

Linnorm Package

March 8, 2016

Type Package

Title Linear model and normality based normalization method (Linnorm)

Version 1.0

Date 2016-03-08

Maintainer Ken Shun Hang Yip <shunyip@bu.edu>

Description Normalization method for RNA-seq Expression data or large scale count data for parametric tests.

License MIT + file LICENSE

Imports Rcpp (>= 0.12.2),
RcppArmadillo,
MASS

LinkingTo Rcpp, RcppArmadillo, MASS

RoxygenNote 5.0.1

R topics documented:

GammaSim	1
Linnorm	2
LogNormSim	3
NBSim	4
PoissonSim	5

Index	7
--------------	----------

GammaSim	<i>This function simulates a RNA-seq dataset with the Gamma distribution based on the input dataset.</i>
----------	--

Description

This function simulates a RNA-seq dataset with the Gamma distribution based on the input dataset.

Usage

```
GammaSim(thisdata, NumRep = 3, NumDiff = 5000, NumFea = 20000,
  showinfo = FALSE, MaxLibSizeLog2FC = 0.5)
```

Arguments

thisdata	Matrix: The matrix or data frame that contains your dataset. Each row is a gene and each column is a replicate. Undefined values such as NA are not supported. This program assumes that all columns are replicates of the same sample.
NumRep	Integer: The number of replicates. This is half of the number of output samples. Defaults to 3.
NumDiff	Integer: The number of Differentially Changed Features. Defaults to 5000.
NumFea	Integer: The number of Total Features. Defaults to 20000.
showinfo	Logical: should we show data information on the console? Defaults to FALSE.
MaxLibSizeLog2FC	Double: The maximum library size difference from the mean that is allowed, in terms of log 2 fold change. Set to 0 to prevent program from generating library size differences. Defaults to 0.5.

Value

This function returns a list that contains a matrix of count data in integer raw count and a vector that indicates the genes that are differentially expressed. In the matrix, each row is a gene and each column is a replicate. The first NumRep (see parameter) of the columns belong to sample 1, and the last NumRep (see parameter) of the columns belong to sample 2. There will be NumFea (see parameter) number of columns.

Examples

```
simulateddata <- GammaSim(expMatrix)
```

Linnorm

Linnorm Function

Description

This function performs the Linear model and normality based normalization method (Linnorm) for RNA-seq expression data or large scale count data.

Usage

```
Linnorm(datamatrix, showinfo = FALSE, method = "default",
  perturbation = 10, minZeroPortion = 2/3)
```

Arguments

datamatrix	The matrix or data frame that contains your dataset. Each row is a feature (or Gene) and each column is a sample (or replicate). Undefined values such as NA are not supported.
method	"default" or "lambda" The program will output the transformed matrix if the method is "default". If the method is "lambda", the program will output a lambda value ($\text{normalizedExp} = \log_{1p}(\text{lambda} * \text{Matrix})$).
perturbation	integer ≥ 2 . To search for an optimal minimal deviation parameter (please see the article), Linnorm uses the iterated local search algorithm which perturbs away from the initial local minimum. The range of the area searched in each perturbation is exponentially increased as the area get further away from the initial local minimum, which is determined by their index. This range is calculated by $10 * (\text{perturbation}^{\text{index}})$.
minZeroPortion	double $\geq 0, \leq 1$. Features without at least this portion of non-zero values will not be used in the calculation of normalizing parameter. Defaults to 2/3.

Value

This function returns either a transformed data matrix or a lambda value. Please see the "method" parameter.

Examples

```
normalizedExp <- Linnorm(expMatrix, showinfo = "TRUE")
```

LogNormSim	<i>This function simulates a RNA-seq dataset with the Log Normal distribution based on the input dataset.</i>
------------	---

Description

This function simulates a RNA-seq dataset with the Log Normal distribution based on the input dataset.

Usage

```
LogNormSim(thisdata, NumRep = 3, NumDiff = 5000, NumFea = 20000,
  showinfo = FALSE, MaxLibSizeLog2FC = 0.5)
```

Arguments

thisdata	Matrix: The matrix or data frame that contains your dataset. Each row is a gene and each column is a replicate. Undefined values such as NA are not supported. This program assumes that all columns are replicates of the same sample.
NumRep	Integer: The number of replicates. This is half of the number of output samples. Defaults to 3.

NumDiff	Integer: The number of Differentially Changed Features. Defaults to 5000.
NumFea	Integer: The number of Total Features. Defaults to 20000.
showinfo	Logical: should we show data information on the console? Defaults to FALSE.
MaxLibSize _{log2FC}	Double: The maximum library size difference from the mean that is allowed, in terms of log 2 fold change. Set to 0 to prevent program from generating library size differences. Defaults to 0.5.

Value

This function returns a list that contains a matrix of count data in integer raw count and a vector that indicates the genes that are differentially expressed. In the matrix, each row is a gene and each column is a replicate. The first NumRep (see parameter) of the columns belong to sample 1, and the last NumRep (see parameter) of the columns belong to sample 2. There will be NumFea (see parameter) number of columns.

Examples

```
simulateddata <- LogNormSim(expMatrix)
```

NBSim	<i>This function simulates a RNA-seq dataset with the Negative Binomial distribution based on the input dataset.</i>
-------	--

Description

This function simulates a RNA-seq dataset with the Negative Binomial distribution based on the input dataset.

Usage

```
NBSim(thisdata, NumRep = 3, NumDiff = 5000, NumFea = 20000,
      showinfo = FALSE, MaxLibSizelog2FC = 0.5)
```

Arguments

thisdata	Matrix: The matrix or data frame that contains your dataset. Each row is a gene and each column is a replicate. Undefined values such as NA are not supported. This program assumes that all columns are replicates of the same sample.
NumRep	Integer: The number of replicates. This is half of the number of output samples. Defaults to 3.
NumDiff	Integer: The number of Differentially Changed Features. Defaults to 5000.
NumFea	Integer: The number of Total Features. Defaults to 20000.
showinfo	Logical: should we show data information on the console? Defaults to FALSE.
MaxLibSize _{log2FC}	Double: The maximum library size difference from the mean that is allowed, in terms of log 2 fold change. Set to 0 to prevent program from generating library size differences. Defaults to 0.5.

Value

This function returns a list that contains a matrix of count data in integer raw count and a vector that indicates the genes that are differentially expressed. In the matrix, each row is a gene and each column is a replicate. The first NumRep (see parameter) of the columns belong to sample 1, and the last NumRep (see parameter) of the columns belong to sample 2. There will be NumFea (see parameter) number of columns.

Examples

```
simulateddata <- NBSim(expMatrix)
```

PoissonSim	<i>This function simulates a RNA-seq dataset with the Poisson distribution based on the input dataset.</i>
------------	--

Description

This function simulates a RNA-seq dataset with the Poisson distribution based on the input dataset.

Usage

```
PoissonSim(thisdata, NumRep = 3, NumDiff = 5000, NumFea = 20000,
  showinfo = FALSE, MaxLibSizeLog2FC = 0.5)
```

Arguments

thisdata	Matrix: The matrix or data frame that contains your dataset. Each row is a gene and each column is a replicate. Undefined values such as NA are not supported. This program assumes that all columns are replicates of the same sample.
NumRep	Integer: The number of replicates. This is half of the number of output samples. Defaults to 3.
NumDiff	Integer: The number of Differentially Changed Features. Defaults to 5000.
NumFea	Integer: The number of Total Features. Defaults to 20000.
showinfo	Logical: should we show data information on the console? Defaults to FALSE.
MaxLibSizeLog2FC	Double: The maximum library size difference from the mean that is allowed, in terms of log 2 fold change. Set to 0 to prevent program from generating library size differences. Defaults to 0.5.

Value

This function returns a list that contains a matrix of count data in integer raw count and a vector that indicates the genes that are differentially expressed. In the matrix, each row is a gene and each column is a replicate. The first NumRep (see parameter) of the columns belong to sample 1, and the last NumRep (see parameter) of the columns belong to sample 2. There will be NumFea (see parameter) number of columns.

Examples

```
simulateddata <- PoissonSim(expMatrix)
```

Index

- *Topic **Binomial**
 - NBSim, 4
- *Topic **CPM**
 - Linnorm, 2
- *Topic **Count**
 - GammaSim, 1
 - Linnorm, 2
 - LogNormSim, 3
 - NBSim, 4
 - PoissonSim, 5
- *Topic **Expression**
 - GammaSim, 1
 - Linnorm, 2
 - LogNormSim, 3
 - NBSim, 4
 - PoissonSim, 5
- *Topic **FPKM**
 - Linnorm, 2
- *Topic **Gamma**
 - GammaSim, 1
- *Topic **Linnorm**
 - Linnorm, 2
- *Topic **Log**
 - LogNormSim, 3
- *Topic **Negative**
 - NBSim, 4
- *Topic **Normal**
 - LogNormSim, 3
- *Topic **Parametric**
 - Linnorm, 2
- *Topic **Poisson**
 - PoissonSim, 5
- *Topic **RNA-seq**
 - GammaSim, 1
 - Linnorm, 2
 - LogNormSim, 3
 - NBSim, 4
 - PoissonSim, 5
- *Topic **RPKM**
 - Linnorm, 2
- *Topic **Raw**
 - GammaSim, 1
 - Linnorm, 2
 - LogNormSim, 3
 - NBSim, 4
 - PoissonSim, 5
- *Topic **Simulation**
 - GammaSim, 1
 - LogNormSim, 3
 - NBSim, 4
 - PoissonSim, 5
- *Topic **TPM**
 - Linnorm, 2
- *Topic **distribution**
 - GammaSim, 1
 - LogNormSim, 3
 - NBSim, 4
 - PoissonSim, 5
- *Topic **normalization**
 - Linnorm, 2
- *Topic **transformation**
 - Linnorm, 2
- GammaSim, 1
- Linnorm, 2
- LogNormSim, 3
- NBSim, 4
- PoissonSim, 5