Package ‘UNDO’

Type Package

Title Unsupervised Deconvolution of Tumor-Stromal Mixed Expressions

Version 1.14.0

Date 2014-07-17

Author Niya Wang <wangny@vt.edu>

Maintainer Niya Wang <wangny@vt.edu>

Depends R (>= 2.15.2), methods, BiocGenerics, Biobase

Imports MASS, boot, nnls, stats, utils

biocViews Software

Description UNDO is an R package for unsupervised deconvolution of tumor and stromal mixed expression data. It detects marker genes and deconvolutes the mixing expression data without any prior knowledge.

License GPL-2

NeedsCompilation no

R topics documented:

UNDO-package .............................................................. 2
BiologicalMixMCF7HS27 ............................................... 2
calc_E1 ................................................................. 3
dimension_reduction ................................................... 4
gene_expression_input ............................................... 5
marker_gene_selection ............................................... 5
mixing_matrix_computation ......................................... 6
NumericalMixingMatrix ................................................. 7
NumericalMixMCF7HS27 ............................................... 8
PureMCF7HS27 .......................................................... 9
two_source_deconv .................................................... 10

Index 11
**UNDO-package**  
*Implementation of UNDO (unsupervised deconvolution of tumor-stromal mixed expressions)*

---

**Description**

This package contains main function "two_source_deconv" to implement the deconvolution of mixed tumor-stromal expressions in a completely unsupervised way. The prior knowledge of mixing matrix or pure expression is not needed. The package detects marker genes and calculate the mixing matrix and pure expressions automatically.

**Details**

Package: UNDO  
Type: Package  
Version: 1.7.3  
Date: 2014-04-30  
License: GPL version 2 or later  

two_source_deconv(ExpressionData,lowper=0.4,highper=0.1,epsilon1=0.01,epsilon2=0.01,A=NULL,S1=NULL,S2=NULL,return=0)

**Author(s)**

Niya Wang <wangny@vt.edu>

**Examples**

data(NumericalMixMCF7HS27)

X <- NumericalMixMCF7HS27

deconvResult <- two_source_deconv(X, lowper = 0.4, highper = 0.1, epsilon1 = 0.1, epsilon2 = 0.1, A = NULL, S1=NULL,S2=NULL,return=0)

---

**BiologicalMixMCF7HS27  MCF7 and HS27 biologically mixed**

---

**Description**

Expression data from MCF7 and HS27 biologically mixing

**Usage**

data(BiologicalMixMCF7HS27)
calc_E1

function calculating the E1 measurement

Description

A function used to calculate the E1 measurement when the real mixing matrix is provided

Usage

calc_E1(A, Aest)

Arguments

A  
real mixing matrix

Aest  
estimated mixing matrix
Value

E1 measurement (numeric)

Author(s)

Niya Wang <wangny@vt.edu>

Examples

```r
A <- matrix(rnorm(10), 2, 2)
Aest <- matrix(rnorm(10), 2, 2)
E1 <- calc_E1(A, Aest)  # to calculate the similarity of two random 2x2 matrix
```

---

dimension_reduction  
*Dimension reduction function*

Description

When the number of input samples is larger than 2, this function is called to reduce the dimension to 2 by using PCA.

Usage

```r
dimension_reduction(X)
```

Arguments

- **X**
  - gene expression data matrix

Value

- **X**
  - **dimenMatrix**
    - the dimension reduction matrix used to recover the mixing matrix for all the samples

Author(s)

Niya Wang (wangny@vt.edu)

Examples

```r
X <- matrix(rnorm(5000), 1000, 5)
dimenResult <- dimension_reduction(X)
```
gene_expression_input

Detect whether the input gene expression data are valid

Description
Check the input gene expression data to see whether they are nonempty, nonnegative, etc.

Usage
gene_expression_input(x)

Arguments
x  gene expression data matrix with row representing genes/probe sets, and column representing samples.

Value
If the input is valid, the output will be the same as the input; otherwise, if the input contains NA, the corresponding rows will be deleted. If the input contains negative value, the algorithm will stop and give error information.

Author(s)
Niya Wang (wangny@vt.edu)

Examples
gene_expression <- matrix(runif(2000),1000,2)
valid_gene_expression <- gene_expression_input(gene_expression)

marker_gene_selection

Select marker genes in two sources

Description
Select the marker genes in tumor and stroma in an unsupervised way

Usage
marker_gene_selection(X, lowper, highper, epsilon1, epsilon2)
mixing_matrix_computation

Arguments

- **X**: gene expression data
- **lowper**: The percentage of genes the user wants to remove with lowest norm. The range should be between 0 and 1.
- **highper**: The percentage of genes the user wants to remove with highest norm. The range should be between 0 and 1.
- **epsilon1**: Influence the number of marker genes. With increasing of epsilon1, the number of marker genes in source 1 will increase. The value should be positive.
- **epsilon2**: Influence the number of marker genes. With increasing of epsilon1, the number of marker genes in source 2 will increase. The value should be positive.

Value

- **a1**: The slope of marker genes in source 1
- **a2**: The slope of marker genes in source 2
- **MG1**: The gene list of marker genes in source 1
- **MG2**: The gene list of marker genes in source 2
- **dimenMatrix**: dimension reduction matrix

Author(s)

Niya Wang (wangny@vt.edu)

Examples

```r
X <- matrix(runif(20000), 10000, 2)
MG_set <- marker_gene_selection(X, 0.4, 0.1, 0.1, 0.1)
```

mixing_matrix_computation

*Calculate and scale the mixing matrix*

Description

Calculate the mixing matrix based on the output from marker_gene_selection(), and scale the mixing matrix to make the sum of proportions from tumor and stroma equal to 1. The pure expression levels of tumor and stroma are also computed.

Usage

```r
mixing_matrix_computation(X, a1, a2, dimenMatrix)
```
NumericalMixingMatrix

Arguments

X Gene expression data matrix
a1 The slope of marker genes in source 1
a2 The slope of marker genes in source 2
dimenMatrix The dimension reduction matrix used to recover mixing matrix for all the samples

Value

Aest estimated mixing matrix
Sest estimated pure gene expression of two sources

Author(s)

Niya Wang (wangny@vt.edu)

Examples

a1<- matrix(runif(2),2,1)
a2<- matrix(runif(2),2,1)
X <- 1000*matrix(runif(20000),10000,2)
dimenMatrix <- NULL
Deconv <- mixing_matrix_computation(X, a1, a2, dimenMatrix)

NumericalMixingMatrix  mixing matrix of data NumericalMixMCF7HS27

Description

real mixing matrix of data NumericalMixMCF7HS27

Usage

data(NumericalMixingMatrix)

Format

The format is: num [1:2, 1:2] 0.775 0.15 0.225 0.85 - attr(*, "dimnames")=List of 2 ..$ : NULL ..$
: chr [1:2] "V1" "V2"

Examples

data(NumericalMixingMatrix)
str(NumericalMixingMatrix)
**Description**

Expression data from MCF7 and HS27 numerically mixing

**Usage**

```r
data(NumericalMixMCF7HS27)
```

**Format**

The format is: Formal class 'ExpressionSet' [package "Biobase"] with 7 slots ..@ experimentData:
  Formal class 'MIAME' [package "Biobase"] with 13 slots ..@ name : chr 
  ..@ lab : chr 
  ..@ contact : chr 
  ..@ abstract : chr 
  ..@ pubMedIds : chr 
  ..@ samples : list() 
  ..@ preprocessings : list() 
  ..@ normControls : list() 
  ..@ other : list() 
  ..@ __classVersion__:Formal class 'Versions' [package "Biobase"] with 1 slots ..@ assayData: ..@ phenoData: ..@ featureData: ..@ annotation: ..@ protocolData: ..@ __classVersion__:Formal class 'Versions' [package "Biobase"] with 4 slots ..@ varMetadata: ..

**Examples**

```r
data(NumericalMixMCF7HS27)
str(NumericalMixMCF7HS27)
```
### Description

pure MCF7 and HS27 expression data

### Usage

```r
data(PureMCF7HS27)
```

### Format

The format is: Formal class 'ExpressionSet' [package "Biobase"] with 7 slots

- `.experimentData`: Formal class 'MIAME' [package "Biobase"] with 13 slots
  - `@ name`: chr 
  - `@ lab`: chr
  - `@ abstract`: chr
  - `@ pubMedIds`: chr
  - `@ samples`: list
  - `@ hybridizations`: list
  - `@ normControls`: list
  - `@ preprocessing`: list

- `. assayData`:<environment: 0x000000000e979d20>

- `. phenoData`: Formal class 'AnnotatedDataFrame' [package "Biobase"] with 4 slots

- `. featureData`: Formal class 'AnnotatedDataFrame' [package "Biobase"] with 4 slots

- `. annotation`: chr "HG-U133A"

### Examples

```r
data(PureMCF7HS27)
str(PureMCF7HS27)
```
two_source_deconv

Main function to call other subfunction to deconvolute the mixed expression data.

Description

This is the main function that is to call all the other subfunctions and realize the deconvolution of mixed expression data. When the real mixing matrix exist, it will also compare the estimated mixing matrix and real mixing matrix and give the E1 measurement.

Usage

two_source_deconv(ExpressionData, lowper = 0.4, highper = 0.1, epsilon1 = 0.01, epsilon2 = 0.01, A = NULL, S1 = NULL, S2 = NULL, return = 0)

Arguments

ExpressionData  gene expression data matrix/ExpressionSet object
lowper          The percentage of genes the user wants to remove with lowest norm. The range should be between 0 and 1.
highper         The percentage of genes the user wants to remove with highest norm. The range should be between 0 and 1.
epsilon1        Influence the number of marker genes. With increasing of epsilon1, the number marker genes in source 1 will increase. The value should be positive.
epsilon2        Influence the number of marker genes. With increasing of epsilon1, the number marker genes in source 2 will increase. The value should be positive.
A               real mixing matrix if existing
S1              Pure expression profile of first source if existing
S2              Pure expression profile of second source if existing
return          if it is equal to 0, do not return estimated S; otherwise, return the estimated S.

Value

Aest            estimated mixing matrix
E1              E1 measurement between real and estimated mixing matrix

Author(s)

Niya Wang (wangny@vt.edu)

Examples

data(NumericalMixMCF7HS27)
X <- NumericalMixMCF7HS27
deconvResult <- two_source_deconv(X, lowper = 0.4, highper = 0.1, epsilon1 = 0.1, epsilon2 = 0.1, A = NULL, S1 = NULL, S2 = NULL)
Index

*Topic datasets
  BiologicalMixMCF7HS27, 2
  NumericalMixingMatrix, 7
  NumericalMixMCF7HS27, 8
  PureMCF7HS27, 9
*Topic methods
  UNDO-package, 2
*Topic package
  UNDO-package, 2

BiologicalMixMCF7HS27, 2
calc_E1, 3
dimension_reduction, 4
gene_expression_input, 5
marker_gene_selection, 5
mixing_matrix_computation, 6
NumericalMixingMatrix, 7
NumericalMixMCF7HS27, 8
PureMCF7HS27, 9
two_source_deconv, 10
UNDO (UNDO-package), 2
UNDO-package, 2