Package 'GeneScape'

June 13, 2023

Type Package

Title Simulation of Single Cell RNA-Seq Data with Complex Structure

Version 1.0

Date 2023-6-4

Maintainer Qi Gao <qi.gao@duke.edu>

Description Simulating single cell RNA-seq data with complicated structure. This package is developed based on the Splat method (Zappia, Phipson and Oshlack (2017) <doi:10.1186/s13059-017-1305-0>). 'GeneScape' incorporates additional features to simulate single cell RNA-seq data with complicated differential expression and correlation structures, such as sub-cell-types, correlated genes (pathway genes) and hub genes.

Encoding UTF-8

License GPL (>= 3)

Imports MASS (>= 7.3-53.1), corpcor (>= 1.6.10), stats

RoxygenNote 7.2.3

NeedsCompilation no

Author Qi Gao [aut, cre]

Repository CRAN

Date/Publication 2023-06-13 09:00:09 UTC

R topics documented:

`fcsim` *fcsim*

Description

This function similate differential expression fold change level

Usage

```
fcsim(n.gene, de.id, fc.loc, fc.scale)
```

Arguments

`n.gene`	total number of genes
`de.id`	index of differentially expressed genes
`fc.loc`	location parameter for fold change (log-normal distribution)
`fc.scale`	scale parameter for fold change (log-normal distribution)

References

Zappia, L., Phipson, B., & Oshlack, A. (2017). Splatter: Simulation of single-cell RNA sequencing data. Genome Biology, 18(1). https://doi.org/10.1186/s13059-017-1305-0

`GeneScape` *GeneScape*

Description

This function simulate single cell RNAseq data with complicated differential expression and correlation structure.

Usage

```
GeneScape(
  nCells = 6000,
  nGroups = NULL,
  groups = NULL,
  lib.size.loc = 9.3,
  lib.size.scale = 0.25,
  de.fc.mat = NULL,
  nGenes = 5000,
  gene.mean.shape = 0.3,
  gene.mean.rate = 0.15,
  gene.means = NULL,
  de.n = 50,
```

```
  de.share = NULL,
  de.id = NULL,
  de.fc.loc = 0.7,
  de.fc.scale = 0.2,
  add.sub = FALSE,
  sub.major = NULL,
  sub.prop = 0.1,
  sub.group = NULL,
  sub.de.n = 20,
  sub.de.id = NULL,
  sub.de.common = FALSE,
  sub.de.fc.loc = 0.7,
  sub.de.fc.scale = 0.2,
  add.cor = FALSE,
  cor.n = 4,
  cor.size = 20,
  cor.cor = 0.7,
  cor.id = NULL,
  band.width = 10,
  add.hub = FALSE,
  hub.n = 10,
  hub.size = 20,
  hub.cor = 0.4,
  hub.id = NULL,
  hub.fix = NULL,
  drop = FALSE,
  dropout.location = -2,
  dropout.slope = -1
)
```

Arguments

nCells	number of cells
nGroups	number of cell groups
groups	group information for cells
lib.size.loc	location parameter for library size (log-normal distribution)
lib.size.scale	scale parameter for library size (log-normal distribution)
de.fc.mat	differential expression fold change matrix, could be generated by this function
nGenes	number of genes
gene.mean.shape	shape parameter for mean expression level (Gamma distribution)
gene.mean.rate	rate parameter for mean expression level (Gamma distribution)
gene.means	mean gene expression levels
de.n	number of differentially expressed genes in each cell type. Should be a integer or a vector of length nGroups

de.share	number of shared DE genes between neighbor cell types. Should be a vector of length (nGroups - 1)
de.id	the index of genes that are DE across cell types. Should be a list of vectors. Each vector corresponds to a cell type. With non-null value of de.id, de.n and de.share would be ignored.
de.fc.loc	the location parameter for the fold change of DE genes. Should be a number, a vector of length nGroups
de.fc.scale	the scale parameter for fold change (log-normal distribution). Should be a number or a vector of length nGroups
add.sub	whether to add sub-cell-types
sub.major	the major cell types correspond to the sub-cell-types
sub.prop	proportion of sub-cell-types in the corresponding major cell type
sub.group	cell index for sub-cell-types. With non-null sub.group specified, sub.prop would be ignored.
sub.de.n	number of differentially expressed genes in each sub-cell-type compared to the corresponding major cell type. Should be a integer or a vector of length sub.major
sub.de.id	the index of additional differentially expressed genes between sub-cell-types and the corresponding major cell types
sub.de.common	whether the additional differential expression structure should be same for all sub-cell-types
sub.de.fc.loc	similar to de.fc.loc, but for addtional differentially expressed genes in sub-cell-types
sub.de.fc.scale	similar to de.fc.scale, but for addtional differentially expressed genes in sub-cell-types
add.cor	whether to add pathways (correlated genes)
cor.n	number of pathways included. Should be a integer.
cor.size	number of correlated genes (length of pathway). Should be a number or a vector of length cor.n
cor.cor	correlation parameters
cor.id	gene index of correlated (pathway) genes. Should be a list of vectors, with each vector represents a pathway. With non-null value of cor.id, cor.n would be ignored.
band.width	No correlation exists if distance of 2 genes are further than band_width in a pathway
add.hub	whether to add hub genes
hub.n	number of hub genes included. Should be a integer.
hub.size	number of genes correlated to the hub gene. Should be a number or a vector of length hub.n
hub.cor	correlation parameters between hub genes and their correlated genes

hub.id	gene index of hub genes. Should be a list of vectors. With non-null value of hub.id, hub.n would be ignored.
hub.fix	user defined genes correlated to hub genes (others are randomly selected). Should be a list of vectors of length hub.n or same as hub.id.
drop	whether to add dropout
dropout.location	dropout mid point (the mean expression level at which the probability is equal to 0.5, same as splat. Could be negative)
dropout.slope	how dropout proportion changes with increasing expression

Details

Compared to splat method in Splatter R package, this function can fix the number and position of differentially expressed genes, have more complicated differential expression structure, add sub-cell-types, correlated genes (AR(1) correlation structure with bound, mimicking pathways) and hub genes.

Value

A list of observed data, true data (without dropout), differential expression rate and hub gene indices.

References

Zappia, L., Phipson, B., & Oshlack, A. (2017). Splatter: Simulation of single-cell RNA sequencing data. Genome Biology, 18(1). https://doi.org/10.1186/s13059-017-1305-0

Examples

```
set.seed(1)
data <- GeneScape()
```

Index