Package 'easyEWAS'

January 3, 2025

Title Perform and visualize EWAS analysis

Version 1.0.0

Description easyEWAS is an R package designed for conducting Epigenome-Wide Association Study (EWAS) and visualizing results. Users only need to provide sample data, and methylation values to easily perform EWAS analysis. The package supports two statistical methods,linear models and linear mixed-effects models. It utilizes the CMplot package to generate visualizations, including Manhattan plots and QQ plots, based on EWAS results. Additionally, the use of parallel computing significantly improves computational efficiency, making it suitable for researchers less experienced in EWAS or parallel computing.

License GPL (>= 3)

Encoding UTF-8

Roxygen list(markdown = TRUE)

RoxygenNote 7.3.2

Imports boot,
boot.pval,
CMplot,
ddpcr,
doParallel,
dplyr,
foreach,
lmerTest,
magrittr,
parallel,
stringr,
survival,
tictoc,
vroom,
writexl,
clusterProfiler,
org.Hs.eg.db,
ggplot2,
DMRcate,
sva,
R.utils,
BiocParallel

Suggests knitr,
rmarkdown

VignetteBuilder knitr

Depends R (>= 4.4.0)

LazyData true

Contents

batchEWAS 2
bootEWAS 3
dmrEWAS 4
enrichEWAS 5
initEWAS 6
loadEWAS 7
plotEWAS 8
startEWAS 11
transEWAS 12

Index **13**

`batchEWAS` *Perform batch effect correction*

Description

Perform batch effect correction based on the function `ComBat` form R package **sva**. It requires that the "batches" in the data set are known. It uses either parametric or non-parametric empirical Bayes frameworks for adjusting data for batch effects.

Usage

```
batchEWAS(input,adjustVar = NULL,batch = NULL, plot = TRUE, par.prior = TRUE,
mean.only = FALSE,ref.batch = NULL)
```

Arguments

`input`	An R6 class integrated with all the information.
`adjustVar`	(Optional) Names of the variate of interest and other covariates besides batch, with each name separated by a comma. Ensure that when correcting for batch effects, the effects of other factors are appropriately considered and adjusted for.Ensure there are no space. e.g. "cov1,cov2".
`batch`	Name of the batch variable.
`plot`	(Optional) TRUE give prior plots with black as a kernel estimate of the empirical batch effect density and red as the parametric.
`par.prior`	(Optional) TRUE indicates parametric adjustments will be used, FALSE indicates non-parametric adjustments will be used.
`mean.only`	(Optional) Default to FALSE. If TRUE ComBat only corrects the mean of the batch effect (no scale adjustment).
`ref.batch`	(Optional) NULL If given, will use the selected batch as a reference for batch adjustment.

Value

input, An R6 class object integrating all information.

Examples

```
## Not run:
res <- initEWAS(outpath = "default")
res <- loadEWAS(input = res, ExpoData = "default", MethyData = "default")
res <- transEWAS(input = res, Vars = "cov1", TypeTo = "factor")
res <- batchEWAS(input = res, batch = "batch", par.prior=TRUE, ref.batch = NULL)

## End(Not run)
```

---

bootEWAS *Perform Bootstrap-based Internal Validation*

Description

Users can perform internal validation of the identified differently methylated sites based on the bootstrap method.

Usage

```
bootEWAS(input, filterP = "PVAL", cutoff = 0.05, CpGs = NULL, times = 500,
bootCI = "perc",filename = "default")
```

Arguments

input	An R6 class integrated with all the information obtained from the startEWAS or plotEWAS function.
filterP	The name of the p value columns such as "PVAL", "FDR", and "Bonfferoni." Users use this P-value to screen for significance sites and further conduct internal validation.
cutoff	The cutoff value of the P-value used to filter for further internal validation. The default is 0.05.
CpGs	The name of the methylation site specified by the user for bootstrap analysis, separated by commas. Be careful not to have spaces, such as "cpg1,cpg2".
times	Number of bootstrap times specified by the user. The default value is 100 times.
bootCI	A vector of character strings representing the type of interval to base the test on. The value should be one of "norm", "basic", "stud", "perc" (the default), and "bca".
filename	User-customized .csv file name for storing bootstrap results. If "default", it will be named as "bootresult".

Value

input, An R6 class object integrating all information.

Examples

```
## Not run:
res <- initEWAS(outpath = "default")
res <- loadEWAS(input = res, ExpoData = "default", MethyData = "default")
res <- transEWAS(input = res, Vars = "cov1", TypeTo = "factor")
res <- startEWAS(input = res, chipType = "EPICV2", model = "lm", expo = "default", adjustP = TRUE)
res <- bootEWAS(input = res, filterP = "PVAL", cutoff = 0.05, times = 100)

## End(Not run)
```

`dmrEWAS` *Perform Differentially Methylated Region analysis*

Description

Perform differential methylation analysis based on the R package **DMRcate**. Computes a kernel estimate against a null comparison to identify significantly DMRs.

Usage

```
dmrEWAS(input, chipType = "EPICV2", what = "Beta", expo = NULL, cov = NULL, genome = "hg38",
lambda=1000, C = 2, filename = "default",fdrCPG = 0.05, pcutoff = "fdr", min.cpgs = 2,
epicv2Filter = "mean")
```

Arguments

`input`	An R6 class integrated with all the information.
`chipType`	The Illumina chip versions for user measurement of methylation data, including "450K","EPICV1", and "EPICV2". The default is "EPICV2".
`what`	Types of methylation values, including "Beta" and "M". Default to "Beta".
`epicv2Filter`	Strategy for filtering probe replicates that map to the same CpG site. "mean" takes the mean of the available probes; "sensitivity" takes the available probe most sensitive to methylation change; "precision" either selects the available probe with the lowest variation from the consensus value (most precise), or takes the mean if that confers the lowest variation instead, "random" takes a single probe at random from each replicate group.
`expo`	Name of the exposure variable used in the DMR analysis.
`cov`	Name(s) of covariate(s) used in the DMR analysis, with each name separated by a comma. Ensure there are no space. e.g. "cov1,cov2,cov3".
`genome`	Reference genome for annotating DMRs. Can be one of "hg19" or "hg38".
`fdrCPG`	Used to individually assess the significance of each CpG site. If the FDR-adjusted p-value of a CpG site is below the specified fdrCPG threshold, the site will be marked as significant. The default value is 0.05.
`pcutoff`	Used to determine the threshold for DMRs. It is strongly recommended to use the default (fdr), unless you are confident about the risk of Type I errors (false positives).
`lambda`	If the distance between two significant CpG sites is greater than or equal to lambda, they will be considered as belonging to different DMRs. The default value is 1000 nucleotides, meaning that if the distance between two significant CpG sites exceeds 1000 nucleotides, they will be separated into different DMRs.

C	Scaling factor for bandwidth. Gaussian kernel is calculated where lambda/C = sigma. Empirical testing shows for both Illumina and bisulfite sequencing data that, when lambda=1000, near-optimal prediction of sequencing-derived DMRs is obtained when C is approximately 2, i.e. 1 standard deviation of Gaussian kernel = 500 base pairs. Cannot be < 0.2.
min.cpgs	Minimum number of consecutive CpGs constituting a DMR. Default to 2.
filename	User-customized .csv file name for storing DMR results. If "default", it will be named as "DMRresult".

Value

input, An R6 class object integrating all information.

Examples

```
## Not run:
res <- initEWAS(outpath = "default")
res <- loadEWAS(input = res, ExpoData = "default", MethyData = "default")
res <- transEWAS(input = res, Vars = "cov1", TypeTo = "factor")
res <- dmrEWAS(input = res, filename = "default", chipType = "EPICV2", what = "Beta", expo = "var",
cov = "cov1,cov2", genome = "hg38",)

## End(Not run)
```

---

enrichEWAS *Enrichment analyses*

Description

Perform GO or KEGG enrichment analysis based on the **clusterProfiler** package.

Usage

```
enrichEWAS(input, filename = "default", method = "GO", ont = "MF", pool = FALSE,
filterP = "PVAL", cutoff = 0.05, plot = TRUE, plotType = "dot", plotcolor = "pvalue",
showCategory= NULL, pvalueCutoff = 0.05, pAdjustMethod = "BH", qvalueCutoff = 0.2)
```

Arguments

input	An R6 class integrated with all the information obtained from the startEWAS or plotEWAS or bootEWAS function.
filename	User-customized .xlsx file name for storing EWAS results. If "default" is chosen, it will be named as "enrichresult".
method	Methods of enrichment analysis, including "GO" and "KEGG".
filterP	The name of the p value columns such as "PVAL", "FDR", and "Bonfferoni." Users use this P-value to screen for significance sites and further conduct enrichment analysis.
cutoff	The cutoff value of the P-value used to filter for further enrichment analysis. The default is 0.05.

ont	When choosing GO enrichment analysis, select the GO sub-ontology for which the enrichment analysis will be performed. One of "BP", "MF", and "CC" sub-ontologies, or "ALL" for all three. Default to "BP".
pool	If ont='ALL', whether pool 3 GO sub-ontologies.
plot	Whether the results of enrichment analysis need to be visualized, the default is TRUE
plotType	Whether to draw a bar plot ("bar") or a dot plot ("dot"), the default is "dot".
plotcolor	It is the vertical axis of the picture of the enrichment analysis results. Users can choose "pvalue" or "p.adjust" or "qvalue". The default is "pvalue".
showCategory	The number of categories which will be displayed in the plots.
pvalueCutoff	The p-value threshold used to filter enrichment results. Only results that pass the p-value test (i.e., those smaller than this value) will be reported. This value refers to the p-value before adjustment. The p-value represents the probability of observing the current level of enrichment under the assumption of no enrichment. The smaller the p-value, the more significant the enrichment result.
pAdjustMethod	The p-value adjustment method used for multiple hypothesis testing, aimed at reducing false positives caused by multiple comparisons. One of "holm", "hochberg", "hommel", "bonferroni", "BH", "BY", "fdr", "none".
qvalueCutoff	qvalue cutoff on enrichment tests to report as significant. The q-value is the result of controlling the false discovery rate (FDR) and represents the proportion of false positives that may occur when conducting multiple tests.Tests must pass i) pvalueCutoff on unadjusted pvalues, ii) pvalueCutoff on adjusted pvalues and iii) qvalueCutoff on qvalues to be reported. The default is 0.2.

Value

input, An R6 class object integrating all information.

Examples

```
## Not run:
res <- initEWAS(outpath = "default")
res <- loadEWAS(input = res, ExpoData = "default", MethyData = "default")
res <- transEWAS(input = res, Vars = "cov1", TypeTo = "factor")
res <- startEWAS(input = res, chipType = "EPICV2", model = "lm", expo = "default", adjustP = TRUE)
res <- plotEWAS(input = res, pval = "PVAL")
res <- bootEWAS(input = res, filterP = "PVAL", cutoff = 0.05, times = 100)
res <- enrichEWAS(input = res, method = "GO", filterP = "PVAL", cutoff = 0.05, pAdjustMethod = "BH")

## End(Not run)
```

---

initEWAS *Initialize the EWAS module*

Description

This function is designed to generate an R6 class for storing all the data and results of EWAS analysis, and also to create a folder on the local computer for storing the analysis results.

Usage

```
initEWAS(outpath = "default")
```

Arguments

outpath	The user-specified path is used to store a generated folder named "EWASresult" that contains all the analysis results. If "default" is specified, the folder will be generated in the current working directory.

Value

input, an R6 class object integrating all information.

Examples

```
## Not run:
res <- initEWAS(outpath = "default")

## End(Not run)
```

---

loadEWAS *Load all data files for EWAS module*

Description

Upload sample data and methylation data for EWAS analysis.

Usage

```
loadEWAS(input, ExpoPath = NULL, MethyPath = NULL, ExpoData = "default",
MethyData = "default")
```

Arguments

input	An R6 class integrated with all the information obtained from the initEWAS function.
ExpoPath	The path to store the user's sample data. Each row represents a sample, and each column represents a variable (exposure variable or covariate). Both .csv and .xlsx file types are supported. The first column must be the sample ID, which must be consistent with the IDs in the methylation data.
MethyPath	The path to store the user's methylation data. Each row represents a CpG site, and each column represents a sample. Both .csv and .xlsx file types are supported. The first column must be the CpG probes. The sample IDs must be consistent with the IDs in the sample data.
ExpoData	The data.frame of the user-supplied sample data that has been loaded into the R environment. If default, the example data inside the package is used. The first column must be the sample name.
MethyData	The data.frame of the user-supplied methylation data that has been loaded into the R environment. If default, an example of methylation data inside the package is loaded. The first column must be the CpG site name.

Value

input, an R6 class object integrating all information.

Examples

```
## Not run:
res <- initEWAS(outpath = "default")
res <- loadEWAS(input = res, ExpoData = "default", MethyData = "default")

## End(Not run)
```

`plotEWAS` *Visualize the results of EWAS analysis*

Description

Visualize EWAS results based on the CMplot package, including Manhattan plots, QQ plots, etc. Please note that this function only supports plotting a single-layer circular Manhattan plot. Additionally, the meaning of each parameter in this function is exactly the same as in CMplot For more detailed information or to create multi-layer circular Manhattan plots, please refer to CMplot (https://cran.r-project.org/web/packages/CMplot/index.html).

Usage

```
plotEWAS(input, p = "PVAL", threshold=NULL, file=c("jpg","pdf","tiff","png"),
col=c("#4197d8","#f8c120","#413496","#495226","#d60b6f","#e66519","#d581b7","#83d3ad",
"#7c162c","#26755d"),LOG10=TRUE,pch=19,type="p",band=1,axis.cex=1,axis.lwd=1.5,
lab.cex=1.5,lab.font=2,plot.type=c("m","c","q","d"),r=0.3,cex=c(0.5,1,1),ylab="",
ylab.pos=3,xticks.pos=1,threshold.col="red", threshold.lwd=1,threshold.lty=2,
amplify=FALSE,signal.cex=1.5,signal.pch=19,signal.col=NULL,signal.line=2, highlight=NULL,
highlight.cex=1,highlight.pch=19,highlight.type="p",highlight.col="red",highlight.text=NULL,
highlight.text.col="black",highlight.text.cex=1,highlight.text.font=3,chr.labels=NULL,
chr.border=FALSE,chr.labels.angle=0,cir.axis=TRUE,cir.axis.col="black",cir.axis.grid=TRUE,
conf.int=TRUE,conf.int.col=NULL, file.name="",dpi=300,height=NULL,width=NULL,main="",
main.cex=1.5,main.font=2,box=FALSE,verbose=FALSE)
```

Arguments

`input`	An R6 class integrated with all the information obtained from the startEWAS function.
`p`	The user needs to specify the name of the p value selected for the result visualization.
`threshold`	The significant threshold.If threshold = 0 or NULL, then the threshold line will not be added.
`file`	The format of the output image file, including "jpg","pdf","tiff", and "png".
`col`	A vector specifies the colors for the chromosomes. If the length of col is shorter than the number of chromosomes, the colors will be applied cyclically.
`LOG10`	logical, whether to change the p-value into log10(p-value) scale.
`pch`	a integer, the shape for the points, is the same with "pch" in plot.

type	a character, could be "p" (point), "l" (cross line), "h" (vertical lines) and so on, is the same with "type" in plot.
band	a number, the size of space between chromosomes, the default is 1.
axis.cex	a number, controls the size of ticks labels of X/Y-axis and the ticks labels of axis for circle plot.
axis.lwd	a number, controls the thickness of X/Y-axis lines and the thickness of axis for circle plot.
lab.cex	a number, controls the size of labels of X/Y-axis and the labels of chromosomes for circle plot.
lab.font	a number, controls the font of labels of all axis.
plot.type	a character or vector, only "d", "c", "m", "q" can be used. if plot.type="d", SNP density will be plotted; if plot.type="c", only circle-Manhattan plot will be plotted; if plot.type="m",only Manhattan plot will be plotted; if plot.type="q",only Q-Q plot will be plotted; if plot.type=c("m","q"), Both Manhattan and Q-Q plots will be plotted.
r	a number, the radius for the circle (the inside radius), the default is 1.
cex	a number or a vector, the size for the points, is the same with "size" in plot, and if it is a vector, the first number controls the size of points in circle plot(the default is 0.5), the second number controls the size of points in Manhattan plot (the default is 1), the third number controls the size of points in Q-Q plot (the default is 1)
ylab	a character, the labels for y axis.
ylab.pos	the distance between ylab and yaxis.
xticks.pos	the distance between labels of x ticks and x axis.
threshold.col	a character or vector, the color for the line of threshold levels, it can also control the color of the diagonal line of QQplot.
threshold.lwd	a number or vector, the width for the line of threshold levels, it can also control the thickness of the diagonal line of QQplot.
threshold.lty	a number or vector, the type for the line of threshold levels, it can also control the type of the diagonal line of QQplot
amplify	logical, CMplot can amplify the significant points, if TRUE, then the points bigger than the minimal significant level will be amplified, the default: amplify=TRUE.
signal.cex	a number, if amplify=TRUE, users can set the size of significant points.
signal.pch	a number, if amplify=TRUE, users can set the shape of significant points.
signal.col	a character, if amplify=TRUE, users can set the colour of significant points, if signal.col=NULL, then the colors of significant points will not be changed.
signal.line	a number, the thickness of the lines of significant CpGs cross the circle.
highlight	a vector, names of CpGs which need to be highlighted.
highlight.cex	a vector, the size of points for CpGs which need to be highlighted.
highlight.pch	a vector, the pch of points for CpGs which need to be highlighted.
highlight.type	a vector, the type of points for CpGs which need to be highlighted.
highlight.col	a vector, the col of points for CpGs which need to be highlighted.
highlight.text	a vector, the text which would be added around the highlighted CpGs.

highlight.text.col
: a vector, the color for added text.

highlight.text.cex
: a value, the size for added text.

highlight.text.font
: text font for the highlighted CpGs

chr.labels
: a vector, the labels for the chromosomes of density plot and Manhattan plot.

chr.border
: a logical, whether to plot the dot line between chromosomes.

chr.labels.angle
: a value, rotate tick labels of x-axis for Manhattan plot (-90 < chr.labels.angle < 90).

cir.axis
: a logical, whether to add the axis of circle Manhattan plot.

cir.axis.col
: a character, the color of the axis for circle.

cir.axis.grid
: logical, whether to add axis grid line in circles.

conf.int
: logical, whether to plot confidence interval on QQ-plot.

conf.int.col
: character or vector, the color of confidence interval of QQplot.

file.name
: a character or vector, the names of output files.

dpi
: a number, the picture resolution for '.jpg', '.npg', and '.tiff' files. The default is 300.

height
: the height of output files.

width
: the width of output files.

main
: character of vector, the title of the plot for manhattan plot and qqplot.

main.cex
: size of title.

main.font
: font of title.

box
: logical, this function draws a box around the current plot.

verbose
: whether to print the log information.

Value

input, An R6 class object integrating all information.

Examples

```
## Not run:
res <- initEWAS(outpath = "default")
res <- loadEWAS(input = res, ExpoData = "default", MethyData = "default")
res <- transEWAS(input = res, Vars = "cov1", TypeTo = "factor")
res <- startEWAS(input = res, chipType = "EPICV2", model = "lm", expo = "var", adjustP = TRUE)
res <- plotEWAS(input = res, p = "PVAL")

## End(Not run)
```

`startEWAS` *Perform EWAS Analysis*

Description

Perform EWAS analysis to obtain the coefficient value, standard deviation and significance p value (or adjust p value) of each site.

Usage

```
startEWAS(input, filename ="default", model = "lm", expo = "default",
cov = NULL,random = NULL, time = NULL, status = NULL, chipType = "EPICV2",
adjustP = TRUE, core = "default")
```

Arguments

`input`	An R6 class integrated with all the information obtained from the loadEWAS or transEWAS function.
`filename`	User-customized .csv file name for storing EWAS results. If "default" is chosen, it will be named as "ewasresult".
`model`	The statistical models used for EWAS analysis include "lm" (general linear regression), "lmer" (linear mixed-effects model), and "cox" (Cox proportional hazards model). The default model is "lm".
`expo`	Name of the exposure variable used in the EWAS analysis.
`cov`	Name(s) of covariate(s) used in the EWAS analysis, with each name separated by a comma. Ensure there are no space. e.g. "cov1,cov2,cov3".
`random`	Random intercept item name, used only when selecting the "lmer" model.
`time`	When the user selects the Cox proportional risk model, the name of the time variable needs to be specified.
`status`	When the user selects the Cox proportional risk model, the name of the status variable needs to be specified.
`adjustP`	Whether to calculate adjusted p-values(FDR and Bonferroni correction). The default is set to TRUE.
`chipType`	The Illumina chip versions for user measurement of methylation data, including "27K", "450K","EPICV1", "EPICV2", and "MSA". The default is "EPICV2".
`core`	The number of cores used during parallel computation. If set to default, it calculates the maximum number of available physical cores minus 1 and treats this as an operational kernel.

Value

input, An R6 class object integrating all information.

Examples

```
## Not run:
res <- initEWAS(outpath = "default")
res <- loadEWAS(input = res, ExpoData = "default", MethyData = "default")
res <- transEWAS(input = res, Vars = "cov1", TypeTo = "factor")
res <- startEWAS(input = res, filename = "default", chipType = "EPICV2", model = "lm",
expo = "var", cov = "cov1,cov2",adjustP = TRUE, core = "default")

## End(Not run)
```

`transEWAS` *Convert variable type of sample data*

Description

Transform the variable types of sample data to the types specified by users.

Usage

```
transEWAS(input, Vars = "default", TypeTo = "factor")
```

Arguments

`input`	An R6 class integrated with all the information obtained from the loadEWAS function.
`Vars`	Variable names that the user wants to convert types for, with each variable name separated by a comma. Ensure there are no spaces. e.g. "var1,var2,var3".
`TypeTo`	The type of variable that the function allows to be converted, including numeric and factor.

Value

input, An R6 class object integrating all information.

Examples

```
## Not run:
res <- initEWAS(outpath = "default")
res <- loadEWAS(input = res, ExpoData = "default", MethyData = "default")
res <- transEWAS(input = res, Vars = "cov1", TypeTo = "factor")

## End(Not run)
```

Index

batchEWAS, 2
bootEWAS, 3

CMplot, *8*
ComBat, *2*

dmrEWAS, 4

enrichEWAS, 5

initEWAS, 6

loadEWAS, 7

plot, *8, 9*
plotEWAS, 8

startEWAS, 11

transEWAS, 12