Package 'scImpute'

January 12, 2018

Type Package

Title Accurate and robust imputation of single-cell RNA sequencing data

Version 0.0.5

Date 2018-01-12

Author Wei Vivian Li, Jingyi Jessica Li

Maintainer Wei Vivian Li <liw@ucla.edu>

Description scRNA-seq analysis is complicated by the excess of zero or near zero counts in the data, which are the so-called dropouts due to low amounts of mRNA within each individual cell. scImpute is developed to simultaneously determine which expression values are affected by dropout events in scRNA-seq data and perform imputation only on dropout entries.

Depends R (>= 3.3.2), parallel, stats, kernlab, penalized, utils, doParallel, foreach

License GPL

RoxygenNote 6.0.1

Suggests knitr,
rmarkdown

VignetteBuilder knitr

R topics documented:

---

scimpute	*use scImpute to impute dropout values in scRNA-seq data*

Description

use scImpute to impute dropout values in scRNA-seq data

Usage

```
scimpute(count_path, infile = "csv", outfile = "csv", out_dir,
  labeled = FALSE, drop_thre = 0.5, Kcluster = NULL, labels = NULL,
  ncores = 5)
```

1

Arguments

count_path   A character specifying the full path of the raw count matrix;

infile   A character specifying the type of file storing the raw count matrix; can be either "csv" or "txt". The input file shoule have rows representing genes and columns representing cells, with its first row as cell names and first column as gene names.

outfile   A character specifying the type of file storing the imputed count matrix; can be either "csv" or "txt".

out_dir   A character specifying the full path of the output directory, which is used to store all intermdediate and final outputs.

labeled   A logical value indicating whether cell type information is available. labels must be specified if labeled = TRUE.

drop_thre   A number between 0 and 1, specifying the threshold to determine dropout values.

Kcluster   An integer specifying the number of cell subpopulations. This parameter can be determined based on prior knowledge or clustering of raw data. Kcluster is used to determine the candidate neighbors of each cell.

labels   A character vector specifying the cell type of each column in the raw count matrix. Only needed when labeled = TRUE. Each cell type should have at least two cells for imputation.

ncores   A integer specifying the number of cores used for parallel computation.

Value

scImpute returns a vector giving the column indices of outlier cells. It saves the imputed count matrix to scimpute.csv or scimpute.txt (depending on outfile) to out_dir.

Author(s)

Wei Vivian Li, <liw@ucla.edu>

Jingyi Jessica Li, <jli@stat.ucla.edu>

References

https://www.biorxiv.org/content/early/2017/05/24/141598

Index