Package 'ExiMiR'

September 24, 2012

Type Package

Title R functions for the normalization of Exiqon miRNA array data

Version 1.4.0

Date 2011-11-02

Author Sylvain Gubian <DL.RSupport@pmi.com>, Alain Sewer <DL.RSupport@pmi.com>, PMP SA

Maintainer Sylvain Gubian <DL.RSupport@pmi.com>

Depends R (>= 2.10), Biobase (>= 2.5.5), affy (>= 1.26.1)

Imports affyio(>= 1.13.3), Biobase(>= 2.5.5), preprocessCore(>= 1.10.0)

Description This package contains functions for reading raw data in ImaGene TXT format obtained from Exiqon miRCURY LNA arrays, annotating them with appropriate GAL files, and normalizing them using a spike-in probe-based method.

License GPL-2

Collate make.gal.env.R read.exi.header.R read.exi.data.R read.exi.R NormiR.R summarize.miR.R

biocViews Microarray, OneChannel, DualChannel, Preprocessing,GeneExpression, Transcription

LazyLoad yes

R topics documented:

ExiMiR-package 2
cdfenv 2
galenv 3
GSE19183 3
GSE20122 3
make.gal.env 4
norm.miR 4
NormiR 6
ReadExi 8
summarize.miR 9

Index **10**

ExiMiR-package *R functions for the normalization of Exiqon miRNA array data*

Description

This package contains functions for reading raw data in ImaGene TXT format obtained from Exiqon miRCURY LNA arrays, annotating them with appropriate GAL files, and normalizing them using a spike-in probe-based method.

Details

Package:	ExiMiR
Type:	Package
Version:	0.99.0
Date:	2010-10-29
License:	GPL-2
LazyLoad:	yes

Author(s)

Sylvain Gubian, Alain Sewer, PMP SA
Maintainer: DL.RSupport@pmi.com

cdfenv *R environement for GEO series GSE19183*

Description

The cdfenv environment is a hash table for the annotation of the Affymetrix Genechip miRNA-1.0 used in the GEO series GSE19183.

Details

This cdfenv environment is based on a CDF annotation file provided directly by Affymetrix, as explained in the vignette of the ExiMiR package. It has been generated by the make.cdf.env function from the package makecdfenv.

galenv *R environement for GEO series GSE20122*

Description

The galenv environment is a hash table for the annotation of the Exiqon miRCURY LNA arrays used in the GEO series GSE20122.

Details

This galenv environment is based on a GAL annotation file provided directly by Exiqon, as explained in the vignette of the ExiMiR package. It has been generated by the make.gal.env function from ExiMiR.

GSE19183 *Affybatch object for raw data from GEO series GSE19183*

Description

The Affybatch object GSE19183 contains the raw expression data obtained from the CEL files of the GEO series GSE19183.

Details

The Affybatch object GSE19183 has been generated using the ReadAffy function from the package affy and its annotation is provided by the cdfenv environment contained in ExiMiR.

GSE20122 *Affybatch object for raw data from GEO series GSE20122*

Description

The Affybatch object GSE20122 contains the raw expression data obtained from the TXT files of the GEO series GSE20122 performed on the Exiqon miRCURY platform.

Details

The Affybatch object GSE20122 has been generated using the ReadExi function from the package ExiMiR and its annotation is provided by the galenv environment contained in ExiMiR as well.

make.gal.env *GAL Environment Maker*

Description

Reads an Exiqon GAL file and creates an environment used as a hash table for the probeset mapping location

Usage

```
make.gal.env(filename=NULL, gal.path=getwd(), verbose=FALSE)
```

Arguments

filename	Character vector. Filename of the GAL file
gal.path	Character vector. Path to the GAL file.
verbose	Logical. If TRUE, messages are shown.

Details

This function is designed similarly to make.cdf.env from the makecdfenv package. If no filename is provided as argument, the function tries to read the first GAL file in the input path. The returned environment is a hash table. For every probeset name we have a matrix with 2 columns. The first column contains the PM locations and the second column the MM locations. For PM only chips the MM column will have NAs.

Value

An environment, used as a hash table.

Author(s)

Sylvain Gubian, Alain Sewer, PMP SA

Examples

```
# The folder 'Exiqon' contains a GAL file
## Not run: galenv <- make.gal.env(gal.path='Exiqon')
```

norm.miR *miRNA raw data normalization function (low level)*

Description

A function which normalizes miRNA probe level intensities stored in an AffyBatch object. It uses the spike-in probe-based method by default. In case the spike-in probe-based method can not be applied, median normalization is executed instead. Several options allow to force the exectution of the spike-in probe-based normalization and to fine-tune the resulting correction functions.

Usage

```
norm.miR( abatch,
method=c("spikein","mean","median"),
figures.show=TRUE,
figures.output=c("display","file"),
min.corr=0.5,
loess.span=-1,
extrap.points=2,
extrap.method=c("mean","linear"),
force.zero=FALSE,
cover.ext=0.5,
cover.int=1/3,
max.log2span=1,
verbose=TRUE)
```

Arguments

abatch	An AffyBatch object.
method	Character vector. By default, spikein method is used. mean or median can also be selected and are used in case the 'spike-in' method can not be applied.
figures.show	Logical. Default value is TRUE. Control figures are generated for the spikein method.
figures.output	Character vector. By default, display is used. Figures are shown to the screen. Using file generates the figures in PDF format in the working directory.
min.corr	Numeric. Default value is 0.5. Minimal allowed value for the average of the off-diagonal elements of the Pearson correlation matrix of the spike-in probeset intensities across the arrays.
loess.span	Numeric. Default value is -1, which corresponds to a loess smoothing neighbourhood spanning a fraction 3/(number of spike-in probesets) of the total number of points. Other positive values are allowed, see the span argument of the R loess function
extrap.points	Numeric. Default value is 2. The number of spike-in probesets used in the high-intensity extrapolation of the normalization correction function.
extrap.method	Character vector. Default value is mean. The method used for the high-intensity extrapolation of the normalization correction function.
force.zero	Logical. Default value is FALSE. If TRUE, it forces the normalization correction functions to have zero values at the lower end of the probe intensity range.
cover.ext	Numeric. Default value is 1/2. Minimal allowed relative coverage of the spike-in probesets intensities. It is computed as the ratio between the intensity range covered by the spike-in probes and the one covered by all probes on the array.
cover.int	Numeric. Default value is 1/3. Maximal allowed relative intensity interval between two consecutive spike-in probesets. It is computed as the largest intensity difference between two consecutive spike-in probesets divided by the overall probe intensity range.
verbose	Logical. Default is TRUE; some details are provided on the console.
max.log2span	Numeric. Default value is 1. Gives the maximal (log2) intensity interval allowed for the probes belonging to one spike-in probeset.

Value

An AffyBatch object with expression data normalized.

Author(s)

Sylvain.Gubian, Alain.Sewer, PMP SA

Examples

```
data(galenv)
data(GSE20122)
abatch.spike <- norm.miR(GSE20122)
# Apply the affy method hist on the generated AffyBatch object abatch.spike
layout(matrix(c(1,2), 1, 2, byrow = TRUE))
hist(GSE20122)
hist(abatch.spike)
layout(1)
```

NormiR *miRNA raw data normalization function (high level)*

Description

This function converts an AffyBatch object into an ExpressionSet object performing both normalization and summarization. By default it uses the spike-in probe-based normalization method and the median summarization. In case the spike-in probe-based method cannot be applied, a median normalization is executed instead. Several options allow to force the execution of the spike-in probe-based normalization and to fine-tune the resulting correction functions.

Usage

```
NormiR( abatch,
method=c("spikein","mean","median"),
background.correct=FALSE,
verbose=TRUE,
figures.show=TRUE,
figures.output=c("display","file"),
out.type=c("ExpressionSet", "data.frame"),
min.corr=0.5,
loess.span=-1,
extrap.points=2,
extrap.method=c("mean","linear"),
force.zero=FALSE,
cover.ext=0.5,
cover.int=1/3,
max.log2span=1)
```

Arguments

abatch	AffyBatch object
method	Character vector. By default, spikein method is used. mean or median can also be selected and are used in case the 'spike-in' method can not be applied.
background.correct	Logical. Default value is FALSE. If TRUE, the rma background correction is applied.
verbose	Logical. Default value is TRUE; some details are provided on the console
figures.show	Logical. Default value is TRUE. Control figures are generated for the spikein method.
figures.output	Character vector. By default, display is used. Figures are shown to the screen. Using file generates the figures in PDF format in the working directory.
out.type	Character vector. Default value is ExpressionSet. The object type output by NormiR.
min.corr	Numeric. Default value is 0.5. Minimal allowed value for the average of the off-diagonal elements of the Pearson correlation matrix of the spike-in probeset intensities across the arrays.
loess.span	Numeric. Default value is -1, which corresponds to a loess smoothing neighbourhood spanning a fraction 3/(number of spike-in probesets) of the total number of points. Other positive values are allowed, see the span argument of the R loess function
extrap.points	Numeric. Default value is 2. The number of spike-in probesets used in the high-intensity extrapolation of the normalization correction function.
extrap.method	Character vector. Default value is mean. The method used for the high-intensity extrapolation of the normalization correction function.
force.zero	Logical. Default value is FALSE. If TRUE, it forces the normalization correction functions to have zero values at the lower end of the probe intensity range.
cover.ext	Numeric. Default value is 1/2. Minimal allowed relative coverage of the spike-in probesets intensities. It is computed as the ratio between the intensity range covered by the spike-in probes and the one covered by all probes on the array.
cover.int	Numeric. Default value is 1/3. Maximal allowed relative intensity interval between two consecutive spike-in probesets. It is computed as the largest intensity difference between two consecutive spike-in probesets divided by the overall probe intensity range.
max.log2span	Numeric. Default value is 1. Gives the maximal (log2) intensity interval allowed for the probes belonging to one spike-in probeset.

Details

See accompanying vignette.

Value

An ExpressionSet object or a data.frame object, depending on the out.type option

Author(s)

Sylvain Gubian, Alain Sewer, PMP SA

Examples

```
data(galenv)
data(GSE20122)
eset.spike <- NormiR(GSE20122)
eset.spike
```

ReadExi *Exiqon 'txt' files reader*

Description

This function reads Exiqon 'txt' files and create an AffyBatch object.

Usage

```
ReadExi( txtfile.path= getwd(),
galname= NULL,
description = NULL,
notes = '',
rm.background = FALSE,
verbose=TRUE)
```

Arguments

txtfile.path	Character vector. Path to the folder which contains samplesinfo.txt and Exiqon 'txt' files
galname	Character vector. Name of a GAL environment generated by the ExiMiR make.gal.env function.
description	a MIAME object.
notes	notes.
rm.background	Logical. Default value is FALSE. If TRUE, the background median is subtracted from the signal median.
verbose	Logical. Default value is TRUE; some details are provided on the console

Details

Exiqon 'txt' files are supplied with a samplesinfo.txt description file which lists the names of the samples files per channel. The txtfile.path argument should be a folder that contains 'txt' files and a samplesinfo.txt file. If not, the ReadExi function stops. The galname argument should be the name of the GAL environment created with make.gal.env function. If galname is not provided, an hashed environment is created based on the annotation that the 'txt' file contains.

Value

An AffyBatch object.

Warning

The image method from the AffyBatch object might not work properly when the galname argument is not assigned.

Author(s)

Sylvain Gubian, Alain Sewer, PMP SA

See Also

AffyBatch, make.gal.env

Examples

```
# The folder 'Exiqon' contains the file 'samplesinfo.txt' and
# the corresponding raw data files in TXT format
## Not run: ebatch <- ReadExi(txtfile.path='Exiqon')
# If the GAL environment has already created by the function make.gal.env
## Not run: ebatch <- ReadExi(galenv='galenv, 'txtfile.path='Exiqon')
```

---

summarize.miR *ExiMiR summarization function.*

Description

Apply median summarization on the given AffyBatch object according to the GAL or CDF environment

Usage

```
summarize.miR(abatch, out.type=c("ExpressionSet","data.frame"))
```

Arguments

abatch	An AffyBatch Object.
out.type	Character vector. By default, the output is an ExpressionSet. data.frame can be also used.

Details

The GAL or CDF environment hash is used to gather probes median intensity values into the probeset record.

Value

An ExpressionSet or a data.frame depending on the out.type argument.

Author(s)

Sylvain.Gubian, Alain Sewer, PMP SA

Examples

```
data(galenv)
data(GSE20122)
abatch.spike <- norm.miR(GSE20122, figures.show=FALSE)
eset.spike <- summarize.miR(abatch.spike)
data.spike <- summarize.miR(abatch.spike, out.type="data.frame")
```

Index

cdfenv, 2

ExiMiR *(*ExiMiR-package*)*, 2
ExiMiR-package, 2

galenv, 3
GSE19183, 3
GSE20122, 3

make.gal.env, 4

norm.miR, 4
NormiR, 6

ReadExi, 8

summarize.miR, 9