

ExiMiR

March 24, 2012

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

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.

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

<code>abatch</code>	AffyBatch object
<code>method</code>	Character vector. By default, <code>spikein</code> method is used. <code>mean</code> or <code>median</code> can also be selected and are used in case the 'spike-in' method can not be applied.
<code>background.correct</code>	Logical. Default value is <code>FALSE</code> . If <code>TRUE</code> , the <code>rma</code> background correction is applied.
<code>verbose</code>	Logical. Default value is <code>TRUE</code> ; some details are provided on the console
<code>figures.show</code>	Logical. Default value is <code>TRUE</code> . Control figures are generated for the <code>spikein</code> method.
<code>figures.output</code>	Character vector. By default, <code>display</code> is used. Figures are shown to the screen. Using <code>file</code> generates the figures in PDF format in the working directory.
<code>out.type</code>	Character vector. Default value is <code>ExpressionSet</code> . The object type output by NormiR.
<code>min.corr</code>	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.
<code>loess.span</code>	Numeric. Default value is -1, which corresponds to a loess smoothing neighbourhood spanning a fraction $3/(\text{number of spike-in probesets})$ of the total number of points. Other positive values are allowed, see the <code>span</code> argument of the R <code>loess</code> function
<code>extrap.points</code>	Numeric. Default value is 2. The number of spike-in probesets used in the high-intensity extrapolation of the normalization correction function.
<code>extrap.method</code>	Character vector. Default value is <code>mean</code> . The method used for the high-intensity extrapolation of the normalization correction function.
<code>force.zero</code>	Logical. Default value is <code>FALSE</code> . If <code>TRUE</code> , it forces the normalization correction functions to have zero values at the lower end of the probe intensity range.
<code>cover.ext</code>	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.
<code>cover.int</code>	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.
<code>max.log2span</code>	Numeric. Default value is 1. Gives the maximal (\log_2) 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

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

cdfenv

R environment 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 environment 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`.

`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

<code>filename</code>	Character vector. Filename of the GAL file
<code>gal.path</code>	Character vector. Path to the GAL file.
<code>verbose</code>	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 execution 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

<code>abatch</code>	An <code>AffyBatch</code> object.
<code>method</code>	Character vector. By default, <code>spikein</code> method is used. <code>mean</code> or <code>median</code> can also be selected and are used in case the 'spike-in' method can not be applied.
<code>figures.show</code>	Logical. Default value is <code>TRUE</code> . Control figures are generated for the <code>spikein</code> method.
<code>figures.output</code>	Character vector. By default, <code>display</code> is used. Figures are shown to the screen. Using <code>file</code> generates the figures in PDF format in the working directory.
<code>min.corr</code>	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.
<code>loess.span</code>	Numeric. Default value is -1, which corresponds to a loess smoothing neighbourhood spanning a fraction $3/(\text{number of spike-in probesets})$ of the total number of points. Other positive values are allowed, see the <code>span</code> argument of the R <code>loess</code> function
<code>extrap.points</code>	Numeric. Default value is 2. The number of spike-in probesets used in the high-intensity extrapolation of the normalization correction function.
<code>extrap.method</code>	Character vector. Default value is <code>mean</code> . The method used for the high-intensity extrapolation of the normalization correction function.
<code>force.zero</code>	Logical. Default value is <code>FALSE</code> . If <code>TRUE</code> , it forces the normalization correction functions to have zero values at the lower end of the probe intensity range.
<code>cover.ext</code>	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.
<code>cover.int</code>	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.
<code>verbose</code>	Logical. Default is <code>TRUE</code> ; some details are provided on the console.
<code>max.log2span</code>	Numeric. Default value is 1. Gives the maximal (\log_2) 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)
```

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

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

Details

The GAL or CDF environment hash is used to gather probes median intensity values into the probe-set 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`, 5

`ExiMiR` (*ExiMiR-package*), 1

`ExiMiR-package`, 1

`galenv`, 5

GSE19183, 1

GSE20122, 2

`make.gal.env`, 6

`norm.miR`, 6

`NormiR`, 2

`ReadExi`, 4

`summarize.miR`, 8