Package 'rMAT'

October 18, 2017

Type Package

Title R implementation from MAT program to normalize and analyze tiling arrays and ChIP-chip data.

Version 3.26.0

Author Charles Cheung and Arnaud Droit and Raphael Gottardo

Maintainer Arnaud Droit <arnaud.droit@crchuq.ulaval.ca> and Raphael Gottardo <rgottard@fhcrc.org>

Description This package is an R version of the package MAT and contains functions to parse and merge Affymetrix BPMAP and CEL tiling array files (using C++ based Fusion SDK and Bioconductor package affxparser), normalize tiling arrays using sequence specific models, detect enriched regions from ChIP-chip experiments. Note: users should have GSL and GenomeGraphs installed. Windows users: 'consult the README file available in the inst directory of the source distribution for necessary configuration instructions'. Snow Leopard users can take advantage of increase speed with Grand Central Dispatch!

Year 2009.

License Artistic-2.0

Depends R(>= 2.9.0), BiocGenerics (>= 0.1.3), IRanges (>= 1.13.10), Biobase (>= 2.15.1), affxparser

Imports stats, methods, BiocGenerics, IRanges, Biobase, affxparser, stats4

Suggests GenomeGraphs, rtracklayer

SystemRequirements GSL (GNU Scientific Library)

biocViews Microarray, Preprocessing

URL http://www.rglab.org

NeedsCompilation yes

R topics documented:

`BPMAPCelParser` *BPMAP and CEL files Reader*

Description

One-step reading of BPMAP and CEL files, using Fusion SDK and affxparser.

Usage

```
BPMAPCelParser(BPMAPFileName, CelFileNames, genomeName=NULL, verbose=FALSE,groupName="",seqName="
```

Arguments

`BPMAPFileName`	String containing the full filename of the BPMAP file.
`CelFileNames`	Vector of strings containing full filenames of CEL files. i.e. c("F1.CEL", "F2.CEL"
`genomeName`	String containing the genome name used.
`groupName`	String containing the group of genome name used.
`seqName`	String containing the group of sequence name (e.g. chromosome) used.
`verbose`	If verbose is selected, the progress and additional information will be displayed while the function is running

Details

This function returns an object of class `tilingSet` containing all necessary information: probe sequences, genomic positions, chromosomes as well as the probe intensities.

Value

An object of class `tilingSet`.

Author(s)

Charles Cheung, <cykc@interchange.ubc.ca> and Raphael Gottardo, <rgottard@fhcrc.org>
Arnaud Droit, <arnaud.droit@crchuq.ulaval.ca>

See Also

`affyTile` for information about the package.

Examples

```
#####################################################
#The data are in inst/doc folder in rMAT package
#####################################################

pwd<-"" #INPUT FILES- BPMAP, ARRAYS, etc.
path<- system.file("extdata", "Sc03b_MR_v04_10000.bpmap",package="rMAT")

bpmapFile<-paste(pwd,path,sep="")

pathCEL<- system.file("extdata", "Swr1WTIP_Short.CEL",package="rMAT")
arrayFile<-paste(pwd,c(pathCEL),sep="")

# Show the all the different sequences
ReadBPMAPAllSeqHeader(bpmapFile)

# create a tiling Set from the corresponding data
# This will only grep the sequences with Sc
ScSet<-BPMAPCelParser(bpmapFile, arrayFile, verbose=FALSE,groupName="Sc")

# show the object
show(ScSet)

# summarize its content
summary(ScSet)
```

---

callEnrichedRegions *Detection of enriched regions*

Description

This function is used to locate putative enriched regions.

Usage

```
callEnrichedRegions(MatScore, dMax=600, dMerge=300, nProbesMin=8, method="score", threshold=5, ve
```

Arguments

MatScore	This object contains an Range Data file
dMax	An integer value. The sliding window side of which the adjacent probes are to average upon in order to compute the rMAT score.
dMerge	An integer value. The maximum size to merge adjacent probes and categorize them as one region for scores of adjacent probes uniformly above the input threshold.
nProbesMin	An integer value. The minimum number of probes to average upon. If the number of probes within the interval is less than nProbesMin, the rMAT score of the region will not be computed.

method	A character string value equal to "score", "pValue" or "FDR". "score" denotes the method of calling enriched regions based sliding widow scores. "pValue" denotes the method of calling enriched regions based on p-values. Method "FDR" uses an FDR procedure to call regions. See Details below.
threshold	An integer value. The threshold of rMAT Score to be labeled as an enriched region. For method=1 or 3, the higher the score, the more confident we are about enriched regions. For method=2, the lower the score, the more confident we are about enriched regions.
verbose	A logical value. If verbose is TRUE, progress information would be displayed.

Details

For more details on the calculation of the rMAT score, pvalues, etc, please refer to the following paper: Johnson et al. Model-based analysis of tiling-arrays for ChIP-chip. Proc Natl Acad Sci USA (2006) vol. 103 (33) pp. 12457-62

Author(s)

Charles Cheung, <cykc@interchange.ubc.ca> and Raphael Gottardo, <rgottard@fhcrc.org> Arnaud Droit, <arnaud.droit@crchuq.ulaval.ca>

See Also

NormalizeProbes, computeMATScore.

Examples

```
#####################################################
#The data are in inst/doc folder in rMAT package.
#####################################################


pwd<-"" #INPUT FILES- BPMAP, ARRAYS, etc.
path<- system.file("extdata", "Sc03b_MR_v04_10000.bpmap",package="rMAT")

bpmapFile<-paste(pwd,path,sep="")

pathCEL<- system.file("extdata", "Swr1WTIP_Short.CEL",package="rMAT")
arrayFile<-paste(pwd,c(pathCEL),sep="")


# Show the all the different sequences
ReadBPMAPAllSeqHeader(bpmapFile)

# create a tiling Set from the corresponding data
# This will only grep the sequences with Sc
ScSet<-BPMAPCelParser(bpmapFile, arrayFile, verbose=FALSE,groupName="Sc")

# show the object
show(ScSet)

# summarize its content
summary(ScSet)
```

```
ScSetNorm<-NormalizeProbes(ScSet, method="MAT",robust=FALSE, all=FALSE, standard=TRUE, verbose=FALSE)

RD<-computeMATScore(ScSetNorm,cName=NULL, dMax=600, verbose=TRUE)
Enrich<-callEnrichedRegions(RD,dMax=600, dMerge=300, nProbesMin=8, method="score", threshold=1, verbose=FALS
```

`computeMATScore` *Detection of enriched regions*

Description

This function is used to compute the rMAT scores following normalization of expression values in order to locate putative enriched regions. This function is now defunct now defunct and you should instead use 'computeMATScore'.

Usage

```
computeMATScore(tilingSet, cName=NULL, dMax=600, verbose=FALSE)
```

Arguments

`tilingSet`	This object contains an ExpressionSet
`cName`	Unique identifier of control name
`dMax`	An integer value. The sliding window side of which the adjacent probes are to average upon in order to compute the rMAT score.
`verbose`	A logical value. If verbose is TRUE, progress information would be displayed.

Details

For more details on the calculation of the rMAT score, pvalues, etc, please refer to the following paper: Johnson et al. Model-based analysis of tiling-arrays for ChIP-chip. Proc Natl Acad Sci USA (2006) vol. 103 (33) pp. 12457-62

Value

The rMAT Score, pValues, and regions. For the regions vector, let 0 denotes the unenriched region. If an enriched region is found, the interval of the region is labeled by a none 0 value. The first region detected is abeled 1 and the next regions are subsequently incremented.

Author(s)

Charles Cheung, <cykc@interchange.ubc.ca> and Raphael Gottardo, <rgottard@fhcrc.org> Arnaud Droit, <arnaud.droit@crchuq.ulaval.ca>

See Also

`NormalizeProbes`, `computeMATScore`, `callEnrichedRegions` for normalizing expression values before computing the rMAT enriched regions.

Examples

```
####################################################
#The data are in inst/doc folder in rMAT package.
####################################################


pwd<-"" #INPUT FILES- BPMAP, ARRAYS, etc.
path<- system.file("extdata", "Sc03b_MR_v04_10000.bpmap",package="rMAT")

bpmapFile<-paste(pwd,path,sep="")

pathCEL<- system.file("extdata", "Swr1WTIP_Short.CEL",package="rMAT")
arrayFile<-paste(pwd,c(pathCEL),sep="")


# Show the all the different sequences
ReadBPMAPAllSeqHeader(bpmapFile)

# create a tiling Set from the corresponding data
# This will only grep the sequences with Sc
ScSet<-BPMAPCelParser(bpmapFile, arrayFile, verbose=FALSE,groupName="Sc")

# show the object
show(ScSet)

# summarize its content
summary(ScSet)

ScSetNorm<-NormalizeProbes(ScSet, method="MAT",robust=FALSE, all=FALSE, standard=TRUE, verbose=FALSE)

RD<-computeMATScore(ScSetNorm,cName=NULL, dMax=600, verbose=TRUE)
```

`MATScore` *Detection of enriched regions*

Description

This function is used to compute the rMAT scores following normalization of expression values in order to locate putative enriched regions. This function is now defunct now defunct and you should instead use 'computeMATScore'.

Usage

```
MATScore(tilingSet, cName="NULL", dMax=600,nProbesMin=8, dMerge=300,method="score",threshold=5,v
```

Arguments

`tilingSet`	This object contains an ExpressionSet
`cName`	Unique identifier of control name

dMax	An integer value. The sliding window side of which the adjacent probes are to average upon in order to compute the rMAT score.
nProbesMin	An integer value. The minimum number of probes to average upon. If the number of probes within the interval is less than nProbesMin, the rMAT score of the region will not be computed.
dMerge	An integer value. The maximum size to merge adjacent probes and categorize them as one region for scores of adjacent probes uniformly above the input threshold.
method	A character string value equal to "score", "pValue" or "FDR". "score" denotes the method of calling enriched regions based sliding widow scores. "pValue" denotes the method of calling enriched regions based on p-values. Method "FDR" uses an FDR procedure to call regions. See Details below.
threshold	An integer value. The threshold of rMAT Score to be labeled as an enriched region. For method=1 or 3, the higher the score, the more confident we are about enriched regions. For method=2, the lower the score, the more confident we are about enriched regions.
verbose	A logical value. If verbose is TRUE, progress information would be displayed.
bedName	This file file includes columns "chromosome rMATScore region pValue" for each probe.

Details

For more details on the calculation of the rMAT score, pvalues, etc, please refer to the following paper: Johnson et al. Model-based analysis of tiling-arrays for ChIP-chip. Proc Natl Acad Sci USA (2006) vol. 103 (33) pp. 12457-62

Value

The rMAT Score, pValues, and regions. For the regions vector, let 0 denotes the unenriched region. If an enriched region is found, the interval of the region is labeled by a none 0 value. The first region detected is labeled 1 and the next regions are subsequently incremented.

Author(s)

Charles Cheung, <cykc@interchange.ubc.ca> and Raphael Gottardo, <rgottard@fhcrc.org>
Arnaud Droit, <arnaud.droit@crchuq.ulaval.ca>

See Also

NormalizeProbes, computeMATScore, callEnrichedRegions for normalizing expression values before computing the rMAT enriched regions.

Examples

```
##################################################
#The data are in inst/doc folder in rMAT package.
##################################################


#pwd<-"" #INPUT FILES- BPMAP, ARRAYS, etc.
#path<- system.file("extdata", "Sc03b_MR_v04_10000.bpmap",package="rMAT")
```

```
#bpmapFile<-paste(pwd,path,sep="")

#pathCEL<- system.file("extdata", "Swr1WTIP_Short.CEL",package="rMAT")
#arrayFile<-paste(pwd,c(pathCEL),sep="")


# Show the all the different sequences
#ReadBPMAPAllSeqHeader(bpmapFile)

# create a tiling Set from the corresponding data
# This will only grep the sequences with Sc
#ScSet<-BPMAPCelParser(bpmapFile, arrayFile, verbose=FALSE,groupName="Sc")

# show the object
#show(ScSet)

# summarize its content
#summary(ScSet)


#ScSetNorm<-NormalizeProbes(ScSet, method="MAT",robust=FALSE, all=FALSE, standard=TRUE, verbose=FALSE)

#ScScore<- MATScore(ScSetNorm, cName=NULL, dMax=600,nProbesMin=8, dMerge=300,method="score",threshold=5,verb
```

NormalizeProbes *Normalize tiling array data using sequence information*

Description

This function is used to normalize tiling array data using sequence information. Users can chose between two different normalization methods. Please refer to the arguments section below.

Usage

```
NormalizeProbes(tilingSet, method="MAT", robust=FALSE,
                all=FALSE, standard=TRUE, verbose=FALSE)
```

Arguments

tilingSet	This object contains an ExpressionSet and has the following additional slots
method	The normalization method to be used. User can choose from "MAT", or "PairBinned". As an upgrade to MAT, the Pair option also takes into account of the interaction between adjacent pairs along the probe as covariates for linear regression.
robust	A logical value. If TRUE, reweighted least-squares estimates are computed.
all	A logical value. If not using all probes to compute (for faster computation and memory efficiency) the regression parameters, then use the minimum of 300,000 or number of probes, whichever is less.
standard	Typical methods.
verbose	A logical value. If verbose is TRUE, progress information would be displayed.

Details

For the original rMAT normalization: method is set to be rMAT in string, robust is set to be false, copyNumber is set to be your copy number's vector, rMATScaling is set to be true, and logTransform is set to be true for untransformed data. The output can be saved as BAR file if the BAR argument specifies a filename, or as a parsed BAR file if argument output specifies a filename.

For more details on normalization, please refer to the following paper: Johnson et al. Model-based analysis of tiling-arrays for ChIP-chip. Proc Natl Acad Sci USA (2006) vol. 103 (33) pp. 12457-62

Value

The matrix of normalized expression values.

Author(s)

Charles Cheung, <cykc@interchange.ubc.ca> and Raphael Gottardo, <rgottard@fhcrc.org> Arnaud Droit, <arnaud.droit@crchuq.ulaval.ca>

See Also

`PairInMatrix()` for generating neighbouring pair-codes from sequences and `affyTile` for information about the package.

Examples

```
#####################################################
#The data are in inst/doc folder in rMAT package.
#####################################################



pwd<-"" #INPUT FILES- BPMAP, ARRAYS, etc.
path<- system.file("extdata", "Sc03b_MR_v04_10000.bpmap",package="rMAT")

bpmapFile<-paste(pwd,path,sep="")

pathCEL<- system.file("extdata", "Swr1WTIP_Short.CEL",package="rMAT")
arrayFile<-paste(pwd,c(pathCEL),sep="")


# Show the all the different sequences
ReadBPMAPAllSeqHeader(bpmapFile)

# create a tiling Set from the corresponding data
# This will only grep the sequences with Sc

ScSet <- BPMAPCelParser(bpmapFile, arrayFile, verbose=FALSE, groupName="Sc")
ScSetNorm <- NormalizeProbes(ScSet, method="MAT", robust=FALSE, all=FALSE,
    standard=TRUE, verbose=FALSE)
```

ReadBPMAPAllSeqHeader *Reading All the BPMAP Sequence Header*

Description

Reading the header of a specified sequence in the BPMAP file. Several sequences could be stored in a single Affymetrix Tiling Array. For example, an array could contain probes from Chromosome 21 and Chromosome 22. The sequenceNum uniquely specifies a sequence. Information about this sequence could be determined in this function. The total number of sequences a tiling array contains can be determined in ReadBPMAPHeader(fileName). The sequenceNum indexes from 0 to (total number of sequences -1).

Usage

```
ReadBPMAPAllSeqHeader(fileName)
```

Arguments

fileName	the full path of the BPMAP file to be read.

Details

The BPMAP Sequence Header gives information about the design of the tiling array.

Value

A list of vectors containing SeqName, GroupName, version, npnrobeMapping, seqNum, and NumHits.

Author(s)

Charles Cheung, <cykc@interchange.ubc.ca> and Raphael Gottardo, <rgottard@fhcrc.org>
Arnaud Droit, <arnaud.droit@crchuq.ulaval.ca>

See Also

BPMAPCelParser() for an one-step BPMAP/CEL parser and affyTile for information about the package.

Examples

```
#####################################################
#The data are in inst/doc folder in rMAT package.
#####################################################


pwd<-"" #INPUT FILES- BPMAP, ARRAYS, etc.
path<- system.file("extdata", "Sc03b_MR_v04_10000.bpmap",package="rMAT")

bpmapFile<-paste(pwd,path,sep="")

pathCEL<- system.file("extdata", "Swr1WTIP_Short.CEL",package="rMAT")
arrayFile<-paste(pwd,c(pathCEL),sep="")
```

```
# Show the all the different sequences
ReadBPMAPAllSeqHeader(bpmapFile)
```

`show,tilingSet-method` *show Method for tiling set object*

Description

This methods show the content of tilinSet objets

Usage

```
## S4 method for signature 'tilingSet'
show(object)
```

Arguments

`object`	Object returned of class `tilingSet`

Details

TilingSet contais an ExpressionSet and has the following additional slots: genomeName, featureSequence, featurePosition, featureChromosome, featureCopyNumber, featureSequence.

Author(s)

Charles Cheung, <cykc@interchange.ubc.ca> and Raphael Gottardo, <rgottard@fhcrc.org> Arnaud Droit, <arnaud.droit@crchuq.ulaval.ca>

See Also

`BPMAPCelParser`, `NormalizeProbes`

`summary,tilingSet-method`
Summary Method for MAT Object

Description

This methods summarize tilinSet object

Usage

```
## S4 method for signature 'tilingSet'
summary(object)
```

Arguments

object	A tilingSet object.

Details

This function will give a basic summary of a tilingSet object including chromosome/genome information.

Author(s)

Charles Cheung, <cykc@interchange.ubc.ca> and Raphael Gottardo, <rgottard@fhcrc.org>
Arnaud Droit, <arnaud.droit@crchuq.ulaval.ca>

See Also

BPMAPCelParser, NormalizeProbes

---

tilingSet *This object contains an ExpressionSet*

Description

This object contains an ExpressionSet and has the following additional slots: genomeName, featureSequence, featurePosition, featureChromosome, featureCopyNumber

Details

Tiling set objects can also be combined using the rbind methods. This is particularly useful when several arrays span a genome/chromosome.

Author(s)

Charles Cheung, <cykc@interchange.ubc.ca> and Raphael Gottardo, <rgottard@fhcrc.org>
Arnaud Droit, <arnaud.droit@crchuq.ulaval.ca>

References

W. E. Johnson, Li, W., Meyer, C. A., Gottardo, R., Carroll, J. S., Brown, M., and Liu, X. S. (2006). Model-based analysis of tiling-arrays for ChIP-chip. PNAS 103:12457-12462.

See Also

BPMAPCelParser, NormalizeProbes

Examples

```
featureChromosome=factor(c("chr1","chr1","chr1","chr1"))
featurePosition=c(as.integer(47193),as.integer(47197),as.integer(47201),
    as.integer(47205))
featureCopyNumber=c(as.integer(1),as.integer(1),as.integer(1),as.integer(1))
a=5.379897
exprs=matrix(a,nrow=4)
genomeName="Sc03b_MR_v04_10000"
featureSequence=c("TCATCAAGGGAAGAGAGTCTCTCAG","TGATCATCACGGGACTTCTGGTTTA","CGGGACTTCTGGTTTATGGAACTAT","ACT

newSet <- new('tilingSet', featureChromosome=featureChromosome,
    featurePosition=featurePosition,featureCopyNumber=featureCopyNumber,
    exprs=exprs, genomeName=genomeName, featureSequence=featureSequence)
```

Index