Package 'IMA'

September 7, 2011

Type Package

Title IMA:Illumina Methylation Analyzer

Version 1.2.2

Date 2011-08-25

Depends R (>= 2.12.0), methods, utils,stats,WriteXLS,limma,dplR,MASS,bioDist,preprocessCore

Author Dan Wang, Li Yan, Qiang Hu,Dominic J Smiraglia, Song Liu

Maintainer Dan Wang<dan.wang@roswellpark.org>

Description IMA(Illumina Methylation Analyzer) is a package designed
to automate the pipeline for analyzing site-level and
region-level methylation changes in epigenetic studies
utilizing the 450K DNA methylation microarray.

License GPL

LazyLoad yes

R topics documented:

IMA-package *IMA(Illumina Methylation Analyzer) is a package designed to automate the pipeline for analyzing site-level and region-level methylation changes in epigenetic studies utilizing the 450K DNA methylation microarray*

Description

IMA automates the tasks commonly required for the differential analysis of epigenetic data sets utilizing the 450K DNA methylation microarray. The package makes use of Illumina methylation annotation for region definition, as well as several Bioconductor packages for various preprocessing and differential testing steps. Written in open source R environment, it provides the flexibility for users to adopt, extend and customize the functionality for their specific needs. It can be used as an automatic pipeline to analyze specific regions as well as specific sites for downstream functional exploration and hypothesis generation.

Details

Package:	IMA
Type:	Package
Version:	1.0
Date:	2011-08-25
License:	GPL
LazyLoad:	yes

Author(s)

Dan Wang, Li Yang, Qiang Hu,Dominic J Smiraglia,Song Liu Maintainer: Dan Wang <wangdan412@gmail.com>

See Also

IMA.methy450R,IMA.methy450PP,regionswrapper,testfunc,sitetest,indexregionfunc

Examples

```
###Users specify the data paths/paramters in the pipeline.R before
#run the comman below
###R --no-save < pipeline.R##This will automately produce the result
#of methylation change
###the pipeline.R file could be found here
###http://ima.r-forge.r-project.org/
```

IMA.methy450PP *Data preprocessing and quality control*

Description

It allows user to choose several filtering step or modify filtering criteria for specific quality control purpose. These include whether or not to filter probes based on detetion Pvalue; whether or not to remove the loci from the X chromosome; whether or not to transfer the raw β vlaue using square arcsine; whether or not to perform normalization; whether or not to remove the loci containing missing β values. The user can choose the preprocessing routes and corresponding cutoffs in the argument of this function.

Usage

```
IMA.methy450PP(data,na.omit = TRUE,normalization=FALSE,transfm = FALSE,
samplefilterdetectP = c(FALSE,1e-5),samplefilterperc = 0.75,
sitefilterdetectP = c(FALSE,0.05),sitefilterperc = 0.75,
locidiff = c(FALSE,0.01), Xchrom = TRUE)
```

Arguments

data	an exprmethy450 class returned by the lumi.methy450R function
na.omit	if TRUE remove the sites containing missing value
normalization	if TRUE, quantile normalization performed
transfm	if TRUE, arcsin square root of beta matrix is performed
samplefilterdetectP	Default is False, i.e, no sample filtering.Otherwise, chose the cut off of detection Pvalue.
samplefilterperc	Remove the sample having over specified percentage of sites with detection Pvalue greater than **samplefilterdetectP**.
sitefilterdetectP	Default is False, i.e. no sites filtering by detection pvalue. Otherwise,chose the cut off of detection Pvalue.
sitefilterperc	Remove the sites having specified percentage of samples with detection Pvalue greater than **sitefilterdetectP**.
locidiff	if FALSE, don't filter sites by the difference of group β value. Otherwise, remove the sites has difference β value greater than specited value.
Xchrom	if TRUE, Remove the sites on chromosome X

Details

It allows user to choose several filtering step or modify filtering criteria for specific quality control purpose. By default, IMA will filter out loci with missing β value, from the X chromosome or with median detection P-value greater than specified value across the samples. The option for sample level quality control is also provided (B. C. Christensen et al., 2011). Although the raw β values will be analyzed as recommended by Illumina, the user can choose Arcsine squareroot transformation. Quantile normalization is also available for cross sample normalization.

Value

This function will return a methy450batch class including:

`bmatrix`	a matrix of β value for individual sites
`detectP`	a matrix of detection pvalue for individual sites
`annot`	a matrix of annotation information for individual sites
`groupinfo`	a list of sample ID and penotype of each sample
`TSS1500Ind`	a list of sites ID belonging to the TSS1500 region of each gene
`TSS200Ind`	a list of sites ID belonging to the TSS200 region of each gene
`UTR5Ind`	a list of sites ID belonging to the 5' UTR region of each gene
`EXON1Ind`	a list of sites ID belonging to the 1st EXON of each gene
`UTR3Ind`	a list of sites ID belonging to the 3' UTR region of each gene
`GENEBODYInd`	a list of sites ID belonging to the gene body region of each gene
`ISLANDInd`	a list of sites ID belonging to the ISLAND region of each UCSC_CPG_ISLAND
`NSHOREInd`	a list of sites ID belonging to the N Shore region of each UCSC_CPG_ISLAND
`SSHOREInd`	a list of sites ID belonging to the S Shore region of each UCSC_CPG_ISLAND
`NSHELFInd`	a list of sites ID belonging to the N Shelf region of each UCSC_CPG_ISLAND
`SSHELFInd`	a list of sites ID belonging to the S Shelf region of each UCSC_CPG_ISLAND

Author(s)

Dan Wang, Li Yang, Qiang Hu, Dominic J Smiraglia, Song Liu

Maintainer: Dan Wang <wangdan412@gmail.com>

See Also

`IMA.methy450R`,`regionswrapper`,`testfunc`,`sitetest`,`indexregionfunc`

Examples

```
setwd(system.file("extdata", package="IMA"))
MethyFileName = "SampleMethFinalReport.txt"
PhenoFileName = "SamplePhenotype.txt"
data = IMA.methy450R(file = MethyFileName,columnGrepPattern=list(beta=".AVG_Beta",detectp=".
dataf = IMA.methy450PP(data,na.omit = TRUE,normalization=FALSE,transfm =FALSE,samplefilterde
#This command will remove the sites have following characters:
```

```
#1.containg miss value
#2.with median detection pvalue great than 0.05
#3.on X chromsome
#This command will also remove the samples with 75% sites having detection pvalue greater th
```

`IMA.methy450R` *Load methylation 450k data*

Description

This function will load the methylation 450k data. The input for the package consists basically of two files containing value methylation data (including annotation) produced by BeadStudio or GenomeStudio, and sample phenotype data prepared by the user. *IMA.methy450R* load the input files with a single command described below and a *exprmethy450* object will be created, which includes the following features: β value matrix, locus annotion, detection Pvalue, sample phenotype information. Besides, baisc quality control information will be outputed in the QC.pdf, which include unsupervised sample clustering using all loci, boxplot for value of each sample, and barplot showing the percent of loci with detection Pvalue smaller than 1e-5 in each sample.

Usage

```
IMA.methy450R(file , columnGrepPattern = list(beta =".AVG_Beta"
, detectp = ".Detection.Pval"), groupfile)
```

Arguments

`file`	This is the input file containing 450k methylation data including β value, detection pvalue and annotation information. The file can be be produced from the illumina BeadStudio or GenomeStudio software.
`columnGrepPattern`	Specify the columns with corresponding characters from the file produced by llumina BeadStudio or GenomeStudio software will be used.
`groupfile`	This is the input file containing the penotype data.

Value

This function will return an *exprmethy450* class including these features: a β value matrix, an annotation matrix, a detection pvalue matrix and a list of sample phenotype. A QC.pdf file will be output as well which including the basic quality control infromation.

`bmatrix`	β value matrix for each site in each sample
`detectP`	a matrix of detection pvalue for each ~450k site on each sample
`annot`	a matrix of annotation for each ~450k site
`groupinfo`	a list of sample ID and penotype for each sample

Author(s)

Dan Wang, Li Yang, Qiang Hu,Dominic J Smiraglia,Song Liu

Maintainer: Dan Wang <wangdan412@gmail.com>

See Also

IMA.methy450PP,regionswrapper,testfunc,sitetest,indexregionfunc

Examples

```
setwd(system.file("extdata", package="IMA"))
MethyFileName = "SampleMethFinalReport.txt"
PhenoFileName = "SamplePhenotype.txt"
data = IMA.methy450R(file = MethyFileName,columnGrepPattern=list(beta=".AVG_Beta",detectp=".
```

---

indexregionfunc *Region level Methylation Index Calculation*

Description

For each specific region of a gene, IMA will collect the loci within it and derive an index of overall region methylation value. Currently, there are three different index metrics implemented in IMA: mean, median, and Tukey's Biweight robust average. By default, the median beta values will be used as the region's methylation index for further analysis.

Usage

```
indexregionfunc(indexlist , beta, indexmethod=c("mean","median","tbrm"))
```

Arguments

indexlist	specify which annotated region you are interested
beta	β matrix for the ~450k sites
indexmethod	The methods available to derive the index of overall region methylation value for each region. The options are mean, median, and Tukey's Biweight robust average.

Value

return a matrix of the index of overall methylation value for each annotated region

Author(s)

Dan Wang; Li Yan; Qiang Hu;Dominic J Smiraglia; Song Liu Maintainer: Dan Wang <wangdan412@gmail.com>

References

https://r-forge.r-project.org/projects/ima/

Examples

```
#data(fullannot)
#indlists = c("BRCA1", "MLH1", "CCNE1", "PTEN", "PALB2")
#index = match("TSS1500Ind",names(fullannot))
#annot = fullannot[[index]]
#index = match(indlists,names(annot))
#indexlist = annot[index]
#beta = data@bmatrix;
#eset = indexregionfunc(indexlist,beta,"mean");
#testfunc(eset,concov = "OFF",testmethod="limma",Padj="BH",gcase ="g2",gcontrol="g1",grouple
```

`outputDMfunc` *Output the differentially methylated sites/regions*

Description

This function will output the testing result with user-specified significance criteria.

Usage

```
outputDMfunc(out, rawpcut = 0.05, adjustpcut = 0.05, betadiffcut = 0.14)
```

Arguments

`out`	this is the output by function of sitetest, testfunc or regionwrapper with three columns for each site/region: pvalue, adjusted pvalue and β difference.
`rawpcut`	either null or define the cut off of the raw pvalue
`adjustpcut`	either null or define the cut off of the adjusted pvalue
`betadiffcut`	either null or define the cut off of the β difference between two groups.

Value

return a matrix with "P-Value", "Adjust Pval","beta-Difference" in seperate columns for each site/region

Author(s)

Dan Wang, Li Yang, Qiang Hu,Dominic J Smiraglia,Song Liu Maintainer: Dan Wang <wangdan412@gmail.com>

See Also

`IMA.methy450R`,`regionswrapper`,`IMA.methy450PP`,`testfunc`,`indexregionfunc`

Examples

```
#sitetest = sitetest(dataf,gcase=gcase,gcontrol=gcontrol,test = testmethod,Padj=Padj,outputD
#sitetest = outputDMfunc(sitetest,rawpcut = 0.05,adjustpcut =0.05,betadiffcut = 0.14)
```

`regionswrapper` *Differential methylation test on all 11 annotated regions*

Description

This function will output all the differential methylation test results for the 11 annotated regions, and store the result in separated sheets of an excel file.

Usage

```
regionswrapper(dataf, indexmethod = c("mean", "median", "tbrm"), gcase = "g2",
gcontrol = "g1",testmethod = c("wilcox", "limma", "pooled", "satterthwaite"),
Padj = "BH", concov = c("OFF", "ON"),list11excel,list11Rdata, rawpcut = NULL,
adjustpcut = NULL, betadiffcut = NULL)
```

Arguments

`dataf`	a methy450batch class returned by the lumi.methy450PP function
`indexmethod`	the methods available to derive an index of overall methylation level for each region
`gcase`	the string name of case group in the sample.txt file
`gcontrol`	the string name of control group in the sample.txt file
`testmethod`	Wilcoxon rank-sum test (default), Student's t-test ("pooled" or "satterthwaite") or empirical Bayes statistics (G. K. Smyth, 2004).
`Padj`	the method available for multiple testing correction. Users could choose any method provided in the p.adjust function of R stat package.
`concov`	if "ON", covariates is continuous variable, and the robust linear regression would be used for testing the association between methylation level and continuous phenotye (e.g., age).
`list11excel`	the name of output excel file containing the differential testing result for each annotated region
`list11Rdata`	the name of Rdata containing the differential testing result for each annotated region
`rawpcut`	either null or define the cut off of the raw pvalue
`adjustpcut`	either null or define the cut off of the adjusted pvalue
`betadiffcut`	either null or define the cut off of the beta difference between two groups.

Value

Return an excel file containing the differential testing result for each annotated region and a Rdata containing the differential testing result for each annotated region

Author(s)

Dan Wang, Li Yang, Qiang Hu,Dominic J Smiraglia,Song Liu

Maintainer: Dan Wang <wangdan412@gmail.com>

See Also

IMA.methy450PP,IMA.methy450R,testfunc,sitetest,indexregionfunc

Examples

```
setwd(system.file("extdata", package="IMA"))
MethyFileName = "SampleMethFinalReport.txt"
PhenoFileName = "SamplePhenotype.txt"
data = IMA.methy450R(file = MethyFileName,columnGrepPattern=list(beta=".AVG_Beta",detectp=".
dataf = IMA.methy450PP(data,na.omit = TRUE,normalization=FALSE,transfm = FALSE,samplefilterd
regionswrapper(dataf,indexmethod ="mean",gcase = "g2",gcontrol="g1",testmethod = "wilcox",Pa
```

---

sitetest *Site level Differential Methylation Analysis*

Description

For each specific site, Wilcoxon rank-sum test (default), Student's t-test ("pooled" or"satterthwaite") and empirical Bayes statistics (G. K. Smyth, 2004) are available for inference in differential testing. Robust linear regression is available as an option to infer methylation change associated with continuous variable (e.g., age). A variety of multiple testing correction algorithms is available, including stringent Bonferroni correction and widely used false discovery rate control. Users can specify the significance criteria in parameter file.

Usage

```
sitetest(dataf, gcase = "g2", gcontrol = "g1", testmethod = c("wilcox",
"limma", "pooled","satterthwaite"), Padj = "BH", concov = "OFF",
rawpcut = NULL, adjustpcut = NULL, betadiffcut = NULL)
```

Arguments

dataf	a methy450batch class returned by the lumi.methy450PP function
gcase	the string name of case group in the sample.txt file
gcontrol	the string name of control group in the sample.txt file
testmethod	Wilcoxon rank-sum test (default), Student's t-test ("pooled" or "satterthwaite"), or empirical Bayes statistics (G. K. Smyth, 2004).
Padj	The methods available for multiple testing correction. Users could choose any method provided in the p.adjust function of R stat package.

concov	if "ON", covariates is continuous variable, and the robust linear regression would be used for testing the association between methylation level and continuous phenotye (e.g., age).
rawpcut	either null or define the cut off of the raw pvalue
adjustpcut	either null or define the cut off of the adjusted pvalue
betadiffcut	either null or define the cut off of the beta difference between two groups.

Value

return a matrix with "P-Value", "Adjust Pval","beta-Difference" in seperate columns for each site

Author(s)

Dan Wang, Li Yang, Qiang Hu,Dominic J Smiraglia,Song Liu Maintainer: Dan Wang <wangdan412@gmail.com>

See Also

IMA.methy450R,regionswrapper,IMA.methy450PP,testfunc,indexregionfunc

Examples

```
#sitetest = sitetest(dataf,gcase=gcase,gcontrol=gcontrol,testmethod = testmethod,Padj=Padj,r
```

---

testfunc *Differential methylation testing function*

Description

Choose the testing function for differential methylation inference

Usage

```
testfunc(eset, concov = c("ON", "OFF"), testmethod = c("wilcox", "limma",
"pooled", "satterthwaite"),Padj = c("holm", "hochberg", "hommel", "bonferroni",
"BH", "BY", "fdr", "none"), grouplev, gcase = "g1",gcontrol = "g2")
```

Arguments

eset	a beta matrix
concov	if "ON", covariates is continuous variable,"grouplev" should be the continuous observation.
testmethod	Wilcoxon rank-sum test (default), Student's t-test ("pooled" or "satterthwaite"), or empirical Bayes statistics (G. K. Smyth, 2004).
Padj	The methods available for multiple testing correction. Users could choose any method provided in the p.adjust function of R stat package.

grouplev	specify the vector for covariate:if concov is "ON", this should be the continuous values, if concov is "OFF", this should be categorical values.
gcase	the string name of case group in the sample.txt file
gcontrol	the string name of control group in the sample.txt file

Value

return a matrix with "P-Value", "Adjust Pval","beta-Difference" in seperate columns for each site or annotated region

Author(s)

Dan Wang, Li Yang, Qiang Hu,Dominic J Smiraglia,Song Liu

Maintainer: Dan Wang <wangdan412@gmail.com>

See Also

IMA.methy450R,regionswrapper,IMA.methy450PP,sitetest,indexregionfunc

Examples

```
#data("fullannotInd")
#indlists = c("BRCA1", "MLH1", "CCNE1", "PTEN", "PALB2")
#annot = fullannot[[match("TSS1500Ind",names(fullannot))]]
#indexlist = annot[match(indlists,names(annot))]
#eset = indexregionfunc(indexlist,data@bmatrix,"mean");
#testfunc(eset,concov = "OFF",testmethod="limma",Padj="BH",gcase ="g2",gcontrol="g1",
#grouplev=grouplev)
```

Index