Package 'LncDM'

July 16, 2016

Type Package

Title The Different Methylation Sites, Elements and Regions of lncRNA

Version 1.0

Date 2016-06-21

Author Hui Zhi,Yanxun Su,Xin Li

Maintainer Hui Zhi<zhihui013201@gmail.com>

Description The package can get the methylation matrix and identify the differential methylation sites,elements or the regions who are in the elements of lncRNA etc. This package gets a matrix of methylation values of the samples. Then it can calculate the methylation values of elements. This package uses linear and t test to identify the differential methylation or the sites,elements and regions that related with the phenotype.

Collate loaddata.R dms.R regionLevel.R dme.R dmr.R

License GPL-2

Depends R (>= 3.0.1), methods

Imports beanplot, reshape, gplots, WriteXLS, MASS, impute, limma, preprocessCore, grDevices, graphics, stats, utils

LazyLoad yes

biocViews Statistics, Methylation, DMR, LncRNA

NeedsCompilation no

R topics documented:

dme *Find the differential methylation elements of lincRNA, protein coding gene ,processed transcript and pseudogene*

Description

Find the differential methylation elements or the elements that related with the phenotype. The elements are belong to lincRNA, protein coding gene, processed transcript and pseudogene.

Usage

```
dme(data,classes=c("lincRNA","gene","processed_transcript","pseudogene"),contin=c(
"ON","OFF"),testmethod = c("wilcox","limma", "t.test", "satterthwaite"), Padj = c(
"holm", "hochberg", "hommel", "bonferroni", "BH", "BY", "fdr", "none"), gcase =
"case", gcontrol = "control", paired = FALSE,rawpcut = 0.05, adjustpcut = 0.05,
 betadiffcut = 0.14,num)
```

Arguments

data	The objects of class LincMethy450 which return from loaddata. The beta matrixs of sites. A site per row and A sample per column.
classes	Whose element will be calculated.
contin	If phenotype is continuous,contin is 'ON',use linear regression to find the elements that related with the phenotype.
testmethod	The method to do the test to find dme while contin is 'OFF' which means phenotype is discontinuous.
Padj	The method of multiple testing adjustment to adjust P value.
gcase	The name of case group while contin is 'OFF'.
gcontrol	The name of case group while contin is 'OFF'.
paired	Whether compare in pairs while do t.test.
rawpcut	It is the threshold for cutting raw P value.
adjustpcut	It is the threshold for cutting adjust P value.
betadiffcut	The minimum differential between two groups' means while contin is 'OFF'.
num	The number which is the parameter of of elements to plot.

Details

dme is designed to find differential methylation elements or the transcripts' elements that related with the continuous phenotype. If contin is 'ON', it means the phenotype is continuous, and linear regression will be used. If the phenotype isn't continuous, test such as t test will be used.

Value

dme will return two excel files that one contains the transcripts' elements whose P value less than rawpcut, adjust P less than adjustpcut and the differ of the means of two groups more than betadiffer, while another is the beta matrix of these significant elements. There are box plot for most significative elements and heat map all significative elements.

Author(s)

Hui Zhi<zhihui013201@gmail.com>,Yanxun Su<hmu_yanxunsu@163.com>,Xin Li<lixin920126@163.com>

See Also

See Also dms and dmr

Examples

```
Dir <- system.file("extdata/localdata",package="LncDM")
dir.create(paste(Dir,"/dme",sep=""))
###user can set the dir of their own
setwd(paste(Dir,"/dme",sep=""))
###load the result of loaddata()
load(paste(Dir,"/loadData.Rdata",sep=""))
###dme is based on the result of the regionLevel()
Region <- regionLevel(data=loadData,indexmethod = "mean",classes="lincRNA")
dme(data=Region,classes="lincRNA",contin="OFF",testmethod = "t.test", Padj = "fdr",
gcase = "case", gcontrol = "control", paired = FALSE,rawpcut = 0.05, adjustpcut = 0.05,
betadiffcut = 0.3,num=1)
```

---

dmr *Find the differential methylation regions in the elements of lincRNA, protein coding gene ,processed transcript and pseudogene*

Description

Find the differential methylation regions or the regions that related with the phenotype who are in the elements of lincRNA etc.

Usage

```
dmr(data,contin=c("ON","OFF"),classes=c("lincRNA","gene","processed_transcript",
"pseudogene"),testmethod = c("wilcox","limma", "t.test", "satterthwaite"), Padj =
c("holm", "hochberg", "hommel", "bonferroni", "BH", "BY", "fdr", "none"), gcase =
"case", gcontrol = "control", paired = FALSE,rawpcut = 0.05, adjustpcut = 0.05,
 betadiffcut = 0.3,num,sole=FALSE)
```

Arguments

data	The objects of class LincMethy450 which return from loaddata. The beta matrixs of sites. A site per row and A sample per column.
classes	Whose regions will be calculated.
contin	If phenotype is continuous,contin is 'ON',and use linear regression to find the regions that related with the phenotype.
testmethod	The method to do the test to find dmr while contin is 'OFF' which means phenotype is discontinuous.
Padj	The method of multiple testing adjustment to adjust P value.
gcase	The name of case group while contin is 'OFF'.
gcontrol	The name of case group while contin is 'OFF'.

paired	Whether compare in pairs while do t.test.
rawpcut	It is the threshold for cutting raw P value.
adjustpcut	It is the threshold for cutting adjust P value.
betadiffcut	The minimum differential between two groups' means while contin is 'OFF'.
num	The number which is the box plot of different methylation regions to plot.
sole	Whether keep the no duplicate records.

Details

dmr is designed to find differential methylation regions or the regions in the elements of lincRNA etc. that related with the continuous phenotype. If contin is 'ON', it means the phenotype is continuous, and linear regression will be used. If the phenotype isn't continuous, test such as t test will be used.

Value

dmr will return two txt files that one contains the transcripts' elements and the region of chromosome whose P value less than rawpcut, adjust P less than adjustpcut and the differ of the means of two groups more than betadiffer, while another is the beta matrix of these significant regions. There are box plot for most significative regions and heat map all significative regions.

Author(s)

Hui Zhi<zhihui013201@gmail.com>,Yanxun Su<hmu_yanxunsu@163.com>,Xin Li<lixin920126@163.com>

See Also

See Also dms and dme

Examples

```
Dir <- system.file("extdata/localdata",package="LncDM")
dir.create(paste(Dir,"/dmr",sep=""))
###user can set the dir of their own
setwd(paste(Dir,"/dmr",sep=""))
###load the result of loaddata()
load(paste(Dir,"/loadData.Rdata",sep=""))
dmr(data=loadData,contin="OFF",classes="lincRNA",testmethod = "t.test", Padj = "fdr",
gcase = "case", gcontrol = "control", paired = FALSE,rawpcut = 0.05, adjustpcut = 0.05,
betadiffcut = 0.3,num=1,sole=FALSE)
```

---

dms *Find the differential methylation CpG sites in the lincRNA, protein coding gene ,processed transcript and pseudogene*

Description

Find the differential methylation sites or the sites that related with the phenotype who are in the lincRNA etc.

Usage

```
dms(data,contin=c("ON","OFF"),classes=c("lincRNA","gene","processed_transcript",
"pseudogene"),testmethod = c("wilcox","limma", "t.test", "satterthwaite"), Padj =
c("holm", "hochberg", "hommel", "bonferroni", "BH", "BY", "fdr", "none"), gcase =
"case", gcontrol = "control", paired = FALSE,rawpcut = 0.05, adjustpcut = 0.05,
 betadiffcut = 0.3,XY=c(FALSE,"X","Y",c("X","Y")),tlog=FALSE,num)
```

Arguments

data	The objects of class LincMethy450 which return from loaddata. The beta matrixs of sites. A site per row and A sample per column.
classes	Whose CoG sites will be calculated
contin	If phenotype is continuous,contin is 'ON',and use linear regression to find the sites that related with the phenotype.
testmethod	The method to do the test to find dms while contin is 'OFF' which means phenotype is discontinuous.
Padj	The method of multiple testing adjustment to adjust P value.
gcase	The name of case group while contin is 'OFF'.
gcontrol	The name of case group while contin is 'OFF'.
paired	Whether compare in pairs while do t.test.
rawpcut	The threshold for cutting raw P value.
adjustpcut	The threshold for cutting adjust P value.
betadiffcut	The minimum differential between two groups' means while contin is 'OFF'.
num	The number of sites,elements or regions to plot.
XY	The chromosomes that where CpG sites were filtered.
tlog	Whether do the log transform for the P values.

Details

dms is designed to find differential methylation sites or the CpG sites that related with the continuous phenotype. If contin is 'ON', which means the phenotype is continuous, and linear regression will be used. If the phenotype isn't continuous, test such as t test will be used.

Value

dms will return two txt files that one contains the sites whose P value less than rawpcut, adjust P less than adjustpcut and the differ of the means of two groups more than betadiffer while another is the beta matrix of dms. There are box plot for most significative sites,heat map and Manhattan of all significative sites.

Author(s)

Hui Zhi<zhihui013201@gmail.com>,Yanxun Su<hmu_yanxunsu@163.com>,Xin Li<lixin920126@163.com>

See Also

See Also dme and dmr

Examples

```
Dir <- system.file("extdata/localdata",package="LncDM")
dir.create(paste(Dir,"/dms",sep=""))
###user can set the dir of their own
setwd(paste(Dir,"/dms",sep=""))
###load the result of loaddata()
load(paste(Dir,"/loadData.Rdata",sep=""))
dms(data=loadData,contin="OFF",classes="lincRNA",testmethod = "t.test", Padj = "fdr",
gcase = "case", gcontrol = "control", paired = FALSE,rawpcut = 0.05, adjustpcut = 0.05,
betadiffcut = 0.3,XY=c(FALSE,"X","Y"),tlog=FALSE,num=1)
```

`loaddata` *Load and preprocessed raw data*

Description

Read the file of signal_intensities, calculate the beta value, filter the unqualified samples and sites. Plot the heat map, box plot, density plot and density bean plot of CpG sites, and bar plot for detect P-value of samples.

Usage

```
loaddata(fileDir,is_beta=FALSE,beta_method=c("M/(M+U)","M/(M+U+100)"),groupfile
,samplefilter = FALSE,contin=c("ON","OFF"),samplefilterperc = 0.75, XYchrom =
c(FALSE, "X","Y", c("X", "Y")),sitefilter = FALSE, sitefilterperc = 0.75,
 filterDecetP =0.05, normalization  = FALSE,transfm = c(FALSE, "arcsinsqr", "logit")
,snpfilter=c(FALSE,"within_10","prob_snp"),gcase="case",gcontrol ="control",skip=0
,imputation=c("mean","min","knn"),knn.k=10)
```

Arguments

`fileDir`	The folder name of samples' signal_intensities files.
`is_beta`	Logical. The signal_intensities is beta value or not.
`beta_method`	The method for calculating the beta.
`groupfile`	The name of phenotype file.
`samplefilter`	Logical. Filter the samples whose most detection P values aren't significative or not.
`contin`	'ON' means the phenotype is continuous,just like age etc. 'OFF' means it is discontinued.
`samplefilterperc`	A number in [0,1]. The samples whose percent of the significative detection P values less than this number will be filtered.
`XYchrom`	The CpG sites in X or Y chromosome should or shuoldn't be filtered.
`sitefilter`	Logical. Filter the sites whose most detection P values aren't significative or not.
`sitefilterperc`	A number in [0,1]. The sites whose percent of the significative detection P values less than this number will be filtered.
`filterDecetP`	Threshold: value of significative detection P. Always 0.05 or 0.01.

normalization	Logical. Normalization for the different chips or not.
transfm	Data transformation for beta or not. Contains 'arcsinsqr' and 'logit'.
snpfilter	The CpG sites that contain SNP sites with 10bp or 50bp shuold or shouldn't be filetered.
gcase	The name of case group while contin is 'OFF'.
gcontrol	The name of case group while contin is 'OFF'.
skip	integer: the number of lines of the data file to skip before beginning to read signal_intensities data, the first row must be signal values.
imputation	The method to fill the NA.Contains 'mean', 'min' and 'knn'.
knn.k	The K number if imputation is 'knn'.

Details

Loaddata is designed to load and process the methylated data for the package. It provides two methods to calculate the beta value,which means the ratio of methylation,M/(M+U) and M/(M+U+100),M means the intensity of methylation and U means the intensity of unmethylation. For the methylated data, a file per sample. In the signal_intensities file, there are four columns, CpG ID, Methylated_Intensity, Unmethylated_Intensity and Detection_P_value. The groupfile that explain the phenotype of samples. Distinguish the case or control. The samples that at the same group have the same label. The sample IDs are same as the names of corresponding signal_intensities file (without File Suffixes). Loaddata also call the other function to plot the heat map,box plot, density plot and density bean plot of CpG sites, and bar plot for detect P-value of samples.

Value

Loaddata will return an object of class LincMethy450. And return some plots to describe the information of data.

Author(s)

Hui Zhi<zhihui013201@gmail.com>,Yanxun Su<hmu_yanxunsu@163.com>,Xin Li<lixin920126@163.com>

See Also

See Also dms, dme and dmr

Examples

```
## Not run:
 ##the directory of phenotype and 450k methylation's sample data
 Dir <- system.file("extdata/localdata",package="LncDM")
 setwd(Dir)
 ###phenotype file's name
 groupfile <- "BRCA_pheno.txt"
 ###our methylation data in the subdirectory "Level_2" is just example data, when you
 ###run this function, please prepare complete sample files, and change default directory
 ###to yourself
 loadData <- loaddata(fileDir="Level_2",is_beta=FALSE,beta_method="M/(M+U)",groupfile=groupfile,
 samplefilter = TRUE,contin="OFF",samplefilterperc = 0.75,XYchrom = c(FALSE, "X","Y"),sitefilter = TRUE,
 sitefilterperc = 0.75,filterDecetP=0.05,normalization  = FALSE,transfm = FALSE,snpfilter=c(FALSE,"prob_snp"
 gcase="case",gcontrol="control",skip=2,imputation="knn",knn.k=10)
 ###save the loadData in order to caculate dms,dmr and dme
 save(loadData,file="loadData.Rdata",compress="xz")
```

```
## End(Not run)
```

`regionLevel` *Calculate the beta values of elements in lincRNA etc*

Description

Calculate the beta values of elements in lincRNA etc. based on reannotation information.

LincRNA and processed transcript's elements are TSS1500, TSS200, 1_exon, genebody and intron. Protein coding gene and pseudogene transcript's elements are 5'UTR, 3'UTR, TSS1500, TSS200, 1_exon, genebody and intron.

Usage

```
regionLevel(data,indexmethod = c("mean", "median"),classes=c("gene","lincRNA",
"processed_transcript","pseudogene"))
```

Arguments

`data`	The objects of class LincMethy450 which return from `loaddata`. The beta matrix of sites. A site per row and A sample per column.
`indexmethod`	The method to calculate the beta if there are more than two CpG sites in a element.
`classes`	Whose CpG sites will be calculated.

Details

The function, regionLevel, is designed to calculate the beta value of gene, lincRNA, processed_transcript and pseudogenes' elements. If indexmethod is "mean" ,and the mean of CpG sites' beta value will be regarded as the element's beta. If classes is "gene", only the elements of protein coding genes' transcript will be calculated. If want to calculate more than one class, please do it respectively.

Value

This methods will return a object of class RegionMethy450. It contains some matrix of elements' beta values. A row per transcript, a col per sample.

Author(s)

Hui Zhi<zhihui013201@gmail.com>,Yanxun Su<hmu_yanxunsu@163.com>,Xin Li<lixin920126@163.com>

See Also

See Also `dms`, `dme` and `dmr`

Examples

```
## Not run:
  Dir <- system.file("extdata/localdata",package="LncDM")
  setwd(Dir)
  ###load the result of loaddata()
  #load(paste(Dir,"/loadData.Rdata",sep=""))
  Region <- regionLevel(data=loadData,indexmethod = "mean",classes="lincRNA")
  ###save the region data in order to caculate dme
  save(Region,file="Region.Rdata")

## End(Not run)
```

Index