rMAT

November 11, 2009

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,group
```

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, ⟨raphael.gottardo@ircm.qc.ca⟩
Arnaud Droit, ⟨arnaud.droit@ircm.qc.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("doc/Sc03b_MR_v04_10000.bpmap",package="rMAT")

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

pathCEL<- system.file("doc/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)
```

`MAT` *This object contains an ExpressionSet*

Description

An object summarizing the output from the MAT function with the following slots : genomeName,featurePosition,featureChromosome,score,regIndex,method,threshold

Usage

```
NewMAT<-new('MAT', genomeName="Sc", featurePosition, featureChromosome, score, r
```

Author(s)

Charles Cheung, ⟨cykc@interchange.ubc.ca⟩ and Raphael Gottardo, ⟨raphael.gottardo@ircm.qc.ca⟩
Arnaud Droit, ⟨arnaud.droit@ircm.qc.ca⟩

References

Lo, K., Brinkman, R. R. and Gottardo, R. (2008) Automated Gating of Flow Cytometry Data via Robust Model-based Clustering. *Cytometry A* **73**, 321-332.

See Also

tilingSet

Examples

```
#       genomeName="Sc03b_MR_v04"
#       featurePosition="4"
#       featureChromosome="chr1"
#       score="8.8623"
#       regIndex="177"
#       method=""
#       threshold="5"

#NewMAT<-new('MAT', genomeName="Sc", featurePosition, featureChromosome, score, regIndex,
```

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.

Usage

```
MATScore(tilingSet, cName="NULL", dMax=600,nProbesMin=8, dMerge=300,method="scor
```

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, ⟨raphael.gottardo@ircm.qc.ca⟩
Arnaud Droit, ⟨arnaud.droit@ircm.qc.ca⟩

See Also

`NormalizeProbes()` 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("doc/Sc03b_MR_v04_10000.bpmap",package="rMAT")

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

pathCEL<- system.file("doc/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, ve

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

`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

```
ScSetNorm<-NormalizeProbes(tilingSet, method="MAT",robust=FALSE, all=FALSE, stan
```

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.
`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, ⟨raphael.gottardo@ircm.qc.ca⟩
Arnaud Droit, ⟨arnaud.droit@ircm.qc.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("doc/Sc03b_MR_v04_10000.bpmap",package="rMAT")

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

pathCEL<- system.file("doc/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, ve
```

`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, ⟨raphael.gottardo@ircm.qc.ca⟩
Arnaud Droit, ⟨arnaud.droit@ircm.qc.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("doc/Sc03b_MR_v04_10000.bpmap",package="rMAT")

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

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

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

`show,MAT-method` *show Method for MAT Object*

Description

This methods show the objects MAT or tilinSet

Usage

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

Arguments

`object`	Object returned from `MAT` .

Details

MAT or TilingSet contains an ExpressionSet and has the following additional slots: genomeName, featureSequence, featurePosition, featureChromosome, featureCopyNumber, featureSequence. This methods shows the information.

Author(s)

Charles Cheung, ⟨cykc@interchange.ubc.ca⟩ and Raphael Gottardo, ⟨raphael.gottardo@ircm.qc.ca⟩ Arnaud Droit, ⟨arnaud.droit@ircm.qc.ca⟩

See Also

`MAT`

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

Description

This methods summarize the objects MAT or tilinSet

Usage

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

Arguments

object	Object returned from MAT .

Details

MAT is an object summarizing the output from the MAT function with the following slots : genomeName,featurePosition,featureChromosome,score,regIndex,method,threshold.

TilingSet is an object containing an ExpressionSet with the follonwing slots: genomeName, featureSequence, featurePosition, featureChromosome, featureCopyNumber

Author(s)

Charles Cheung, ⟨cykc@interchange.ubc.ca⟩ and Raphael Gottardo, ⟨raphael.gottardo@ircm.qc.ca⟩
Arnaud Droit, ⟨arnaud.droit@ircm.qc.ca⟩

See Also

MAT

---

tilingSet *This object contains an ExpressionSet*

Description

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

Usage

```
newSet<-new('tilingSet', featureChromosome,featurePosition, featureCopyNumber,
```

Arguments

tilingSet	This object contains an ExpressionSet
genomeName	String containing the genome name used (vector).
featureChromosome	String containing the name of chromosome used (vector).
featurePosition	String containing the Position of the sequences (vector).
featureCopyNumber	String containing the copy number of sequence (vector).
exprs	String containing the expresion data of enriched region (matrix with n column.

`featureSequence`
: String containing the sequence (vector).

`experimentData`
: String containing the type of experiments.

Author(s)

Charles Cheung, ⟨cykc@interchange.ubc.ca⟩ and Raphael Gottardo, ⟨raphael.gottardo@ircm.qc.ca⟩
Arnaud Droit, ⟨arnaud.droit@ircm.qc.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

MAT

Examples

```
featureChromosome=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","CGGGACTTCTGGTT

newSet<-new('tilingSet', featureChromosome=featureChromosome,featurePosition=featurePosit
```

Index