using [splatPopEstimate](#). For details of the eQTL simulation see [splatPopSimulate](#).

splatPopParseEmpirical

splatPopParseEmpirical

Description

Parse splatPop key information from empirical data provided.

Usage

```
splatPopParseEmpirical(
  vcf = vcf,
  gff = gff,
  eqtl = eqtl,
  means = means,
  params = params
)
```

Arguments

- | | |
|--------|---|
| vcf | VariantAnnotation object containing genotypes of samples. |
| gff | Either NULL or a data.frame object containing a GFF/GTF file. |
| eqtl | Either NULL or if simulating population parameters directly from empirical data, a data.frame with empirical/desired eQTL results. To see required format, run ‘mockEmpiricalSet()’ and see eqtl output. |
| means | Either NULL or if simulating population parameters directly from empirical data, a Matrix of real gene means across a population, where each row is a gene and each column is an individual in the population. To see required format, run ‘mockEmpiricalSet()’ and see means output. |
| params | SplatPopParams object containing parameters for population scale simulations. See SplatPopParams for details. |

Details

NOTE: This function will cause some of the parameters in the `splatPopParams` object to be ignored, such as population level gene mean and variance and eQTL parameters.

This function will ignore a number of parameters defined in `splatPopParams`, instead pulling key information directly from provided VCF, GFF, gene means, and eQTL mapping result data provided.

Value

A partial `splatPop` ‘key’

<code>splatPopQuantNorm</code>	<i>Quantile normalize by sample to fit sc expression distribution.</i>
--------------------------------	--

Description

For each sample, expression values are quantile normalized (`qgamma`) using the gamma distribution parameterized from `splatEstimate()`. This ensures the simulated gene means reflect the distribution expected from a sc dataset and not a bulk dataset.

Usage

```
splatPopQuantNorm(params, means)
```

Arguments

<code>params</code>	SplatPopParams object containing parameters for population scale simulations. See SplatPopParams for details.
<code>means</code>	Mean gene expression matrix with eQTL effects.

Value

matrix of quantile normalized gene mean expression levels.

Examples

```
if (requireNamespace("VariantAnnotation", quietly = TRUE) &&
    requireNamespace("preprocessCore", quietly = TRUE)) {
  bulk.means <- mockBulkMatrix(n.genes = 100, n.samples = 100)
  bulk.qnorm <- splatPopQuantNorm(newSplatPopParams(), bulk.means)
}
```

splatPopSimulate	<i>splatPop simulation</i>
------------------	----------------------------

Description

Simulate scRNA-seq count data using the splat model for a population of individuals with correlation structure.

Usage

```
splatPopSimulate(
  params = newSplatPopParams(nGenes = 50),
  vcf = mockVCF(),
  method = c("single", "groups", "paths"),
  gff = NULL,
  eqtl = NULL,
  means = NULL,
  key = NULL,
  counts.only = FALSE,
  sparsify = TRUE,
  verbose = TRUE,
  ...
)
```

Arguments

params	SplatPopParams object containing parameters for population scale simulations. See SplatPopParams for details.
vcf	VariantAnnotation object containing genotypes of samples.
method	which simulation method to use. Options are "single" which produces a single population, "groups" which produces distinct groups (eg. cell types), "paths" which selects cells from continuous trajectories (eg. differentiation processes).
gff	Either NULL or a data.frame object containing a GFF/GTF file.
eqtl	Either NULL or if simulating population parameters directly from empirical data, a data.frame with empirical/desired eQTL results. To see required format, run 'mockEmpiricalSet()' and see eqtl output.
means	Either NULL or if simulating population parameters directly from empirical data, a Matrix of real gene means across a population, where each row is a gene and each column is an individual in the population. To see required format, run 'mockEmpiricalSet()' and see means output.
key	Either NULL or a data.frame object containing a full or partial splatPop key.
counts.only	logical. Whether to save only counts in sce object.
sparsify	logical. Whether to automatically convert assays to sparse matrices if there will be a size reduction.
verbose	logical. Whether to print progress messages.
...	any additional parameter settings to override what is provided in params.

Details

This function is for simulating data in a single step. It consists of a call to [splatPopSimulateMeans](#), which simulates a mean expression level per gene per sample, followed by a call to [splatPopSimulateSC](#), which uses the splat model to simulate single-cell counts per individual. Please see the documentation for those functions for more details.

Value

SingleCellExperiment object containing simulated counts, intermediate values like the gene means simulated in 'splatPopSimulateMeans', and information about the differential expression and eQTL effects assigned to each gene.

See Also

[splatPopSimulateMeans](#), [splatPopSimulateSC](#)

Examples

```
if (requireNamespace("VariantAnnotation", quietly = TRUE) &&
    requireNamespace("preprocessCore", quietly = TRUE)) {
  vcf <- mockVCF()
  gff <- mockGFF()
  sim <- splatPopSimulate(vcf = vcf, gff = gff, sparsify = FALSE)
}
```

splatPopSimulateMeans *splatPopSimulateMeans*

Description

Simulate mean expression levels for all genes for all samples, with between sample correlation structure simulated with eQTL effects and with the option to simulate multiple groups (i.e. cell-types).

Usage

```
splatPopSimulateMeans(
  vcf = mockVCF(),
  params = newSplatPopParams(nGenes = 1000),
  verbose = TRUE,
  key = NULL,
  gff = NULL,
  eqtl = NULL,
  means = NULL,
  ...
)
```

Arguments

vcf	VariantAnnotation object containing genotypes of samples.
params	SplatPopParams object containing parameters for population scale simulations. See SplatPopParams for details.
verbose	logical. Whether to print progress messages.
key	Either FALSE or a data.frame object containing a full or partial splatPop key.
gff	Either NULL or a data.frame object containing a GFF/GTF file.
eqtl	Either NULL or if simulating population parameters directly from empirical data, a data.frame with empirical/desired eQTL results. To see required format, run 'mockEmpiricalSet()' and see eqtl output.
means	Either NULL or if simulating population parameters directly from empirical data, a Matrix of real gene means across a population, where each row is a gene and each column is an individual in the population. To see required format, run 'mockEmpiricalSet()' and see means output.
...	any additional parameter settings to override what is provided in params.

Details

SplatPopParams can be set in a variety of ways. 1. If not provided, default parameters are used. 2. Default parameters can be overridden by supplying desired parameters using [setParams](#). 3. Parameters can be estimated from real data of your choice using [splatPopEstimate](#).

'splatPopSimulateMeans' involves the following steps:

1. Load population key or generate random or GFF/GTF based key.
2. Format and subset genotype data from the VCF file.
3. If not in key, assign expression mean and variance to each gene.
4. If not in key, assign eGenes-eSNPs pairs and effect sizes.
5. If not in key and groups >1, assign subset of eQTL associations as group-specific and assign DEG group effects.
6. Simulate mean gene expression matrix without eQTL effects
7. Quantile normalize by sample to fit single-cell expression distribution as defined in 'splatEstimate'.
8. Add quantile normalized gene mean and cv info the eQTL key.
9. Add eQTL effects to means matrix.

Value

A list containing: 'means' a matrix (or list of matrices if n.groups > 1) with the simulated mean gene expression value for each gene (row) and each sample (column), 'key' a data.frame with population information including eQTL and group effects, and 'condition' a named array containing conditional group assignments for each sample.

See Also

[splatPopParseVCF](#), [splatPopParseGenes](#), [splatPopAssignMeans](#), [splatPopQuantNorm](#), [splatPopQuantNormKey](#), [splatPopQTLEffects](#), [splatPopGroupEffects](#), [splatPopSimMeans](#), [splatPopSimEffects](#),

Examples

```
if (requireNamespace("VariantAnnotation", quietly = TRUE) &&
    requireNamespace("preprocessCore", quietly = TRUE)) {
  means <- splatPopSimulateMeans()
}
```

splatPopSimulateSC *splatPopSimulateSC*

Description

Simulate count data for a population from a fictional single-cell RNA-seq experiment using the Splat method.

Usage

```
splatPopSimulateSC(
  sim.means,
  params,
  key,
  method = c("single", "groups", "paths"),
  counts.only = FALSE,
  conditions = NULL,
  sparsify = TRUE,
  verbose = TRUE,
  ...
)
```

Arguments

sim.means	Matrix or list of matrices of gene means for the population. Output from ‘splatPopSimulateMeans()’.
params	SplatPopParams object containing parameters for population scale simulations. See SplatPopParams for details.
key	data.frame object containing a full or partial splatPop key. Output from ‘splatPopSimulateMeans()’.
method	which simulation method to use. Options are "single" which produces a single cell population for each sample, "groups" which produces distinct groups (eg. cell types) for each sample (note, this creates separate groups from those created in ‘popSimulate’ with only DE effects), and "paths" which selects cells from continuous trajectories (eg. differentiation processes).

counts.only	logical. Whether to return only the counts.
conditions	named array with conditional group assignment for each sample. Output from 'splatPopSimulateMeans()'.
sparsify	logical. Whether to automatically convert assays to sparse matrices if there will be a size reduction.
verbose	logical. Whether to print progress messages.
...	any additional parameter settings to override what is provided in params.

Value

SingleCellExperiment object containing simulated counts, intermediate values like the gene means simulated in 'splatPopSimulateMeans', and information about the differential expression and eQTL effects assigned to each gene.

Examples

```
if (requireNamespace("VariantAnnotation", quietly = TRUE) &&
    requireNamespace("preprocessCore", quietly = TRUE)) {
  params <- newSplatPopParams()
  sim.means <- splatPopSimulateMeans()
  sim <- splatPopSimulateSC(sim.means$means, params, sim.means$key)
}
```

splatSimulate	<i>Splat simulation</i>
---------------	-------------------------

Description

Simulate count data from a fictional single-cell RNA-seq experiment using the Splat method.

Usage

```
splatSimulate(
  params = newSplatParams(),
  method = c("single", "groups", "paths"),
  sparsify = TRUE,
  verbose = TRUE,
  ...
)

splatSimulateSingle(params = newSplatParams(), verbose = TRUE, ...)

splatSimulateGroups(params = newSplatParams(), verbose = TRUE, ...)

splatSimulatePaths(params = newSplatParams(), verbose = TRUE, ...)
```

Arguments

params	SplatParams object containing parameters for the simulation. See SplatParams for details.
method	which simulation method to use. Options are "single" which produces a single population, "groups" which produces distinct groups (eg. cell types), or "paths" which selects cells from continuous trajectories (eg. differentiation processes).
sparsify	logical. Whether to automatically convert assays to sparse matrices if there will be a size reduction.
verbose	logical. Whether to print progress messages.
...	any additional parameter settings to override what is provided in params.

Details

Parameters can be set in a variety of ways. If no parameters are provided the default parameters are used. Any parameters in `params` can be overridden by supplying additional arguments through a call to `setParams`. This design allows the user flexibility in how they supply parameters and allows small adjustments without creating a new `SplatParams` object. See examples for a demonstration of how this can be used.

The simulation involves the following steps:

1. Set up simulation object
2. Simulate library sizes
3. Simulate gene means
4. Simulate groups/paths
5. Simulate BCV adjusted cell means
6. Simulate true counts
7. Simulate dropout
8. Create final dataset

The final output is a `SingleCellExperiment` object that contains the simulated counts but also the values for various intermediate steps. These are stored in the `colData` (for cell specific information), `rowData` (for gene specific information) or `assays` (for gene by cell matrices) slots. This additional information includes:

`colData` **Cell** Unique cell identifier.

Group The group or path the cell belongs to.

ExpLibSize The expected library size for that cell.

Step (paths only) how far along the path each cell is.

`rowData` **Gene** Unique gene identifier.

BaseGeneMean The base expression level for that gene.

OutlierFactor Expression outlier factor for that gene. Values of 1 indicate the gene is not an expression outlier.

GeneMean Expression level after applying outlier factors.

BatchFac[Batch] The batch effects factor for each gene for a particular batch.

DEFac[Group] The differential expression factor for each gene in a particular group. Values of 1 indicate the gene is not differentially expressed.

SigmaFac[Path] Factor applied to genes that have non-linear changes in expression along a path.

assays **BatchCellMeans** The mean expression of genes in each cell after adding batch effects.

BaseCellMeans The mean expression of genes in each cell after any differential expression and adjusted for expected library size.

BCV The Biological Coefficient of Variation for each gene in each cell.

CellMeans The mean expression level of genes in each cell adjusted for BCV.

TrueCounts The simulated counts before dropout.

Dropout Logical matrix showing which values have been dropped in which cells.

Values that have been added by Splatter are named using UpperCamelCase in order to differentiate them from the values added by analysis packages which typically use underscore_naming.

Value

SingleCellExperiment object containing the simulated counts and intermediate values.

References

Zappia L, Phipson B, Oshlack A. Splatter: simulation of single-cell RNA sequencing data. *Genome Biology* (2017).

Paper: [10.1186/s13059-017-1305-0](https://doi.org/10.1186/s13059-017-1305-0)

Code: <https://github.com/Oshlack/splatter>

See Also

[splatSimLibSizes](#), [splatSimGeneMeans](#), [splatSimBatchEffects](#), [splatSimBatchCellMeans](#), [splatSimDE](#), [splatSimCellMeans](#), [splatSimBCVMeans](#), [splatSimTrueCounts](#), [splatSimDropout](#)

Examples

```
# Simulation with default parameters
sim <- splatSimulate()

# Simulation with different number of genes
sim <- splatSimulate(nGenes = 1000)
# Simulation with custom parameters
params <- newSplatParams(nGenes = 100, mean.rate = 0.5)
sim <- splatSimulate(params)
# Simulation with adjusted custom parameters
sim <- splatSimulate(params, mean.rate = 0.6, out.prob = 0.2)
# Simulate groups
sim <- splatSimulate(method = "groups")
# Simulate paths
sim <- splatSimulate(method = "paths")
```

summariseDiff	<i>Summarise diffSCEs</i>
---------------	---------------------------

Description

Summarise the results of `diffSCEs`. Calculates the Median Absolute Deviation (MAD), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and Kolmogorov-Smirnov (KS) statistics for the various properties and ranks them.

Usage

```
summariseDiff(diff)
```

Arguments

`diff` Output from `diffSCEs`

Value

data.frame with MADs, MAEs, RMSEs, scaled statistics and ranks

Examples

```
sim1 <- splatSimulate(nGenes = 1000, batchCells = 20)
sim2 <- simpleSimulate(nGenes = 1000, nCells = 20)
difference <- diffSCEs(list(Splat = sim1, Simple = sim2), ref = "Simple")
summary <- summariseDiff(difference)
head(summary)
```

zinbEstimate	<i>Estimate ZINB-WaVE simulation parameters</i>
--------------	---

Description

Estimate simulation parameters for the ZINB-WaVE simulation from a real dataset.

Usage

```
zinbEstimate(
  counts,
  design.samples = NULL,
  design.genes = NULL,
  common.disp = TRUE,
  iter.init = 2,
  iter.opt = 25,
  stop.opt = 1e-04,
```

```

    params = newZINBParams(),
    verbose = TRUE,
    BPPARAM = SerialParam(),
    ...
)

## S3 method for class 'SingleCellExperiment'
zinbEstimate(
  counts,
  design.samples = NULL,
  design.genes = NULL,
  common.disp = TRUE,
  iter.init = 2,
  iter.opt = 25,
  stop.opt = 1e-04,
  params = newZINBParams(),
  verbose = TRUE,
  BPPARAM = SerialParam(),
  ...
)

## S3 method for class 'matrix'
zinbEstimate(
  counts,
  design.samples = NULL,
  design.genes = NULL,
  common.disp = TRUE,
  iter.init = 2,
  iter.opt = 25,
  stop.opt = 1e-04,
  params = newZINBParams(),
  verbose = TRUE,
  BPPARAM = SerialParam(),
  ...
)

```

Arguments

counts	either a counts matrix or a SingleCellExperiment object containing count data to estimate parameters from.
design.samples	design matrix of sample-level covariates.
design.genes	design matrix of gene-level covariates.
common.disp	logical. Whether or not a single dispersion for all features is estimated.
iter.init	number of iterations to use for initialization.
iter.opt	number of iterations to use for optimization.
stop.opt	stopping criterion for optimization.
params	ZINBParams object to store estimated values in.

verbose	logical. Whether to print progress messages.
BPPARAM	A BiocParallelParam instance giving the parallel back-end to be used. Default is SerialParam which uses a single core.
...	additional arguments passes to zinbFit .

Details

The function is a wrapper around [zinbFit](#) that takes the fitted model and inserts it into a [ZINBParams](#) object. See [ZINBParams](#) for more details on the parameters and [zinbFit](#) for details of the estimation procedure.

Value

ZINBParams object containing the estimated parameters.

Examples

```
if (requireNamespace("zinbwave", quietly = TRUE)) {
  library(scuttle)
  set.seed(1)
  sce <- mockSCE(ncells = 20, ngenes = 100)

  params <- zinbEstimate(sce)
  params
}
```

ZINBParams

The ZINBParams class

Description

S4 class that holds parameters for the ZINB-WaVE simulation.

Parameters

The ZINB-WaVE simulation uses the following parameters:

nGenes The number of genes to simulate.

nCells The number of cells to simulate.

[seed] Seed to use for generating random numbers.

model Object describing a ZINB model.

The majority of the parameters for this simulation are stored in a [ZinbModel](#) object. Please refer to the documentation for this class and its constructor([zinbModel](#)) for details about all the parameters.

The parameters not shown in brackets can be estimated from real data using [zinbEstimate](#). For details of the ZINB-WaVE simulation see [zinbSimulate](#).

zinbSimulate	<i>ZINB-WaVE simulation</i>
--------------	-----------------------------

Description

Simulate counts using the ZINB-WaVE method.

Usage

```
zinbSimulate(params = newZINBParams(), sparsify = TRUE, verbose = TRUE, ...)
```

Arguments

params	ZINBParams object containing simulation parameters.
sparsify	logical. Whether to automatically convert assays to sparse matrices if there will be a size reduction.
verbose	logical. Whether to print progress messages
...	any additional parameter settings to override what is provided in params.

Details

This function is just a wrapper around `zinbSim` that takes a `ZINBParams`, runs the simulation then converts the output to a `SingleCellExperiment` object. See `zinbSim` and the ZINB-WaVE paper for more details about how the simulation works.

Value

SingleCellExperiment containing simulated counts

References

Campbell K, Yau C. Uncovering genomic trajectories with heterogeneous genetic and environmental backgrounds across single-cells and populations. bioRxiv (2017).

Risso D, Perraudeau F, Gribkova S, Dudoit S, Vert J-P. ZINB-WaVE: A general and flexible method for signal extraction from single-cell RNA-seq data bioRxiv (2017).

Paper: [10.1101/125112](https://doi.org/10.1101/125112)

Code: <https://github.com/drisso/zinbwave>

Examples

```
if (requireNamespace("zinbwave", quietly = TRUE)) {  
  sim <- zinbSimulate()  
}
```

Index

addGeneLengths, 3
assays, 15, 62

BASiCS_MCMC, 6
BASiCS_Sim, 8
BASiCSEstimate, 4, 7
BASiCSParams, 7, 8
BASiCSParams-class (BASiCSParams), 7
BASiCSSimulate, 7, 7
BiocParallelParam, 20, 40, 42, 66

colData, 15, 62
compareSCEs, 8, 26, 27
create_synthetic, 29, 30

density, 14
diffSCEs, 10, 26, 27, 64

empirical_lambda, 28

fitdist, 45

getParam, 11
getParam,Params-method (getParam), 11
getParams, 12
ggplot, 9, 11

kersplatEstBCV, 13
kersplatEstimate, 12, 14
kersplatEstLib, 13
kersplatEstMean, 13
kersplatGenNetwork, 17
KersplatParams, 13, 15, 17
KersplatParams-class (KersplatParams), 13
kersplatSample, 15, 17, 18
kersplatSelectRegs, 17
kersplatSetup, 16, 18
kersplatSimAmbientCounts, 16
kersplatSimCellCounts, 16
kersplatSimCellMeans, 14, 16

kersplatSimCounts, 16
kersplatSimGeneMeans, 17
kersplatSimLibSizes, 16
kersplatSimPaths, 14, 17
kersplatSimulate, 14, 18

lambda1_calculator, 48
lambda2_calculator, 48
listSims, 19
lun2Estimate, 19, 21, 22
Lun2Params, 20, 21
Lun2Params-class (Lun2Params), 21
lun2Simulate, 21, 21
lunEstimate, 23, 24
LunParams, 23, 24, 25
LunParams-class (LunParams), 24
lunSimulate, 24, 24

makeCompPanel, 25
makeDiffPanel, 26
makeOverallPanel, 27
mfaEstimate, 28, 29
MFAParams, 28, 29, 30
MFAParams-class (MFAParams), 29
mfaSimulate, 29, 29
minimiseSCE, 30
mockBulkeQTL, 31
mockBulkMatrix, 32
mockEmpiricalSet, 32
mockGFF, 33
mockVCF, 34

newBASiCSParams (newParams), 35
newKersplatParams (newParams), 35
newLun2Params (newParams), 35
newLunParams (newParams), 35
newMFAParams (newParams), 35
newParams, 35
newPhenoParams (newParams), 35
newSCDDParams (newParams), 35

- newSimpleParams (newParams), 35
- newSparseDCParams (newParams), 35
- newSplatParams (newParams), 35
- newSplatPopParams (newParams), 35
- newZINBParams (newParams), 35
- Params, 36
- Params-class (Params), 36
- phenoEstimate, 36, 37
- PhenoParams, 36, 37, 38
- PhenoParams-class (PhenoParams), 37
- phenoSimulate, 37, 38
- pre_proc_data, 48
- preprocess, 40
- rowData, 4, 15, 62
- scDD, 40
- scDDEstimate, 39, 41
- SCDDParams, 40, 42
- SCDDParams-class (SCDDParams), 40
- scDDSimulate, 41, 41
- SerialParam, 20, 40, 42, 66
- setParam, 42, 44
- setParam, BASiCSPParams-method (setParam), 42
- setParam, KersplatParams-method (setParam), 42
- setParam, Lun2Params-method (setParam), 42
- setParam, LunParams-method (setParam), 42
- setParam, Params-method (setParam), 42
- setParam, PhenoParams-method (setParam), 42
- setParam, SCDDParams-method (setParam), 42
- setParam, SplatParams-method (setParam), 42
- setParam, SplatPopParams-method (setParam), 42
- setParam, ZINBParams-method (setParam), 42
- setParams, 35, 44, 59, 62
- setParams, KersplatParams-method (setParams), 44
- setParams, Params-method (setParams), 44
- setParams, SplatParams-method (setParams), 44
- sim_data, 50
- simpleEstimate, 45, 46
- SimpleParams, 45, 46, 47
- SimpleParams-class (SimpleParams), 46
- simpleSimulate, 46, 46
- simulate_phenopath, 38
- simulateSet, 41, 42
- SingleCellExperiment, 4, 8, 15, 30, 38, 41, 42, 50, 62, 67
- sparsedc_cluster, 48
- sparseDCEstimate, 47, 49
- SparseDCParams, 48, 48, 50
- SparseDCParams-class (SparseDCParams), 48
- sparseDCSimulate, 49, 49
- splatEstBCV, 51
- splatEstDropout, 51
- splatEstimate, 50, 53
- splatEstLib, 51
- splatEstMean, 51
- splatEstOutlier, 51
- SplatParams, 51, 54, 62
- SplatParams-class (SplatParams), 51
- splatPopAssignMeans, 60
- splatPopQTLEffects, 60
- splatPopEstimate, 53, 55, 59
- splatPopEstimateEffectSize, 53
- splatPopEstimateMeanCV, 53
- splatPopGroupEffects, 60
- SplatPopParams, 53, 54, 55–57, 59, 60
- SplatPopParams-class (SplatPopParams), 54
- splatPopParseEmpirical, 55
- splatPopParseGenes, 60
- splatPopParseVCF, 60
- splatPopQuantNorm, 56, 60
- splatPopQuantNormKey, 60
- splatPopSimEffects, 60
- splatPopSimMeans, 60
- splatPopSimulate, 55, 57
- splatPopSimulateMeans, 58, 58
- splatPopSimulateSC, 58, 60
- splatSimBatchCellMeans, 63
- splatSimBatchEffects, 63
- splatSimBCVMeans, 63
- splatSimCellMeans, 63
- splatSimDE, 63
- splatSimDropout, 63
- splatSimGeneMeans, 63

splatSimLibSizes, [63](#)
splatSimTrueCounts, [63](#)
splatSimulate, [53](#), [61](#)
splatSimulateGroups (splatSimulate), [61](#)
splatSimulatePaths (splatSimulate), [61](#)
splatSimulateSingle (splatSimulate), [61](#)
summariseDiff, [64](#)

zinbEstimate, [64](#), [66](#)
zinbFit, [66](#)
ZinbModel, [66](#)
zinbModel, [66](#)
ZINBParams, [66](#), [66](#), [67](#)
ZINBParams-class (ZINBParams), [66](#)
zinbSim, [67](#)
zinbSimulate, [66](#), [67](#)