Package

January 11, 2019

Type Package

Title Impute scRNAseq data method

Version 0.0.1

Author Tao Peng

Maintainer The package maintainer Tao Peng <tpengmath@gmail.com>

Description SCRABBLE imputes drop-out data by optimizing an objective function
that consists of three terms. The first term ensures that imputed values for
genes with nonzero expression remain as close to their original values as possible,
thus minimizing unwanted bias towards expressed genes. The second term ensures the
rank of the imputed data matrix to be as small as possible. The rationale is that
we only expect a limited number of distinct cell types in the samples. The third term
operates on the bulk RNA-Seq data. It ensures consistency between the average gene
expression of the aggregated imputed data and the average gene expression of the bulk
RNA-Seq data. We developed a convex optimization algorithm to minimize the objective function.

License GPL-3

Encoding UTF-8

Depends R(>= 3.3)

LazyData true

Imports Rcpp (>= 0.12.13),
rARPACK,
pracma,
ggplot2,
RColorBrewer,
reshape2,
gridExtra

LinkingTo Rcpp, RcppEigen, gridExtra

RoxygenNote 6.1.1

Suggests knitr,rmarkdown,BiocStyle

VignetteBuilder knitr

R topics documented:

---

data   *Test data for scrabble*

Description

"data" is a data list with the length of 3. The first element in the list is generated drop-out scRNAseq data with 800 genes and 1000 cells. The second element in the list is the generated bulk RNAseq data with 800 genes. The third element is the true scRNAseq data without dropouts. The steps of generating the data is shown in Details section.

Usage

```
data_sc <- data[[1]]
data_bulk <- data[[2]]
data_true <- data[[3]]
```

Format

An object of class list of length 3.

Details

The data set was generated from the well-developed R package Splatter.

Author(s)

Tao Peng, Kai Tan

---

scrabble   *Runs SCRABBLE*

Description

SCRABBLE imputes drop-out data by optimizing an objective function that consists of three terms. The first term ensures that imputed values for genes with nonzero expression remain as close to their original values as possible, thus minimizing unwanted bias towards expressed genes. The second term ensures the rank of the imputed data matrix to be as small as possible. The rationale is that we only expect a limited number of distinct cell types in the samples. The third term operates on the bulk RNA-Seq data. It ensures consistency between the average gene expression of the aggregated imputed data and the average gene expression of the bulk RNA-Seq data. We developed 58 a convex optimization algorithm to minimize the objective function.

Usage

```
scrabble(data, parameter, nIter = 60, error_out_threshold = 1e-07,
  nIter_inner = 100, error_inner_threshold = 1e-05)
```

Arguments

data	the input data list. The input data is a list of two datasets, scRNAseq and bulk RNAseq.
parameter	the vector of parameters. The first parameter is the value of alpha in the mathematical model , the second one is the value of beta in the mathematical model.
nIter	the maximum iterations, the default is 60.
error_out_threshold	the threshold of the error between the current imputed matrix and the previous one. Default is 1e-5.
nIter_inner	the maximum interations of calculating the sub-optimization problem. Default is 60.
error_inner_threshold	the threshold of the error between the current updated matrix and the previous one. Default is 1e-5.

Value

A data matrix with the same size of the input scRNAseq data

Examples

```
# Set up the parameter used in SCRABBLE
parameter <- c(1, 1e-6, 1e-4)

# Run SCRABLE
result <- scrabble(data,parameter = parameter)
```

Index