Package 'PREDE'

May 6, 2020

Type Package

Title Deconvolution of heterogeneous tumor samples using partial reference signals

Version 1.2.1

Author Xiaoqi Zheng

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

Depends R (>= 3.5.0), matrixStats, stats, utils, quadprog, gtools

Description Characterizing the tumor heterogeneity, including the molecular profile of each cell population and its proportion in tumor tissues are very important. The present methods are roughly divided into two categories: reference-based methods and reference-free methods. However, in real clinical practices, the deconvolution problems are neither reference-based nor reference-free, but consist of a fraction of known cell types while leaving the rest cell types unknown. Therefore we proposed a partial-reference based deconvolution (PREDE) model based on Non-negative matrix factorization. We performed comprehensive evaluation of the proposed and ongoing deconvolution methods using simulated data and real TCGA tumor samples. We found that our PREDE model could effectively recover the true expression profile and absolute proportion of available and unknown cell types.

License GPL-2

NeedsCompilation no

R topics documented:

`generate_bulk` *generate the mixture samples*

Description

generate the mixture samples based on the profile matrix of cell types and the proportion of cell types.

Usage

```
generate_bulk(celltypes,nSample =100,csd = 0.1)
```

Arguments

celltypes	gene expression profiles or methylation profiles of cell type.
nSample	the number of generated mixture samples.
csd	different levels of noise.

Details

In simulation studies, some gene expression profiles of cancer cell lines were selected as basis matrix, the proportion matrix H is randomly generated under the Dirichlet distribution. The gene expression matrix of mixture samples can be obtained by multiplying W and H matrices, followed by an additional error matrix of normal distribution with mean equal to zero and different levels of noise.

Value

The function generate_bulk returns a list containing the following components:

Y	the profile matrix of the mixture samples.
W	the profile matrix of the cell types.
H	the proportion matrix of the cell types.

Author(s)

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

References

Y. Qin, W. Zhang, S. Nan, N. Wei and X. Zheng (2020). Deconvolution of heterogeneous tumor samples using partial reference signals. Submitted.

Examples

```
## load example data
data(lung_exp)
W <- lung_exp[,1:6]

## generate the mixed samples based on the profile matrix of the cell types W
bulk <- generate_bulk(W,nSample =100,csd = 0.1)
```

GetCelltypeNum *get optimal number of total cell types in the mixture tumor sample*

Description

get the optimal number of total cell types by computing AIC value.

Usage

```
GetCelltypeNum(Y,W = NULL,W1 = NULL,maxK = 50)
```

Arguments

Y	matrix of profiles of mixture samples, which can be gene expression proflies,methylation profile, etc.
W	matrix of the profiles of cell types.
W1	matrix/vector of the profiles of partial reference cell types.
maxK	an upper bound of the number of total cell types.

Details

Before the deconvolution of mixture sample into distinct cell populations based on the partial reference, the number of total cell types should be specified. We obtain the optimal number of total cell types by computing the Akaike information criterion (AIC). Given the range of K, we get a series of values of AIC and take the number of total cell types with minimum aic as the optimal number of total cell types.

Value

a vector for AIC values under different numbers of total cell types, and a value for optimal number of total cell types .

Author(s)

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

References

Y. Qin, W. Zhang, S. Nan, N. Wei and X. Zheng (2020). Deconvolution of heterogeneous tumor samples using partial reference signals. Submitted.

Examples

```
## load example data
data(lung_exp)
W <- lung_exp[,1:6]

## generate the mixed samples
bulk <- generate_bulk(W,nSample =100,csd = 0.1)

## select the feature
```

```
feat <- select_feature(mat = bulk$Y,method = "cv",nmarker = 1000,startn = 0)

## get optimal number of total cell types
OptimalK <- GetCelltypeNum(bulk$Y[feat,],W=NULL,W1=W[feat,1:4],maxK = 10)
```

PREDE *Partial-reference based deconvolution model*

Description

Estimate the profiles of all cell types and the proportion of cell types in tumor samples

Usage

```
PREDE(Y,W = NULL,W1,type = "GE",K,iters = 500,rssDiffStop=1e-10)
```

Arguments

Y	matrix of profiles of tumor samples, which can be gene expression proflies,methylation profile, etc.
W	the profile matrix of cell types.
W1	matrix/vector of the profiles of partial reference cell types.
type	type of tumor sample data. Options are "GE","ME", etc.
K	the number of total cell types.
iters	the maximum number of iterations to execute.
rssDiffStop	the threshold of the interations terminate. The iterations end when either the maximum number of gnerations has been reached or the absolute error is less than the given threshold 'rssDiffStop'.

Details

Partial-reference based decomposition model could estimate the profiles of unknown cell types and the proportions of cell types. The number of total cell types K should be specified when running the function PREDE. The optimal value of K could be determined by the output value AIC of the function PREDE. If the profile of cell types W is non-empty, we can use adjustWH to adjust the order of predicted cell types.

Value

a matrix for profiles of cell types and a matrix for the proportions of cell types.

Author(s)

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

References

Y. Qin, W. Zhang, S. Nan, N. Wei and X. Zheng (2020). Deconvolution of heterogeneous tumor samples using partial reference signals. Submitted.

Examples

```
## load example data
data(lung_exp)
W <- lung_exp[,1:6]

## Partial-reference based deconvolution without total cell types W
bulk <- generate_bulk(W,nSample =100,csd = 0.1)
feat <- select_feature(mat = bulk$Y,method = "cv",nmarker = 1000,startn = 0)
PREDE(bulk$Y[feat,],W1=W[feat,1:4],type = "GE",K=7,iters = 100,rssDiffStop=1e-5)
```

`select_feature` *Select feature for profile matrix of mixed samples*

Description

Select feature for profile matrix of mixed samples before applying PREDE model

Usage

```
select_feature(mat,method,nmarker, startn=0)
```

Arguments

`mat`	matrix of profiles of mixed samples.
`method`	method of selecting features. Options are "random","cv" and "topvar".
`nmarker`	number of selecting features.
`startn`	startn+1 represents the position where the first feature starts.

Details

Since the total number of genes or CpG sites is huge compared with the number of samples and cell populations, it is necessary to select features. nmarker represents the number of selected features. startn+1 represents the position where the first feature starts.

Value

The row in which the selected feature are.

Author(s)

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

References

Y. Qin, W. Zhang, S. Nan, N. Wei and X. Zheng (2020). Deconvolution of heterogeneous tumor samples using partial reference signals. Submitted.

Examples

```
## load example data
data(lung_exp)
W <- lung_exp[,1:6]

## Partial-reference based deconvolution without total cell types W
bulk <- generate_bulk(W,nSample =100,csd = 0.1)
feat = select_feature(mat = bulk$Y,method = "cv",nmarker = 1000,startn = 0)
```

Index