Package 'DFnorm'

August 10, 2017

Type Package

Title Fnorm A distribution-free normalization method for RNA-seq data

Version 1.0

Date 2017-08-09

Author Shengping Yang and Jianrong Wu

Maintainer Shengping Yang <shengping.yang@ttuhsc.edu>

Depends GMD, mclust, Hmisc

Description A non-parametric approach for RNA-seq data normalization.

License GPL (>= 3)

R topics documented:

DFnorm-package   *DFnorm*

Description

DFnorm assumes that highly correlated genes are those consistently expressed across subjects. By using a Spearman correlation, diseased subjects can be classified into different subtypes, and then normalized, which accommodates studies with data collected from a heterogeneous population.

Details

A distribution-free method for RNA-seq data normalization.

author Shengping Yang and Jianrong Wu

Maintainer: Shengping Yang <shengping.yang@ttuhsc.edu>

---

cf.fun   *A function for picking the reference genes.*

Description

Genes can be used as reference genes are identified by using this function.

Usage

```
cf.fun(x, c.t, inc)
```

Arguments

x	A vector of gene expression for one subject.
c.t	A vector of mean or median gene expression
inc	A vector of indicators that define whether a gene has met the minimally required read counts.

Value

A vector of indicators whether a gene will be used as a reference gene.

Examples

```
#cf.fun(x, c.t, inc)
```

---

cluster.k   *A function to find the optimal number of clusters.*

Description

Find the number of optimal clusters.

Usage

```
cluster.k(cdata, ev.thres)
```

Arguments

cdata	A data matrix to be clustered.
ev.thres	The threshold to declare a cluster.

Value

A value, which is the optimal number of clusters.

Examples

```
#cluster.k(cdata, ev.thres)
```

DFnorm   *DFnorm A distribution free normalization method for RNAseq data*

Description

DFnorm assumes that highly correlated genes are those consistently expressed across subjects. By using a Spearman correlation, diseased subjects can be classified into different subtypes, and then normalized, which accommodates studies with data collected from a heterogeneous population.

Usage

```
DFnorm(raw0, TNstatus, cor, kcut, ev.thres, both.tumor, inc.thres)
```

Arguments

raw0	The raw RNA-seq data matrix
TNstatus	A vector indicating disease and normal status, or grouping ids of two dieased groups.
cor	The Spearman correlation matrix, corresponding to the raw data. If not supplied, DFnorm will generate the correlation matrix itself, however, it might take a while depending on the size of the data.
kcut	The number of disease subtypes. It can be a number of known subtypes in the diseased group. If not given, then DFnorm will estimate the best numbers of subtypes.
ev.thres	The threshold for defining disease subtypes.
both.tumor	An indicator of whether both groups are diseased subjects.
inc.thres	The minimal read count number of a gene to be included as a reference gene in the normalization process.

Value

A list with one vector of grouping ids for the normalized data, and one matrix of normalized RNA-seq data.

Examples

```
library("GMD")
#raw0=matrix(rnorm(100000), ncol=100)
#tn=c(rep(1,50), rep(0,50))
#DF=DFnorm(raw0, tn, both.tumor="no")

#normalized.data=DF[[2]]
#new.tn=DF[[1]]
```

F.subtype   *A function to find the largest subtype.*

Description

Find the largest subtype.

Usage

```
F.subtype(tcor, kcut)
```

Arguments

tcor   Transposed correlation matrix.

kcut   The number of clusters.

Value

A vector of names of subject ids in the largest subtype.

Examples

```
#F.subtype(tcor, kcut)
```

---

main.subtype   *A function to find the main subtypes in the diseased group(s).*

Description

Find the largest subtype.

Usage

```
main.subtype(tcor, ev.thres, kcut)
```

Arguments

tcor   Transposed correlation matrix.

ev.thres   The threshold for declaring a cluster.

kcut   The number of clusters.

Value

A vector of ids that belong to the main subtype.

Examples

```
#main.subtype(tcor, ev.thres, kcut)
```

---

mode *A function to find the mode of a distribution.*

Description

Find the mode of a distribution.

Usage

```
mode(x, weigh)
```

Arguments

x   A vector of data from a distribution.

weigh   Weight of each value in the data.

Value

The value of mode.

Examples

```
#mode(x, weigh)
```

---

norm.ck *A function for normalizing data in the control group.*

Description

Normalizing RNA-seq data for the normal control group.

Usage

```
norm.ck(acc, mcsum, inc.thres)
```

Arguments

acc   Raw RNA-seq data of the control group.

mcsum   A vector of mdian gene expression values.

inc.thres   The minimum number of read counts for a genes to be used in the analysis.

Value

A matrix of normalized RNA-seq data.

Examples

```
#norm.ck(acc, mcsum, inc.thres)
```

norm.tumor | *A function for normalizing RNA-seq data for the diseased group.*

Description

Normalizing RNA-seq data for the diseased group.

Usage

```
norm.tumor(att, tcor, ev.thres, kcut, inc.thres)
```

Arguments

att   Raw RNA-seq read counts.

tcor   Transposed correlation matrix.

ev.thres   The threshold for declaring a cluster.

kcut   The number of subgroups. This value can be assigned manually.

inc.thres   The minimum number of read counts for a gene to be used in the analysis.

Value

A matrix of normalized RNA-seq data.

Examples

```
#norm.tumor(att, tcor, ev.thres, kcut, inc.thres)
```

Index