Package 'IMA'

September 5, 2011

Type Package

Title IMA:Illumina Methylation Analyzer

Version 1.0

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 auto-mate the pipeline for analyzing site-level and
region-level methylation changes in epigenetic studies
utilizing the 450K DNA methylation microarray.

License GPL

LazyLoad yes

Repository R-forge

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,sumregionfun

Examples

```
###Users specify the data paths/paramters in the pipeline.R before
#run the comman below
###R --no-save < pipeline.R##This will auto-mate to produce the result
#of methylation
###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, filter
```

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
`filtersample`	Keep the samples with detection Pvalue <10e-5 on more than 75% sites
`filterdetectP`	Remove the sites with detection Pvalue > 0.01 across detectPcut percentage of the samples
`detectPcut`	Remove the sites with detection Pvalue > 0.01 across 75% of the samples
`locidiff`	if TRUE, keep the sites has big variance between two groups.
`locidiffcut`	if locidiff = TRUE, keep the sites has variance great than the locidiffcut between two groups
`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 beta value, from the X chromosome or with median detection P-value greater than 0.05 across the samples. The option for sample level quality control is also provided (B. C. Christensen et al., 2011). Although the raw beta values will be analyzed as recommended by Illumina, the user can choose Arcsine squareroot transformation. Both scaling and quantile normalization options are available for cross sample normalization.

Value

This function will return a methy450batch class including:

bmatrix	a matrix of beta 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 istes ID belonging to the ISLAND region of each UCSC_CPG_ISLAND
NSHOREInd	a list of istes ID belonging to the N Shore region of each UCSC_CPG_ISLAND
SSHOREInd	a list of istes ID belonging to the S Shore region of each UCSC_CPG_ISLAND
NSHELFInd	a list of istes ID belonging to the N Shelf region of each UCSC_CPG_ISLAND
SSHELFInd	a list of istes 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,sumregionfun

Examples

```
#dataf =IMA.methy450PP(data,na.omit=na.omit,normalization=normalization,
#transfm = transfm,filterdetectP = filterdetectP, detectPcut =detectPcut, Xchrom = xchrom)
#About 2 mins for preprocessing ~450k sites
```

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,sumregionfun

Examples

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

---

outputDESfunc *Output the differential expressed sites/regions*

Description

This function will subset the testing result by using the user specified cut off.

Usage

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

Arguments

out	this is the return results by function of sigtest, testfunc,regionwrapper with three columns: pvalue,adjusted pvalue and beta difference.
rawpcut	define the cut off of the raw pvalue
adjustpcut	define the cut off of the adjusted pvalue
betadiffcut	define the cut off of the β difference between two groups.

Value

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

Author(s)

Dan Wang, Li Yang, Qiang Hu,Dominic J Smiraglia,Song Liu Maintainer: Dan Wang <wang-dan412@gmail.com>

See Also

IMA.methy450R,regionswrapper,IMA.methy450PP,testfunc,sumregionfun

Examples

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

`regionswrapper` *Differentially test on the 11 annotated region-level analysis separately*

Description

this function will output all the statistical results for the 11 annotated region-level analysis result in separated sheet of excel file, and the result could be used for further analysis such as GO functional annotation, pathway analysis, network etc. This function will took up to 2 hours to run.

Usage

```
regionswrapper(dataf, sumregion = c("mean", "median", "tbrm"), gcase = "g2", gcontr
```

Arguments

`dataf`	a methy450batch class returned by the lumi.methy450PP function
`sumregion`	the methods can be used to summarize the targed sites 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 and empirical Bayes statistics (G. K. Smyth, 2004).
`Padj`	the method applied for multiple testing correction.users could choose any method in the p.adjust function such as FDR, BH etc.
`concov`	if "ON", covariates is continuous variable, then the robust linear regression would be used for testing.
`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
`outputDES`	if outputDES = TRUE, only out the differential methylated sites by using defined cut off
`rawpcut`	only used when outputDES = TRUE, and define the cut off of the raw pvalue
`adjustpcut`	only used when outputDES = TRUE, and define the cut off of the adjusted pvalue
`betadiffcut`	only used when outputDES = TRUE, and 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 <wang-dan412@gmail.com>

See Also

IMA.methy450PP,IMA.methy450R,testfunc,sitetest,sumregionfun

Examples

```
#regionswrapper(dataf,sumregion=sumregion,gcase=gcase,gcontrol=gcontrol,
#testmethod=testmethod,Padj=Padj,concov=concov,list14excel="../output/list14result.xls")
```

---

sitetest *Differential Methylation Analysis*

Description

For each specific region, Wilcoxon rank-sum test (default), Student's t-test 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. The same statistical inference and multiple test correction procedures described above will also be applied to each single site to obtain site-level differential methylation inference.

Usage

```
sitetest(dataf, gcase = "g2", gcontrol = "g1", testmethod = c("wilcox", "limma", "p
```

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 and empirical Bayes statistics (G. K. Smyth, 2004).
Padj	specify which method applied for multiple testing correction.
concov	if "ON", covariates is continuous variable
rawpcut	define the cut off of the raw pvalue
adjustpcut	define the cut off of the adjusted pvalue
betadiffcut	define the cut off of the beta difference between two groups.

Value

return a matrix with "P-Value", "Adjust Pval","beta-Difference" in 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,sumregionfun

Examples

```
#sitetest = sitetest(dataf,gcase=gcase,gcontrol=gcontrol,test = testmethod,Padj=Padj,outputD
#sitetest = outputDESfunc(sitetest,outputDES = TRUE,rawpcut = 0.05,adjustpcut =0.05,betadiff
```

---

sumregionfun *Methylation Index Calculation*

Description

for each specific region (e.g., promoter), 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

```
sumregionfun(indexlist , beta, sumregion)
```

Arguments

indexlist	specify which annotated region you are interested
beta	beta matrix for each ~450k site
sumregion	The method derive the overall region methylation value for each gene or UCSC_CPG_ISLAND. The options are mean, median, and Tukey's Biweight robust average.

Value

return a summarized matrix for the 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 = sumregionfun(indexlist,beta,"mean");
#testfunc(eset,concov = "OFF",testmethod="limma",Padj="BH",gcase ="g2",gcontrol="g1",grouple
```

`testfunc` *Differential testing function*

Description

pick up the test for differential methylation testing

Usage

```
testfunc(eset, concov = c("ON", "OFF"), testmethod = c("wilcox", "limma", "pooled",
```

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 and empirical Bayes statistics (G. K. Smyth, 2004).
`Padj`	specify which method applied for multiple testing correction.
`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`	specify the string name of phenotype in the case group
`gcontrol`	specify the string name of phenotype in the case group

Value

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

Examples

```
#data(fullannotInd)
#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 = sumregionfun(indexlist,beta,"mean");
#testfunc(eset,concov = "OFF",testmethod="limma",Padj="BH",gcase ="g2",gcontrol="g1",grouple
```

Index