

Package ‘madsim’

December 7, 2016

Type Package

Title A Flexible Microarray Data Simulation Model

Version 1.2.1

Date 2016-12-07

Author Doulaye Dembele

Maintainer Doulaye Dembele <doulaye@igbmc.fr>

Description This function allows to generate two biological conditions synthetic microarray dataset which has similar behavior to those currently observed with common platforms. User provides a subset of parameters. Available default parameters settings can be modified.

License GPL (>= 2.0)

NeedsCompilation no

Repository CRAN

Date/Publication 2016-12-07 15:54:18

R topics documented:

madsim-package	1
madsim	2
madsim_test	4

Index	6
--------------	----------

madsim-package	<i>A Flexible Microarray Data Simulation Model</i>
----------------	--

Description

madsim allows to generate two conditions biological synthetic microarray dataset which known characteristics. These data have similar behavior as those obtained with current microarray platforms.

Details

Package: madsim
Type: Package
Version: 1.2.1
Date: 2016-12-07
License: GPL (>=2.0)

This package has only one function

Author(s)

Doulaye Dembele Maintainer: Doulaye Dembele <doulaye@igbmc.fr>

References

Dembele D. (2013), A Flexible Microarray Data Simulation Model, *Microarrays*, 2013, 2(2):115-130

Examples

```
# set parameters settings
fparams <- data.frame(m1 = 7, m2 = 7, shape2 = 4, lb = 4, ub = 14, pde = 0.02, sym = 0.5);
dparams <- data.frame(lambda1 = 0.13, lambda2 = 2, muminde = 1, sdde = 0.5);
sdn <- 0.4; rseed <- 50;

# generate synthetic data without using real microarray data as seed
mydata <- madsim(mdata=NULL, n=10000, ratio=0, fparams, dparams, sdn, rseed);

# calculate MMplot variables using samples 1 and 12
A <- 0.5*(mydata$xdata[,12] + mydata$xdata[,1]);
M <- mydata$xdata[,12] - mydata$xdata[,1];

# draw MA plot using samples 1 and 12
plot(A,M)
```

madsim

Main function of the package "madsim"

Description

function madsim() allows to generate two biological conditions synthetic microarray dataset with known characteristics. These data have similar behavior as those obtained with current microarray platforms. Hence, they can be used for performance evaluation of data meta-analysis methods.

Usage

```
madsim(mdata = NULL, n = 10000, ratio = 0,
       fparams = data.frame(m1=7,m2=7,shape2=4,lb=4,ub=14,pde=0.02,sym=0.5),
       dparams = data.frame(lambda1=0.13, lambda2=2, muminde=1, sdde=0.5),
       sdn = 0.4, rseed = 50)
```

Arguments

<code>mdata</code>	a data frame with numerical values to be used as seed, its length should be greater than 100. When set to NULL (default) data generated are fully synthetic: <code>mdata = NULL</code>
<code>n</code>	an integer specifying the number of genes in the data generated: <code>n = 10000</code>
<code>ratio</code>	a flag (0,1) allowing to have log2 intensitie or log2 ratio: <code>ratio = 0</code>
<code>fparams</code>	a data frame containing 7 components defining the data lower (lb) and upper bound (ub), the beta distribution shape (shape2) parameter, the percentage of differentially expressed (pde) number of genes and the partition of the number of down and up regulated (sym) genes: <code>fparams=data.frame(m1=7,m2=7,shape2=2,lb=4,ub=14,pde=0.02,sym=0.5)</code>
<code>dparams</code>	a data frame containing 4 components defining how low and high expressed genes are distributed (lambda1), and how changes are for DE genes (lambda2, muminde, sdde): <code>dparams = data.frame(lambda1=0.13,lambda2=2,muminde=1,sdde=0.5)</code>
<code>sdn</code>	a positive scalar used as standard deviation for the additive gaussian noise: <code>sdn = 0.4</code>
<code>rseed</code>	an integer used as seed for generating random number by the computer in use: <code>rseed = 50</code>

Details

User provides a subset of parameters. A detailed description of these parameters is available in the reference given below. Default parameters settings (in arguments above) can be modified.

Value

Returned is a data frame containing 3 components

<code>xdata</code>	a dataset with sizes, the number of rows and columns, specified by input parameters <code>n</code> and <code>m1+m2</code> , respectively
<code>xid</code>	a vector of indexes with values are from the set (0, -1, 1). These values are used for non differentially expressed, down- and up-regulated genes
<code>xsd</code>	a scalar containing the standard deviation of first column of the dataset generated

Author(s)

Doulaye Dembele

References

Dembele D. (2013), A Flexible Microarray Data Simulation Model. *Microarrays*, 2013, 2(2):115-130

Examples

```

# load a sample of real microarray data
data(madsim_test)

# set parameters settings
mdata <- madsim_test$V1;
fparams <- data.frame(m1 = 7, m2 = 7, shape2 = 4, lb = 4, ub = 14,pde=0.02,sym=0.5);
dparams <- data.frame(lambda1 = 0.13, lambda2 = 2, muminde = 1, sdde = 0.5);
sdn <- 0.4; rseed <- 50;

# generate fully synthetic data
mydata1 <- madsim(mdata = NULL, n = 10000, ratio = 0, fparams, dparams, sdn, rseed);

# use true affymetrix data to generate synthetic data
mydata2 <- madsim(mdata = madsim_test, n=10000, ratio=0,fparams,dparams,sdn,rseed);

A1 <- 0.5*(mydata1$xdata[,12] + mydata1$xdata[,1]);
M1 <- mydata1$xdata[,12] - mydata1$xdata[,1];

A2 <- 0.5*(mydata2$xdata[,12] + mydata2$xdata[,1]);
M2 <- mydata2$xdata[,12] - mydata2$xdata[,1];

# draw MA plot using samples 1 and 12
op <- par(mfrow = c())
  plot(A1,M1)
  plot(A2,M2)
par(op)

```

madsim_test

A microarray data sample, one colomn numerical values

Description

A text file containing an example of real microarray which can be used as seed. This dataset is from a Affymetrix GeneChip array (Human Gene 1.0 ST)

Usage

```
data(madsim_test)
```

Format

A data frame with 33297 observations on the following variable.

V1 a numeric vector

Examples

```
# load a sample of real microarray data
data(madsim_test)

# set parameter settings
mdata <- madsim_test$V1;
fparams <- data.frame(m1=7, m2=7, shape2=4, lb=4, ub=14, pde=0.02, sym=0.5);
dparams <- data.frame(lambda1 = 0.13, lambda2 = 2, muminde = 1, sdde = 0.5);
sdn <- 0.4; rseed <- 50;

# generate data using microarray as seed
mydata <- madsim(mdata, n = 10000, ratio = 0, fparams, dparams, sdn, rseed);

# calculate Mplot variables using samples 1 and 12
A <- 0.5*(mydata$xdata[,12] + mydata$xdata[,1]);
M <- mydata$xdata[,12] - mydata$xdata[,1];

# draw Mplot representation using samples 1 and 12
plot(A,M)
```

Index

*Topic **package**

 madsim-package, [1](#)

madsim, [2](#)

madsim-package, [1](#)

madsim_test, [4](#)