Package 'scKWARN'

October 2, 2020

Type Package

Title Single-cell RNA Sequencing Normalization Using A Local Average Technique

Version 1.0.6

Date 2020-10-02

Author Hsu, Chih-Yuan

Maintainer Chih-Yuan Hsu <chih-yuan.hsu@vumc.org>

Description A normalization method for single-cell RNA sequencing data.

License GPL-2 | GPL-3

Imports Rcpp (>= 1.0.1), Matrix, stats, methods

LinkingTo Rcpp

LazyData true

RoxygenNote 6.1.1

Archs x64

R topics documented:

`LocASN` *Single-cell RNA sequencing normalization using a local average technique*

Description

A function of normalizing single cell RNA-seq gene expression.

Usage

```
LocASN(countmatrix, conditions = NULL, filter = FALSE,
       gene_num_gezero = 3, cell_num_gezero = 10,
       numGeneforEst = 2000, divideforFast = FALSE, numDivide = 2)
```

Arguments

countmatrix	Input. Unnormalized count (sparse) matrix (genes by cells).
conditions	Input (Optional). Indicate which cells are sampled from the same conditions. The default value, NULL, denotes all the cells are sampled from the same condition.
filter	Input (Optional). A logic value to indicate if need data filtering. If yes, please see the details of gene_num_gezero and cell_num_gezero for input. The default value if FALSE.
gene_num_gezero	Input (Optional). A threshold (interger) to determine the inclusion of a gene. The gene included needs to be expressed in at least *gene_num_gezero* cells. The default value is *gene_num_gezero* = 3.
cell_num_gezero	Input (Optional). A threshold (interger) to determine the inclusion of a cell. The cell included needs to contain at least *cell_num_gezero* expressed genes. The default value is *cell_num_gezero* = 10.
numGeneforEst	Input (Optional). Use top *numGeneforEst* (integer) highly expressed genes to estimate scaling factors. The default value is *numGeneforEst* = 2000.
divideforFast	Input (Optional). A logic value to indicate if speeding up computation by randomly dividing cells into *numDivide* groups. It is recommended to use for a large number of cells, for example, > 50K cells. The default value is FALSE. Please input an integer in *numDivide* below if *divideforFast* = TRUE
numDivide	Input (Optional). An integer is required if *divideforFast* = TRUE. Suggest *numDivide* = # of cells divided by 5K~10K.

Value

NormalizedData	Matrix (genes by cells). Data matrix after normalization.
scalingFactor	Vector. Cell-specific scaling factors.
delete_genes	Vector. Indeice of the genes deleted.
delete_cells	Vector. Indeice of the cells deleted.

Examples

```
set.seed(12345)
G <- 2000; n <- 600 # G: number of genes, n: number of cells
NB_cell <- function(j) rnbinom(G, size = 0.1, mu = rgamma(G, shape = 2, rate = 2))
countsimdata <- sapply(1:n, NB_cell)
colnames(countsimdata) <- paste("cell", 1:n, sep = "_")
rownames(countsimdata) <- paste("gene", 1:G, sep = "_")
Result <- LocASN(countmatrix = as(countsimdata,"sparseMatrix"))
Result$NormalizedData[1:10,1:10]; Result$scalingFactor[1:10]

#conditions <- c(rep(1,n/2), rep(2,n/2))
#Result2 <- LocASN(countmatrix = countsimdata, conditions = conditions)
#Result2$NormalizedData[1:10,1:10]; Result2$scalingFactor[1:10]
```

Index