Package 'DrImpute'

July 15, 2017

Version 1.0

Date 2017-7-15

Title Imputing Dropout Events in Single-Cell RNA-Sequencing Data

Description R codes for imputing dropout events. Many statistical methods in cell type identification, visualization and lineage reconstruction do not account for dropout events ('PCAReduce', 'SC3', 'PCA', 't-SNE', 'Monocle', 'TSCAN', etc). 'DrImpute' can improve the performance of such software by imputing dropout events.

Author Il-Youp Kwak with contributions from Wuming Gong

Maintainer Il-Youp Kwak <ilyoup.kwak@gmail.com>

Depends R (>= 3.1.0)

Imports Rcpp

Suggests knitr, rmarkdown, devtools, roxygen2, irlba

License GPL-3

VignetteBuilder knitr

URL https://github.com/ikwak2/DrImpute

LinkingTo Rcpp, RcppArmadillo

RoxygenNote 6.0.1

NeedsCompilation yes

Repository CRAN

Date/Publication 2017-07-15 21:00:01 UTC

R topics documented:

DrImpute *Imputing dropout events in single-cell RNA-sequencing data.*

Description

Imputing dropout events in single-cell RNA-sequencing data.

Usage

```
DrImpute(X, ks = 10:15, dists = c("spearman", "pearson"), method = "mean",
  cls = NULL, seed = 1, zerop = 0)
```

Arguments

X	Gene expression matrix (gene by cell).
ks	Number of cell clustering groups. Default set to ks = 10:15.
dists	Distribution matrices to use. Default is set to c("spearman", "pearson"). "euclediadian" can be added as well.
method	Use "mean" for mean imputation, "med" for median imputation.
cls	User can manually provide clustering information. Using different base clusterings. each row represent different clusterings. each column represent each cell.
seed	User can provide a seed.
zerop	zero percentage of resulting imputation is at least zerop.

Value

Imputed Gene expression matrix (gene by cell).

Author(s)

Il-Youp Kwak

References

Il-Youp Kwak, Wuming Gong, Kaoko Koyano-Nakagawa and Daniel J. Garry (2017+) DrImpute: Imputing dropout eveents in single cell RNA sequencing data

Examples

```
data(exdata)
exdata <- preprocessSC(exdata)
exdata <- exdata[1:3000, 1:80]
logdat <- log(exdata+1)
cls <- getCls(logdat)
logdat_imp <- DrImpute(logdat, cls = cls)
```

exdata *Usoskin data*

Description

This data set is subset from Usoskin et al. (2015). Original data is RNA-seq data on 799 cells dissected from the mouse lumbar dorsal root ganglion distributed over a total of nine 96-well plates. We randomly selected 150 cells from the data.

Column names indicate four different cell types, NF, NP, TH, and PEP.

Usage

```
data(exdata)
```

References

Usoskin D et al. Unbiased classification of sensory neuron types by large-scale single-cell RNA sequencing. Nature Neuroscience. Nature Research,2015;18:145-53.

Examples

```
data(exdata)
exdata <- preprocessSC(exdata)
```

getCls *get base clustering results using SC3 based clustering methods.*

Description

Similarity matrix constructed using "pearson", "spearman" or "euclidean". K-means clustering is performed on first few number of principal components of similarity matrix.

Usage

```
getCls(X, ks = 10:15, dists = c("spearman", "pearson"),
  dim.reduc.prop = 0.05)
```

Arguments

X	Log transformed gene expression matrix (Gene by Cell).
ks	Number of cell clustering groups. Default set to ks = 10:15.
dists	Distribution matrices to use. Default is set to c("spearman", "pearson"). "euclidean" can be added as well.
dim.reduc.prop	Proportion of principal components to use for K-means clustering.

Value

A matrix object, Each row represent different clustering results.

Author(s)

Il-Youp Kwak

References

Il-Youp Kwak, Wuming Gong, Kaoko Koyano-Nakagawa and Daniel J. Garry (2017+) DrImpute: Imputing dropout eveents in single cell RNA sequencing data

See Also

DrImpute preprocessSC

Examples

```
data(exdata)
exdata <- preprocessSC(exdata)
exdata <- exdata[1:3000, 1:80]
logdat <- log(exdata+1)
cls <- getCls(logdat)
```

---

preprocessSC *A function for preprocessing gene expression matrix.*

Description

Preprocess gene expression data

Usage

```
preprocessSC(X, min.expressed.gene = 0, min.expressed.cell = 2,
  max.expressed.ratio = 1, normalize.by.size.effect = FALSE)
```

Arguments

X	Gene expression matrix (Gene by Cell).
min.expressed.gene	Cell level filtering criteria. For a given cell, if the number of expressed genes are less than min.expressed.gene, we filter it out.
min.expressed.cell	Gene level filtering criteria. For a given gene, if the number of expressed cells are less than min.expressed.cell, we filter it out.

`max.expressed.ratio`
: Gene level filtering criteria. For a given gene, if the ratio of expressed cells are larger than max.expressed.ratio, we filter it out.

`normalize.by.size.effect`
: Normaize using size factor.

Value

Filtered gene expression matrix

Author(s)

Wuming Gong

References

Il-Youp Kwak, Wuming Gong, Kaoko Koyano-Nakagawa and Daniel J. Garry (2017+) DrImpute: Imputing dropout eveents in single cell RNA sequencing data

See Also

`DrImpute`

Examples

```
data(exdata)
exdata <- preprocessSC(exdata)
```

Index