Package ‘gcrma’

April 10, 2024

Title  Background Adjustment Using Sequence Information
Version  2.74.0
Author  Jean(ZHIJIN) Wu, Rafael Irizarry
        with contributions from
        James MacDonald <jmacdon@med.umich.edu>
        Jeff Gentry
Description  Background adjustment using sequence information
Maintainer  Z. Wu <zwu@stat.brown.edu>
License  LGPL
Depends  R (>= 2.6.0), affy (>= 1.23.2), graphics, methods, stats, utils
Imports  Biobase, affy (>= 1.23.2), affyio (>= 1.13.3), XVector,
        Biostrings (>= 2.11.32), splines, BiocManager
Suggests  affydata, tools, splines, hgu95av2cdf, hgu95av2probe
biocViews  Microarray, OneChannel, Preprocessing
git_url  https://git.bioconductor.org/packages/gcrma
git_branch  RELEASE_3_18
git_last_commit  1ceeafe
git_last_commit_date  2023-10-24
Repository  Bioconductor 3.18
Date/Publication  2024-04-10

R topics documented:

affinity.spline.coefs ........................................ 2
bg.adjust.affinities ........................................ 2
bg.adjust.gcrma ............................................. 3
bg.parameters.ns ............................................. 5
compute.affinities .......................................... 6
fast.bkg ..................................................... 7
gcrma ....................................................... 8
affinity.spline.coefs  Spline coefficients for estimation of affinity from probe sequence

Description

Spline coefficients for estimation of affinity from probe sequence

Usage

data(affinity.spline.coefs)

See Also

compute.affinities

bg.adjust.fullmodel(pms,mms,ncs=NULL,apm,amm,anc=NULL,index.affinities,k=6
* fast + 0.25 * (1 - fast),rho=.7,fast=FALSE)

bg.adjust.fullmodel(pms,mms,ncs=NULL,apm,amm,anc=NULL,index.affinities,k=6
* fast + 0.25 * (1 - fast),rho=.7,fast=FALSE)

bg.adjust.fullmodel(pms,mms,ncs=NULL,apm,amm,anc=NULL,index.affinities,k=6
* fast + 0.25 * (1 - fast),fast=FALSE,nomm=FALSE)

bg.adjust.fullmodel(pms,mms,ncs=NULL,apm,amm,anc=NULL,index.affinities,k=6
* fast + 0.25 * (1 - fast),fast=FALSE,nomm=FALSE)

Arguments

pms          PM intensities after optical background correction, before non-specific-binding correction.
mms          MM intensities after optical background correction, before non-specific-binding correction.
ncs          Negative control probe intensities after optical background correction, before non-specific-binding correction. If ncs=NULL, the MM probes are considered the negative control probes.
index.affinities

The index of pms with known sequences. (For some types of arrays the sequences of a small subset of probes are not provided by Affymetrix.)

apm

Probe affinities for PM probes with known sequences.

amm

Probe affinities for MM probes with known sequences.

anc

Probe affinities for Negative control probes with known sequences. This is ignored when ncs=NULL.

rho

correlation coefficient of log background intensity in a pair of pm/mm probes. Default=.7

k

A tuning parameter. See details.

fast

Logical value. If TRUE a faster add-hoc algorithm is used.

nomm

Logical value indicating if MM intensities are available and will to be used to estimate background.

Details

Assumes PM=background1+signal, mm=background2, (log(background1), log(background2))' follows bivariate normal distribution, signal distribution follows power law. bg.parameters.gcrma and sg.parameters.gcrma provide adhoc estimates of the parameters.

the original gcrma uses an empirical Bayes estimate. this requires a complicated numerical integration. An add-hoc method tries to imitate the empirical Bayes estimate with a PM-B but values of PM-B<k going to k. This can be thought as a shrunken MVUE. For more details see Wu et al. (2003).

Value

a vector of same length as x.

Author(s)

Rafeal Irizarry, Zhijin(Jean) Wu

See Also

gcrma

description

bg.adjust.gcrma

GCRMA background adjust (internal function)

Description

This function performs background adjustment (optical noise and non-specific binding on an AffyBatch project and returns an AffyBatch object in which the PM intensities are adjusted. 
Usage

bg.adjust.gcrma(object, affinity.info=NULL, 
    affinity.source=c("reference","local"), 
    NCprobe=NULL, 
    type=c("fullmodel","affinities","mm","constant"), 
    k=6*fast+0.5*(1-fast), stretch=1.15*fast+1*(1-fast), correction=1, 
    GSB.adjust=TRUE, 
    rho=.7, optical.correct=TRUE, verbose=TRUE, fast=TRUE)

Arguments

object an AffyBatch
affinity.info NULL or an AffyBatch containing the affinities in the exprs slot. This object can be created using the function compute.affinities.
affinity.source
reference: use the package internal Non-specific binding data or local: use the experimental data in object. If local is chosen, either MM probes or a user-defined list of probes (see NCprobes) are used to estimate affinities.
NCprobe Index of negative control probes. When set as NULL, the MM probes will be used. These probes are used to estimate parameters of non-specific binding on each array. These will be also used to estimate probe affinity profiles when affinity.info is not provided.
type "fullmodel" for sequence and MM model. "affinities" for sequence information only. "mm" for using MM without sequence information.
k A tuning factor.
stretch .
correction .
GSB.adjust Logical value. If TRUE, probe effects in specific binding will be adjusted.
rho correlation coefficient of log background intensity in a pair of pm/mm probes. Default=.7
optical.correct Logical value. If TRUE, optical background correction is performed.
verbose Logical value. If TRUE messages about the progress of the function is printed.
fast Logical value. If TRUE a faster ad hoc algorithm is used.

Details

The returned value is an AffyBatch object, in which the PM probe intensities have been background adjusted. The rest is left the same as the starting AffyBatch object.

The tunning factor k will have different meanings if one uses the fast (ad hoc) algorithm or the empirical bayes approach. See Wu et al. (2003)

Value

An AffyBatch.
Author(s)
Rafeal Irizarry

Examples

```r
if(require(affydata) & require(hgu95av2probe) & require(hgu95av2cdf)){
  data(Dilution)
  ai <- compute.affinities(cdfName(Dilution))
  Dil.adj<-bg.adjust.gcrma(Dilution,affinity.info=ai,type="affinities")
}
```

Description
An internal function to be used by `gcrma`

Usage

```r
bg.parameters.ns(x,affinities,affinities2=NULL,affinities3=NULL,span=.2)
```

Arguments

- `x` PM or MM intensities after optical background correction, before non-specific-binding correction.
- `affinities` Probe affinities for probes with known sequences. Used to estimate the function between non-specific binding and affinities.
- `affinities2` Probe affinities for the probes whose expected non-specific binding intensity is to be predicted.
- `affinities3` Probe affinities for another extra group of probes whose expected non-specific binding intensity is to be predicted.
- `span` The span parameter passed to loess function

Value

A vector of the same length as `x`.

Author(s)
Rafeal Irizarry, Zhijin (Jean) Wu

See Also

`gcrma`
compute.affinities  Probe Affinity computation

Description
An internal function to calculate probe affinities from their sequences.

Usage
compute.affinities(cdfname, verbose=TRUE)
compute.affinities2(cdfname, verbose=TRUE)
check.probes(probepackage, cdfname)

Arguments
cdfname  Object of class character representing the name of CDF file associated with the arrays in the AffyBatch.
probepackage  character representing the name of the package with the probe sequence information.
verbose  Logical value. If TRUE messages about the progress of the function is printed.

Details
The affinity of a probe is described as the sum of position-dependent base affinities. Each base at each position contributes to the total affinity of a probe in an additive fashion. For a given type of base, the positional effect is modeled as a spline function with 5 degrees of freedom.

Use compute.affinities2 if there are no MM probes.
check.probes makes sure things are matching as they should.

Value
compute.affinities returns an AffyBatch with the affinities for PM probes in the pm locations and the affinities for MM probes in the mm locations. NA will be added for probes with no sequence information.

Author(s)
Rafeal Irizarry

References

See Also
gcrma, affinity.spline.coefs
**fast.bkg**  
*Internal functions for justGCRMA*

**Description**

These are internal functions for justGCRMA that are called based on memory or speed constraints.

**Usage**

```r
fast.bkg(filenames, pm.affinities, mm.affinities, index.affinities, type, minimum, optical.correct, verbose, k, rho, correction, stretch, fast, cdfname, read.verbose)
mem.bkg(filenames, pm.affinities, mm.affinities, index.affinities, type, minimum, optical.correct, verbose, k, rho, correction, stretch, fast, cdfname, read.verbose)
```

**Arguments**

- **filenames**: A list of cel files.
- **pm.affinities**: Values passed from `compute.affinities`.
- **mm.affinities**: Values passed from `compute.affinities`.
- **index.affinities**: Values passed from `compute.affinities`.
- **type**: "fullmodel" for sequence and MM model. "affinities" for sequence information only. "mm" for using MM without sequence information.
- **minimum**: A minimum value to be used for `optical.correct`.
- **optical.correct**: Logical value. If TRUE, optical background correction is performed.
- **verbose**: Logical value. If TRUE, messages about the progress of the function are printed.
- **k**: A tuning factor.
- **rho**: Correlation coefficient of log background intensity in a pair of pm/mm probes. Default=.7
- **correction**: Logical value. If TRUE, then a faster ad hoc algorithm is used.
- **stretch**: Logical value. If TRUE, then a faster ad hoc algorithm is used.
- **fast**: Logical value. If TRUE, then a faster ad hoc algorithm is used.
- **cdfname**: Used to specify the name of an alternative cdf package. If set to NULL, the usual cdf package based on Affymetrix’ mappings will be used. Note that the name should not include the ‘cdf’ on the end, and that the corresponding probe package is also required to be installed. If either package is missing an error will result.
- **read.verbose**: Logical value. If TRUE, a message is returned as each celfile is read in.
Details

Note that this expression measure is given to you in log base 2 scale. This differs from most of
the other expression measure methods.

The tuning factor 'k' will have different meanings if one uses the fast (add-hoc) algorithm or the
empirical Bayes approach. See Wu et al. (2003)

Value

An ExpressionSet.

Author(s)

James W. MacDonald <jmacdon@med.umich.edu>

See Also

gcrma

Description

This function converts an AffyBatch into an ExpressionSet using the robust multi-array average
(RMA) expression measure with help of probe sequence.

Usage

gcrma(object, affinity.info=NULL, affinity.source=c("reference","local"), NCprobe=NULL,
type=c("fullmodel","affinities","mm","constant"),
k=6*fast+0.5*(1-fast), stretch=1.15*fast+l*(1-fast), correction=1,
GSB.adjust=TRUE, rho=.7, optical.correct=TRUE, verbose=TRUE, fast=TRUE,
subset=NULL, normalize=TRUE, ...)

Arguments

object an AffyBatch
affinity.info NULL or an AffyBatch containing the affinities in the exprs slot. This object
can be created using the function compute.affinities.
affinity.source reference: use the package internal Non-specific binding data or local: use
the experimental data in object. If local is chosen, either MM probes or a
user-defined list of probes (see NCprobes) are used to estimate affinities.
NCprobe

Index of negative control probes. When set as NULL, the MM probes will be used. These probes are used to estimate parameters of non-specific binding on each array. These will be also used to estimate probe affinity profiles when affinity.info is not provided.

type

"fullmodel" for sequence and MM model. "affinities" for sequence information only. "mm" for using MM without sequence information.

k

A tuning factor.

stretch

correction

GSB.adjust

Logical value. If TRUE, probe effects in specific binding will be adjusted.

rho

correlation coefficient of log background intensity in a pair of pm/mm probes. Default=.7

optical.correct

Logical value. If TRUE, optical background correction is performed.

verbose

Logical value. If TRUE messages about the progress of the function is printed.

fast

Logical value. If TRUE a faster ad hoc algorithm is used.

subset

a character vector with the names of the probesets to be used in expression calculation.

normalize

logical value. If 'TRUE' normalize data using quantile normalization.

... further arguments to be passed (not currently implemented - stub for future use).

Details

Note that this expression measure is given to you in log base 2 scale. This differs from most of the other expression measure methods.

The tuning factor k will have different meanings if one uses the fast (add-hoc) algorithm or the empirical Bayes approach. See Wu et al. (2003)

Value

An ExpressionSet.

Author(s)

Rafeal Irizarry

Examples

```r
if(require(affydata) & require(hgu95av2probe) & require(hgu95av2cdf)){
  data(Dilution)
  ai <- compute.affinities(cdfName(Dilution))
  Dil.expr<-gcrma(Dilution,affinity.info=ai,type="affinities")
}
```
gcrma.engine

GCRMA background adjust engine (internal function)

Description

This function adjusts for non-specific binding when all arrays in the dataset share the same probe affinity information. It takes matrices of PM probe intensities, MM probe intensities, other negative control probe intensities (optional) and the associated probe affinities, and return one matrix of non-specific binding corrected PM probe intensities.

Usage

```r
gcrma.engine(pms, mms, ncs = NULL,
              pm.affinities = NULL, mm.affinities = NULL, anc = NULL,
              type = c("fullmodel", "affinities", "mm", "constant"),
              k = 6*fast + 0.5*(1-fast),
              stretch = 1.15*fast + 1*(1-fast), correction = 1,
              GSB.adjust = TRUE, rho = 0.7,
              verbose = TRUE, fast = FALSE)
```

Arguments

- `pms`: The matrix of PM intensities
- `mms`: The matrix of MM intensities
- `ncs`: The matrix of negative control probe intensities. When left as `NULL`, the MMs are considered the negative control probes.
- `pm.affinities`: The vector of PM probe affinities. Note: This can be shorter than the number of rows in `pms` when some probes do not have sequence information provided.
- `mm.affinities`: The vector of MM probe affinities.
- `anc`: The vector of Negative Control probe affinities. This is ignored if MMs are used as negative controls (`ncs=NULL`)
- `type`: "fullmodel" for sequence and MM model. "affinities" for sequence information only. "mm" for using MM without sequence information.
- `k`: A tuning factor.
- `stretch`: .
- `correction`: .
- `GSB.adjust`: Logical value. If `TRUE`, probe effects in specific binding will be adjusted.
- `rho`: correlation coefficient of log background intensity in a pair of pm/mm probes. Default = .7
- `verbose`: Logical value. If `TRUE` messages about the progress of the function is printed.
- `fast`: Logical value. If `TRUE` a faster add-hoc algorithm is used.
**Details**

Note that this expression measure is given to you in log base 2 scale. This differs from most of the other expression measure methods.

The tuning factor $k$ will have different meanings if one uses the fast (add-hoc) algorithm or the empirical bayes approach. See Wu et al. (2003)

**Value**

A matrix of PM intensities.

**Author(s)**

Rafeal Irizarry & Zhijin Wu

**See Also**

gcrma.engine2

---

**gcrma.engine2**

*GCRMA background adjust engine (internal function)*

**Description**

This function adjust for non-specific binding when each array has its own probe affinity information. It takes an AffyBatch object of probe intensities and an AffyBatch of probe affinity, returns one matrix of non-specific binding corrected PM probe intensities.

**Usage**

```r
gcrma.engine2(object, pmIndex=NULL, mmIndex=NULL, NCprobe=NULL, affinity.info, type=c("fullmodel","affinities","mm","constant"), k=6*fast+0.5*(1-fast), stretch=1.15*fast+1*(1-fast), correction=1, GSB.adjust=TRUE, rho=0.7, verbose=TRUE, fast=TRUE)
```

**Arguments**

- `object` an `AffyBatch`. Note: this is an internal function. Optical noise should have been corrected for.

- `pmIndex` Index of PM probes. This will be computed within the function if left `NULL`

- `mmIndex` Index of MM probes. This will be computed within the function if left `NULL`

- `NCprobe` Index of negative control probes. When set as `NULL`, the MM probes will be used. These probes are used to estimate parameters of non-specific binding on each array. These will be also used to estimate probe affinity profiles when `affinity.info` is not provided.
**getCDF**

<table>
<thead>
<tr>
<th>argument</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>affinity.info</td>
<td>NULL or an AffyBatch containing the affinities in the exprs slot. This object can be created using the function <code>compute.affinities</code>.</td>
</tr>
<tr>
<td>type</td>
<td>&quot;fullmodel&quot; for sequence and MM model. &quot;affinities&quot; for sequence information only. &quot;mm&quot; for using MM without sequence information.</td>
</tr>
<tr>
<td>k</td>
<td>A tuning factor.</td>
</tr>
<tr>
<td>stretch</td>
<td>.</td>
</tr>
<tr>
<td>correction</td>
<td>.</td>
</tr>
<tr>
<td>GSB.adjust</td>
<td>Logical value. If TRUE, probe effects in specific binding will be adjusted.</td>
</tr>
<tr>
<td>rho</td>
<td>correlation coefficient of log background intensity in a pair of pm/mm probes. Default=.7</td>
</tr>
<tr>
<td>verbose</td>
<td>Logical value. If TRUE messages about the progress of the function is printed.</td>
</tr>
<tr>
<td>fast</td>
<td>Logical value. If TRUE a faster add-hoc algorithm is used.</td>
</tr>
</tbody>
</table>

**Details**

Note that this expression measure is given to you in log base 2 scale. This differs from most of the other expression measure methods.

The tuning factor k will have different meanings if one uses the fast (add-hoc) algorithm or the empirical bayes approach. See Wu et al. (2003)

**Value**

A matrix of PM intensities.

**Author(s)**

Rafeal Irizarry & Zhijin Wu

**See Also**

gcrma.engine

---

**Description**

These are internal functions that are called by justGCRMA and GCRMA in order to automatically download and install cdf environments and probe packages.

**Usage**

```r
cCDF(cdfname, lib = .libPaths()[1], verbose = TRUE)
ggetProbePackage(probepackage, lib = .libPaths()[1], verbose = TRUE)
```
justGCRMA

Arguments

cdfname  Name of the cdfenv to install.
probepackage  Name of the probe package to install.
lib  Directory of the R library where packages will be installed.
verbose  Output informative comments? Defaults to TRUE

Value

Nothing is returned. These functions are called simply to install environments.

Author(s)

James W. MacDonald, based on getCDFinfo, written by Jeff Gentry.

See Also

getCDFinfo

justGCRMA  Compute GCRMA Directly from CEL Files

Description

This function converts CEL files into an ExpressionSet using the robust multi-array average (RMA) expression measure with help of probe sequences.

Usage

```r
just.gcrma(..., filenames=character(0),
  phenoData=new("AnnotatedDataFrame"),
  description=NULL,
  notes="", compress=getOption("BioC")$affy$compress.cel,
  normalize=TRUE, bgversion=2, affinity.info=NULL,
  type=c("fullmodel","affinities","mm","constant"),
  k=6*fast+0.5*(1-fast), stretch=1.15*fast+1*(1-fast),
  correction=1, rho=0.7, optical.correct=TRUE,
  verbose=TRUE, fast=TRUE, minimum=1, optimize.by =
  c("speed","memory"),
  cdfname = NULL, read.verbose = FALSE)

justGCRMA(..., filenames=character(0),
  widget=getOption("BioC")$affy$use.widgets,
  compress=getOption("BioC")$affy$compress.cel,
  celfile.path=getwd(),
  sampleNames=NULL,
  phenoData=NULL,
```

description=NULL, notes="", normalize=TRUE, bgversion=2, affinity.info=NULL, type=c("fullmodel","affinities","mm","constant"), k=6*fast+0.5*(1-fast), stretch=1.15*fast+1*(1-fast), correction=1, rho=0.7, optical.correct=TRUE, verbose=TRUE, fast=TRUE, minimum=1, optimize.by = c("speed","memory"), cdfname = NULL, read.verbose = FALSE)

Arguments

... file names separated by comma.
filenames file names in a character vector.
widget a logical specifying if widgets should be used.
compress are the CEL files compressed?
phenoData a AnnotatedDataFrame object.
description a MIAME object.
notes notes.
affinity.info NULL or a list of three components: apm,amm and index, for PM probe affinities, MM probe affinities, the index of probes with known sequence, respectively.
type "fullmodel" for sequence and MM model. "affinities" for sequence information only. "mm" for using MM without sequence information.
k A tuning factor.
rho correlation coefficient of log background intensity in a pair of pm/mm probes. Default=.7.
stretch .
correction .
normalize Logical value. If TRUE, then normalize data using quantile normalization.
optical.correct Logical value. If TRUE, then optical background correction is performed.
verbose Logical value. If TRUE, then messages about the progress of the function is printed.
fast Logical value. If TRUE, then a faster add-hoc algorithm is used.
optimize.by "speed" will use a faster algorithm but more RAM, and "memory" will be slower, but require less RAM.
bgversion integer value indicating which RMA background to use 1: use background similar to pure R rma background given in affy version 1.0 - 1.0.2 2: use background similar to pure R rma background given in affy version 1.1 and above.
minimum .
celfile.path a character denoting the path 'ReadAffy' should look for cel files.
sampleNames a character vector of sample names to be used in the 'AffyBatch'.
cdfname Used to specify the name of an alternative cdf package. If set to NULL, the usual cdf package based on Affymetrix' mappings will be used. Note that the name should not include the 'cdf' on the end, and that the corresponding probe package is also required to be installed. If either package is missing an error will result.
read.verbose Logical value. If TRUE, then messages will be printed as each celfile is read in.

Details
This method should require much less RAM than the conventional method of first creating an AffyBatch and then running gcrma.
This is a simpler version than gcrma, so some of the arguments available in gcrma are not available here. For example, it is not possible to use the MM probes to estimate background. Instead, the internal NSB estimates are used (which is also the default for gcrma).
Note that this expression measure is given to you in log base 2 scale. This differs from most of the other expression measure methods.
The tuning factor k will have different meanings if one uses the fast (add-hoc) algorithm or the empirical Bayes approach. See Wu et al. (2003)
fast.bkg and mem.bkg are two internal functions.

Value
An ExpressionSet object.

Author(s)
James W. MacDonald
Index

* datasets
  affinity.spline.coefs, 2
* internal
  fast.bkg, 7
  getCDF, 12
* manip
  bg.adjust.affinities, 2
  bg.adjust.gcrma, 3
  bg.parameters.ns, 5
  compute.affinities, 6
  gcrma, 8
  gcrma.engine, 10
  gcrma.engine2, 11
  justGCRMA, 13

  affinity.spline.coefs, 2, 6
  AffyBatch, 4, 8, 11
  AnnotatedDataFrame, 14
  average.for.PAV(bg.parameters.ns), 5

  base.profiles(compute.affinities), 6
  bg.adjust.affinities, 2
  bg.adjust.constant
    (bg.adjust.affinities), 2
  bg.adjust.fullmodel
    (bg.adjust.affinities), 2
  bg.adjust.gcrma, 3
  bg.adjust.mm(bg.adjust.affinities), 2
  bg.adjust.optical
    (bg.adjust.affinities), 2
  bg.parameters.ns, 5

  check.probes(compute.affinities), 6
  compute.affinities, 2, 4, 6, 8, 12
  compute.affinities2
    (compute.affinities), 6
  compute.affinity.coef
    (compute.affinities), 6

  ExpressionSet, 15

  fast.bkg, 7
  gcrma, 2, 3, 5, 6, 8, 15
  gcrma.engine, 10
  gcrma.engine2, 11
  getCDF, 12
  getProbePackage(getCDF), 12
  GSB.adv(gcrma), 8

  just.gcrma(justGCRMA), 13
  justGCRMA, 13

  left.sigma(bg.parameters.ns), 5

  mem.bkg(fast.bkg), 7
  MIAME, 14

  PAV(bg.parameters.ns), 5
  plotBaseProfiles(compute.affinities), 6