Package 'InfiniumPurify'

January 14, 2017

Type Package

Title Estimate and Account for Tumor Purity in Cancer Methylation Data Analysis

Version 1.3.1

Date 2017-1-8

Author Yufang Qin

Maintainer Yufang Qin <yfqin@shou.edu.cn>

Depends matrixStats

Description The proportion of cancer cells in solid tumor sample, known as the tumor purity, has adverse impact on a variety of data analyses if not properly accounted for. We develop 'InfiniumPurify', which is a comprehensive R package for estimating and accounting for tumor purity based on DNA methylation Infinium 450k array data. 'InfiniumPurify' provides functionalities for tumor purity estimation. In addition, it can perform differential methylation detection and tumor sample clustering with the consideration of tumor purities.

License GPL-2

NeedsCompilation no

Repository CRAN

Date/Publication 2017-01-14 12:12:25

R topics documented:

abbr *abbr*

Description

This data set lists abbreviations for all TCGA cancer types.

Usage

```
abbr
```

Format

A dataframe containing names and abbreviations for all TCGA cancer types.

Source

X. Zheng, N. Zhang, H.J. Wu and H. Wu, Estimating and accounting for tumor purity in the analysis of DNA methylation data from cancer studies. *Genome biology*, accepted.

beta.emp *beta.emp*

Description

An example data set for InfiniumClust and InfiniumPurify.

Usage

```
beta.emp
```

Format

A dataframe containing methylaton beta values for 62 tumor and normal samples.

Source

X. Zheng, N. Zhang, H.J. Wu and H. Wu, Estimating and accounting for tumor purity in the analysis of DNA methylation data from cancer studies. *Genome biology*, accepted.

CancerTypeAbbr *Print abbreviations of cancer types with known iDMCs.*

Description

Print tumor types and their abbreviations with known informative DMCs.

Usage

```
CancerTypeAbbr()
```

Arguments

None.

Author(s)

Xiaoqi Zheng <xqzheng@shnu.edu.cn>.

References

X. Zheng, N. Zhang, H.J. Wu and H. Wu, Estimating and accounting for tumor purity in the analysis of DNA methylation data from cancer studies. *Genome biology*, in revision.

Examples

```
data(abbr)
CancerTypeAbbr()
```

getPurity *Estimate the tumor purity for 450K methylation data*

Description

Estimate the percentage of tumor cells in cancer samples which are mixtures of cancer and normal cells.added a sentence

Usage

```
getPurity(tumor.data,normal.data = NULL,tumor.type = NULL)
```

Arguments

tumor.data	numeric vector/matrix of beta values for tumor samlpes. The names/rownames of tumor.data should be probe names of Infinium 450k array, and colnames should be names of tumor samples.
normal.data	numeric matrix of beta values for normal samlpes. The rownames of normal.data should be probe names of Infinium 450k array, and colnames should be names of normal samples.
tumor.type	cancer type (in abbreviation) of tumor and normal samlpes. Options are "LUAD", "BRCA" and so on. See CancerTypeAbbr for detail.

Details

Arguments normal.data and tumor.type could be null. If either the number of tumor samples or number of normal smaples is less than 20, the tumor.type argument should be specified according to CancerTypeAbbr. If the numbers of tumor and normal samples are both more than 20, tumor.type could be null. In such case, getPurity first identify 1000 iDMCs by Wilcox rank-sum test, then tumor purity for each sample is estimated as the density mode of adjusted methylation levels of iDMCs.

Value

A vector of tumor purities for each tumor sample.

Author(s)

Xiaoqi Zheng <xqzheng@shnu.edu.cn>.

References

N. Zhang, H.J. Wu, W. Zhang, J. Wang, H. Wu and X. Zheng (2015) Predicting tumor purity from methylation microarray data. *Bioinformatics* **31(21)**, 3401-3405.

X. Zheng, N. Zhang, H.J. Wu and H. Wu, Estimating and accounting for tumor purity in the analysis of DNA methylation data from cancer studies. *Genome biology*, accepted.

Examples

```
## load example data
data(beta.emp)

normal.data <- beta.emp[,1:21]
tumor.data <- beta.emp[,22:61]

## call purity for single tumor sample
purity <- getPurity(tumor.data = tumor.data[,1],normal.data = NULL,tumor.type= "LUAD")

## call purity for less than 20 tumor samples
purity <- getPurity(tumor.data = tumor.data[,1:10],normal.data = NULL,tumor.type= "LUAD")

## call purity for more than 20 tumor samples with matched normal samples
purity <- getPurity(tumor.data = tumor.data[,1:40],normal.data = normal.data)
```

iDMC *iDMC*

Description

This data set lists pre-selected iDMCs for all TCGA cancer types.

Usage

```
iDMC
```

Format

A list containing informative Differential methylation CpG sites (iDMC) and their average methylation levels in tumor and normal samples.

Source

X. Zheng, N. Zhang, H.J. Wu and H. Wu, Estimating and accounting for tumor purity in the analysis of DNA methylation data from cancer studies. *Genome biology*, accepted.

InfiniumClust *Tumor sample clustering from Infinium 450k array data*

Description

Clustering of tumor samples into subtypes accounting for tumor purity.

Usage

```
InfiniumClust(tumor.data, purity, K, maxiter = 100, tol = 0.001)
```

Arguments

tumor.data	numeric matrix of beta values for tumor samlpes. The rownames of tumor.data should be probe names of Infinium 450k array, and colnames should be names of tumor samples.
purity	purities for tumor samples. Could be estimated by getPurity, or user specified purities from other tools.
K	the number of clusters.
maxiter	the maximum number of iterations allowed. Default is 100.
tol	tolerance for convergence of EM iterations. Default is 0.001.

Details

An EM based statistical method for subtype classification based on DNA methylation data, while adjusting for tumor purity.

Value

InfiniumClust returns a list consisting oflikelihood `tol.ll` and membership matrix `Z`.

`tol.ll`	total log-likelihood of converged EM algorithm.
`Z`	the membership matrix, where row corresponds to tumor samples and column corresponds to K clusters.

Author(s)

Xiaoqi Zheng <xqzheng@shnu.edu.cn> and Hao Wu <hao.wu@emory.edu>

References

W. Zhang, H. Feng, H. Wu and X. Zheng (2016). Tumor purity improves cancer subtype classification from DNA methylation data. Submitted.

Examples

```
## load example data
data(beta.emp)
normal.data <- beta.emp[,1:21]
tumor.data <- beta.emp[,22:31]

## estimate tumor purity
purity <- getPurity(tumor.data = tumor.data,tumor.type= "LUAD")

## cluster tumor samples accounting for tumor purity
out <- InfiniumClust(tumor.data,purity,K=3, maxiter=5, tol=0.001)
```

`InfiniumDMC` *Differentially Methylation Calling accounting for tumor purity*

Description

Infer differentially methylated CpG sites with the consideration of tumor purities.

Usage

```
InfiniumDMC(tumor.data,normal.data,purity,threshold)
```

Arguments

tumor.data	numeric matrix of beta values for tumor samlpes. The rownames of tumor.data should be probe names of Infinium 450k array, and colnames should be names of tumor samples.
normal.data	numeric matrix of beta values for normal samlpes. The rownames of normal.data should be probe names of Infinium 450k array, and colnames should be names of normal samples.
purity	purities for tumor samples. Could be estimated by getPurity, or user specified purities from other tools.
threshold	probability threshold in control-free DM calling. Default is 0.1.

Details

If normal.data is provided, the function tests each CpG site for differential methylation between tumor and normal samples with the consideration of tumor purities by a generalized linear regression. If normal.data is not provided, the function computes posterior probability to rank CpG sites.

Value

A data frame of statistics, p-values and q-values for all CpG sites.

Author(s)

Xiaoqi Zheng <xqzheng@shnu.edu.cn>.

References

X. Zheng, N. Zhang, H.J. Wu and H. Wu, Estimating and accounting for tumor purity in the analysis of DNA methylation data from cancer studies. *Genome biology*, in revision.

See Also

dmpFinder in the **minfi** package.

Examples

```
## load example data
data(beta.emp)

normal.data <- beta.emp[,1:21]
tumor.data <- beta.emp[,22:61]

## estimate tumor purity
purity <- getPurity(tumor.data = tumor.data,normal.data = normal.data)

## DM calling with normal controls
DMC = InfiniumDMC(tumor.data = tumor.data,normal.data = normal.data,purity = purity)
```

```
## DM calling without normal control
DMC_ctlFree = InfiniumDMC(tumor.data = tumor.data,purity = purity)
```

`InfiniumPurify` *Purify tumor methylomes caused by normal cell contamination.*

Description

Deconvolute purified tumor methylomes accounting for tumor purity.

Usage

```
InfiniumPurify(tumor.data,normal.data,purity)
```

Arguments

`tumor.data`	numeric matrix of beta values for tumor samlpes. The rownames of tumor.data should be probe names of Infinium 450k array, and colnames should be names of tumor samples.
`normal.data`	numeric matrix of beta values for normal samlpes. The rownames of normal.data should be probe names of Infinium 450k array, and colnames should be names of normal samples.
`purity`	purities for tumor samples. Could be estimated by getPurity, or user specified purities from other tools.

Details

The function deconvolutes purified tumor methylomes by a linear regression model.

Value

A matrix of purified beta values for all CpG sites (row) and tumor samples (column).

Author(s)

Xiaoqi Zheng <xqzheng@shnu.edu.cn>.

References

X. Zheng, N. Zhang, H.J. Wu and H. Wu, Estimating and accounting for tumor purity in the analysis of DNA methylation data from cancer studies. *Genome biology*, accepted.

Examples

```
## load example data
data(beta.emp)

normal.data <- beta.emp[,1:21]
tumor.data <- beta.emp[,22:61]

## estimate tumor purity
purity <- getPurity(tumor.data = tumor.data,normal.data = NULL,tumor.type= "LUAD")

## correct tumor methylome by tumor purity
tumor.purified = InfiniumPurify(tumor.data = tumor.data[1:100,],
                                normal.data = normal.data[1:100,],
                                purity = purity)
```

Index