

# Package: madsim (via r-universe)

October 17, 2024

**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

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

**Repository** <https://doulaye.r-universe.dev>

**RemoteUrl** <https://github.com/cran/madsim>

**RemoteRef** HEAD

**RemoteSha** 28cf3bf969dd5e34048f686da19455b9e6cc7cfb

## Contents

madsim-package . . . . .	2
madsim . . . . .	3
madsim_test . . . . .	5

<b>Index</b>	<b>6</b>
--------------	----------

## Description

madsim allows to generate two conditions biological synthetic microarray dataset with 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**

mdata	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: mdata = NULL
n	an integer specifying the number of genes in the data generated: n = 10000
ratio	a flag (0,1) allowing to have log2 intensitie or log2 ratio: ratio = 0
fparams	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: fparams=data.frame(m1=7,m2=7,shape2=2,lb=4,ub=14,pde=0.02,sym=0.5)
dparams	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): dparams = data.frame(lambda1=0.13, lambda2=2, muminde=1, sdde=0.5)
sdn	a positive scalar used as standard deviation for the additive gaussian noise: sdn = 0.4
rseed	an integer used as seed for generating random number by the computer in use: rseed = 50

**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

xdata	a dataset with sizes, the number of rows and columns, specified by input parameters n and m1+m2, respectively
xid	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
xsd	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 column 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 MMplot variables using samples 1 and 12
A <- 0.5*(mydata$xdata[,12] + mydata$xdata[,1]);
M <- mydata$xdata[,12] - mydata$xdata[,1];

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

# Index

\* **package**

    madsim-package, [2](#)

madsim, [3](#)

madsim-package, [2](#)

madsim\_test, [5](#)