Package 'metevalue'

March 28, 2023

Type Package

Title E-Value in the Omics Data Association Studies

Version 0.1.14

Date 2023-03-28

Maintainer Yifan Yang <yfyang.86@hotmail.com>

Description In the omics data association studies, it is common to conduct the p-value corrections to control the false significance. Among those p-value correction methods, E-value is recently studied based on V. Vovk and R. Wang (2021) <doi:10.1214/20-AOS2020>. This package provides e-value calculation for several types of omics data association studies. Currently, four data formats are supported: BiSeq, MDRfinder, methylKit and metilene data. The relevant references are listed below: Katja Hebestreit and Hans-Ulrich Klein (2022) <doi:10.18129/B9.bioc.BiSeq>; Altuna Akalin et.al (2012) <doi:10.18129/B9.bioc.methylKit>.

License Apache License (>= 2)

RoxygenNote 7.2.1

Depends sqldf, psych, dplyr, R (>= 3.5.0)

Encoding UTF-8

Suggests rmarkdown, prettydoc, knitr, ggplot2, tidyr, testthat (>= 3.0.0)

VignetteBuilder knitr

LazyData true

Config/testthat/edition 3

NeedsCompilation no

Author Yifan Yang [aut, cre, cph],
Xiaoqing Pan [aut],
Haoyuan Liu [aut]

Repository CRAN

Date/Publication 2023-03-28 02:40:02 UTC

R topics documented:

`demo_biseq_DMR` *DMR BiSeq Demo Dataset*

Description

The BiSeq dataset for demo purpose. The data are dummy data. It includes 9 columns:

The dummy output for BiSeq illustrating purpose. It is dummy.

Details

- seqnames: Chromosome
- start: The positions of the start sites of the corresponding region
- end: The positions of the end sites of the corresponding region
- strand: Strand
- median.p
- median.meth.group1
- median.meth.group2
- median.meth.diff
- seqnames
- start

- end
- width
- strand
- median.p
- median.meth.group1
- median.meth.group2
- median.meth.diff

Notice that there are "NaN" within the feature columns.

Please check the vignette "metevalue" for details.

`demo_biseq_methyrate` *BiSeq Methyrate Demo Dataset*

Description

The methyrate for BiSeq illustrating purpose. It is dummy.

Details

The data includes 12 columns.

- chr: string Chromosome
- pos: int Position
- g1~g2: methylation rate data in groups, repeat 5 times. Notice that there are "NaN" within the feature columns.

Please check the vignette "metevalue" for details.

`demo_DMRfinder_DMRs` *DMRfinder Output Demo Dataset*

Description

The output dummy dataset for DMRfinder illustrating purpose.

Details

The data includes 6 columns.

- chr: string Chromosome
- pos: int Position
- g1~g2: methylation rate data in groups, repeat 2 times. Notice that there are "NaN" within the feature columns.

Please check the vignette "metevalue" for details.

`demo_DMRfinder_rate_combine`

DMRfinder Methyrate Demo Dataset

Description

The methyrate for BiSeq illustrating purpose. It is dummy.

Details

The data includes 6 columns.

- chr: string Chromosome

- pos: int Position

- g1~g2: methylation rate data in groups, repeat 2 times. Notice that there are "NaN" within the feature columns.

Please check the vignette "metevalue" for details.

`demo_methylkit_methyrate`

Methyrate Dataset

Description

The methyrate dataset samples "myCpG" data from the methylKit (a bioconductor package) for illustrating purpose.

Details

The data includes 6 columns.

- chr: string Chromosome

- pos: int Position

- g1~g2: methylation rate data in groups (4 columns)

Please check the vignette "metevalue" for details.

References

Akalin, Altuna, et al. "methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles." Genome biology 13.10 (2012): 1-9. doi: 10.1186/gb20121310r87

`demo_methylkit_met_all` *Methyrate output dataset from methylKit*

Description

The methyrate dataset samples "myCpG" data from the methylKit (a bioconductor package) for illustrating purpose.

Details

The data includes 7 columns:

- chr: Chromosome

- start: The positions of the start sites of the corresponding region

- end: The positions of the end sites of the corresponding region

- strand: Strand

- pvalue: The adjusted p-value based on BH method in MWU-test

- qvalue: cutoff for qvalue of differential methylation statistic

- methyl.diff: The difference between the group means of methylation level

Please check the vignette "metevalue" for details.

References

Akalin, Altuna, et al. "methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles." Genome biology 13.10 (2012): 1-9. doi: 10.1186/gb20121310r87

`demo_metilene_input` *Metilene Methyrate Demo Dataset*

Description

The methyrate for metilene illustrating purpose. It is dummy.

Details

The data includes 18 columns.

- chr: string Chromosome

- pos: int Position

- g1~g2: methylation rate data in groups, repeat 8 times. Notice that there are "NaN" within the feature columns.

Please check the vignette "metevalue" for details.

demo_metilene_out *Metilene Demo Output Dataset*

Description

The output dummy data for "metilene" meythod illustrating purpose.

Details

The data includes 10 columns.

- V1: string Chromosome
- V2: The positions of the start sites of the corresponding region
- V3: The positions of the end sites of the corresponding region
- V4- V10: data value.

Please check the vignette "metevalue" for details.

evalue_buildin_sql *Build-in data process function*

Description

Build-in data process function

Usage

```
evalue_buildin_sql(a, b, method = "metilene")
```

Arguments

a	data frame of the methylation rate
b	data frame of output data corresponding to the "method" option
method	"metilene" or "biseq", "DMRfinder" or "methylKit"

Value

a data frame combines data frame a and b corresponding to the "method" option

Examples

```
data("demo_metilene_out")
data("demo_metilene_input")
result = evalue_buildin_var_fmt_nm(demo_metilene_input,
                                   demo_metilene_out, method="metilene")
result_sql = evalue_buildin_sql(result$a, result$b, method="metilene")
```

evalue_buildin_var_fmt_nm *Build-in check file format function Perform the format check and data clean for the "metilene" or "biseq", "DMRfinder" or "methylKit" method correspondingly.*

Description

Build-in check file format function Perform the format check and data clean for the "metilene" or "biseq", "DMRfinder" or "methylKit" method correspondingly.

Usage

```
evalue_buildin_var_fmt_nm(a, b, method = "metilene")
```

Arguments

a	data frame of the methylation rate
b	data frame of output data corresponding to the "method" option
method	"metilene" or "biseq", "DMRfinder" or "methylKit"

Value

list(a, b) which contains the cleaned data correspondingly

Examples

```
data("demo_metilene_out")
data("demo_metilene_input")
evalue_buildin_var_fmt_nm(demo_metilene_input,
                          demo_metilene_out, method="metilene")
```

metevalue.biseq *Evalue of the BiSeq data format*

Description

Perform the Evaluation for the BiSeq data. Please check vignette "metevalue" for details.

Usage

```
metevalue.biseq(
  methyrate,
  BiSeq.output,
  adjust.methods = "BH",
  sep = "\t",
  bheader = FALSE
)
```

Arguments

methyrate	is the methyrate file. The columns are (in order): - chr: Chromosome - pos: int Position - g1~g2: methylation rate data in groups
BiSeq.output	is the output file of BiSeq. The columns are (in order): - seqnames: Chromosome - start: The positions of the start sites of the corresponding region - end: The positions of the end sites of the corresponding region - width: The number of CpG sites within the corresponding region - strand: Strand - median.p: The median p-value among CpG sites within the corresponding region - median.meth.group1: The median methylation rate in the first group among CpG sites within the corresponding region - median.meth.group2: The median methylation rate in the second group among CpG sites within the corresponding region - median.meth.diff: The median methylation difference between groups among CpG sites within the corresponding region
adjust.methods	is the adjust methods of e-value. It can be 'bonferroni', 'hochberg', 'holm', 'hommel', 'BH', 'BY'
sep	seperator, default is the TAB key.
bheader	a logical value indicating whether the BiSeq.output file contains the names of the variables as its first line. By default, bheader = FALSE.

Value

a dataframe, the columns are (in order):

- chr: Chromosome

- start: The positions of the start sites of the corresponding region

- end: The positions of the end sites of the corresponding region

- q-value: The adjusted p-value based on BH method in MWU-test

- methyl.diff: The difference between the group means of methylation level

- CpGs: The number of CpG sites within the corresponding region

- p : p-value based on MWU-test

- p2: p-value based on 2D KS-test

- m1: The absolute mean methylation level for the corresponding segment of group 1

- m2: The absolute mean methylation level for the corresponding segment of group 2

- e_value: The e-value of the corresponding region

Examples

```
data("demo_biseq_methyrate")
data("demo_biseq_DMR")
example_tempfiles = tempfile(c("demo_biseq_methyrate", "demo_biseq_DMR"))
tempdir()
#### write to temp file ####
write.table(demo_biseq_methyrate, file=example_tempfiles[1],row.names=FALSE,
            col.names=TRUE, quote=FALSE, sep='\t')
write.table (demo_biseq_DMR, file=example_tempfiles[2],
             sep ="\t", row.names =FALSE, col.names =TRUE, quote =FALSE)
#### compute e-value and its adjustment ####
result = metevalue.biseq(example_tempfiles[1],
                         example_tempfiles[2], bheader = TRUE)
```

`metevalue.biseq.chk` *Check the BiSeq data format*

Description

Check the BiSeq data format

Usage

```
metevalue.biseq.chk(
  input_filename_a,
  input_filename_b,
  sep = "\t",
  bheader = FALSE
)
```

Arguments

`input_filename_a`

metilene input file path. This file is a sep (e.g. TAB) separated file with two key columns and several value columns: For exampe:

chrom pos g1 g1 g1 g1 g1 g1 g1 g1 g2 g2 g2 g2 g2 g2 g2 g2

chrom and pos are keys; g1 g1 g2 g2 are samples in two experiment groups.

input_filename_b
: metilene input file path. This file should stored as a sep(e.g. TAB) separated file with two key columns and several value columns: The columns are (in order):
- chr: Chromosome
- start: The position of the start site of the corresponding region
- end: The position of the end site of the corresponding region
- range: The range of the corresponding region
- strand: Strand
- median.p: The median of p-values in the corresponding region
- median.meth.group1 : The median of methylation level for the corresponding segment of group 1
- median.meth.group2 : The median of methylation level for the corresponding segment of group 2
- median.meth.diff: The median of the difference between the methylation level

sep
: separator, default is the TAB key.

bheader
: a logical value indicating whether the input_filename_b file contains the names of the variables as its first line. By default, bheader = FALSE.

Value

list(file_a, file_b, file_a_b) returns a list with three pr-handled data.frames corresponding to the input_filename_a, input_filename_b file and a A JOIN B file.

Examples

```
data("demo_biseq_DMR")
data("demo_biseq_methyrate")
```

---

metevalue.DMRfinder *Evalue of the DMRfinder data format*

Description

Perform the Evaluation for the DMRfinder data.

Usage

```
metevalue.DMRfinder(
  methyrate,
  DMRfinder.output,
  adjust.methods = "BH",
  sep = "\t",
  bheader = FALSE
)
```

Arguments

`methyrate`	is the methyrate file. - chr: Chromosome - pos: int Position - g1~g2: methylation rate data in groups
`DMRfinder.output`	is the output file of DMRfinder. - chr: Chromosome - start: The positions of the start sites of the corresponding region - end: The positions of the end sites of the corresponding region - CpG: The number of CpG sites within the corresponding region - Control.mu: The average methylation rate in control group - Expt1.mu: The average methylation rate in experiment group - Control.Expt1.diff: The methylation difference between control and experiment groups - Control.Expt1.pval: P-value based on Wald-test.
`adjust.methods`	is the adjust methods of e-value. It can be 'bonferroni', 'hochberg', 'holm', 'hommel', 'BH', 'BY'
`sep`	seperator, default is the TAB key.
`bheader`	a logical value indicating whether the DMRfinder.output file contains the names of the variables as its first line. By default, bheader = FALSE.

Value

a dataframe, the columns are (in order):

- chr: Chromosome
- start: The positions of the start sites of the corresponding region
- end: The positions of the end sites of the corresponding region
- q-value: The adjusted p-value based on BH method in MWU-test
- methyl.diff: The difference between the group means of methylation level
- CpGs: The number of CpG sites within the corresponding region
- p : p-value based on MWU-test
- p2: p-value based on 2D KS-test
- m1: The absolute mean methylation level for the corresponding segment of group 1
- m2: The absolute mean methylation level for the corresponding segment of group 2
- e_value: The e-value of the corresponding region

Examples

```
#### DMRfinder example ####'
data(demo_DMRfinder_rate_combine)
data(demo_DMRfinder_DMRs)
```

`metevalue.DMRfinder.chk` *Check the DMRfinder data format*

Description

Check the DMRfinder data format

Usage

```
metevalue.DMRfinder.chk(
  input_filename_a,
  input_filename_b,
  sep = "\t",
  bheader = FALSE
)
```

Arguments

`input_filename_a`	the combined data of methylation rate file. This file is a sep (e.g. TAB) separated file with two key columns and several value columns: For exampe: chrom pos g1 g1 g1 g1 g1 g1 g1 g1 g2 g2 g2 g2 g2 g2 g2 g2 chrom and pos are keys; g1 g1 g2 g2 are samples in two experiment groups.
`input_filename_b`	the output file of DMRfinder. The columns are (in order): - chr: Chromosome - start: The position of the start sites of the corresponding region - end: The position of the end sites of the corresponding region - CpG: The number of CpG sites within the corresponding region - 'Control:mu': The absolute mean methylation level for the corresponding segment of the control group - 'Exptl:mu': The absolute mean methylation level for the corresponding segment of the experimental group - 'Control->Exptl:diff': The difference between the group means of methylation level - p: p-value
`sep`	separator, default is the TAB key.
`bheader`	a logical value indicating whether the input_filename_b file contains the names of the variables as its first line. By default, bheader = FALSE.

Value

list(file_a, file_b, file_a_b) returns a list with three pr-handled data.frames corresponding to the input_filename_a, input_filename_b file and a A JOIN B file.

Examples

```
data("demo_DMRfinder_rate_combine")
data("demo_DMRfinder_DMRs")
```

`metevalue.methylKit` *Evalue of the methylKit data format*

Description

Perform the Evaluation for the BiSeq data.

Usage

```
metevalue.methylKit(
  methyrate,
  methylKit.output,
  adjust.methods = "BH",
  sep = "\t",
  bheader = FALSE
)
```

Arguments

`methyrate`	is the output file of methylKit, the columns are (in order): - chr: Chromosome - pos: int Position - g1~g2: methylation rate data in groups
`methylKit.output`	is the output data with e-value of each region - chr: Chromosome - start: The positions of the start sites of the corresponding region - end: The positions of the end sites of the corresponding region - strand: Strand - pvalue: The adjusted p-value based on BH method in MWU-test - qvalue: cutoff for qvalue of differential methylation statistic - methyl.diff: The difference between the group means of methylation level
`adjust.methods`	is the adjust methods of e-value. It can be 'bonferroni', 'hochberg', 'holm', 'hommel', 'BH', 'BY'
`sep`	seperator, default is the TAB key.
`bheader`	a logical value indicating whether the input_filename_b file contains the names of the variables as its first line. By default, bheader = FALSE.

Value

a dataframe, the columns are (in order):

- chr: Chromosome
- start: The positions of the start sites of the corresponding region
- end: The positions of the end sites of the corresponding region
- q-value: The adjusted p-value based on BH method in MWU-test
- methyl.diff: The difference between the group means of methylation level
- CpGs: The number of CpG sites within the corresponding region
- p : p-value based on MWU-test
- p2: p-value based on 2D KS-test
- m1: The absolute mean methylation level for the corresponding segment of group 1
- m2: The absolute mean methylation level for the corresponding segment of group 2
- e_value: The e-value of the corresponding region

Examples

```
#### methylKit example ####
data(demo_methylkit_methyrate)
data(demo_methylkit_met_all)
example_tempfiles = tempfile(c("rate_combine", "methylKit_DMR_raw"))
tempdir()
write.table(demo_methylkit_methyrate, file=example_tempfiles[1],
      row.names=FALSE, col.names=TRUE, quote=FALSE, sep='\t')
write.table (demo_methylkit_met_all, file=example_tempfiles[2],
      sep ="\t", row.names =FALSE, col.names =TRUE, quote =FALSE)
result = metevalue.methylKit(example_tempfiles[1], example_tempfiles[2],
      bheader = TRUE)
str(result)
```

---

metevalue.methylKit.chk

Check the methylKit data format

Description

Check the methylKit data format

Usage

```
metevalue.methylKit.chk(
  input_filename_a,
  input_filename_b,
  sep = "\t",
  bheader = FALSE
)
```

Arguments

input_filename_a	the combined data of methylation rate file. This file is a sep (e.g. TAB) separated file with two key columns and several value columns: For exampe: chrom pos g1 g1 g1 g1 g1 g1 g1 g1 g2 g2 g2 g2 g2 g2 g2 g2 chrom and pos are keys; g1 g1 g2 g2 are samples in two experiment groups.
input_filename_b	the output file of methylKit. a methylDiff or methylDiffDB object containing the differential methylated locations satisfying the criteria. The columns are (in order): - chr: Chromosome - start: The position of the start sites of the corresponding region - end: The position of the end sites of the corresponding region - strand: Strand - p: p-value - qvalue: The adjusted p-value based on BH method - meth.diff : The difference between the group means of methylation level
sep	separator, default is the TAB key.
bheader	a logical value indicating whether the input_filename_b file contains the names of the variables as its first line. By default, bheader = FALSE.

Value

list(file_a, file_b, file_a_b) returns a list with three pr-handled data.frames corresponding to the input_filename_a, input_filename_b file and a A JOIN B file.

Examples

```
#### methylKit example ####
data(demo_methylkit_methyrate)
data(demo_methylkit_met_all)
```

---

metevalue.metilene *Evalue of the Metilene data format*

Description

Evalue of the Metilene data format

Usage

```
metevalue.metilene(
  methyrate,
  metilene.output,
  adjust.methods = "BH",
  sep = "\t",
  bheader = FALSE
)
```

Arguments

methyrate	metilene input file path. This file is a sep (e.g. TAB) separated file with two key columns and several value columns in pairs: For exampe: chrom pos g1 g1 g1 g1 g1 g1 g1 g1 g2 g2 g2 g2 g2 g2 g2 g2 chrom and pos are keys; g1 g1 g2 g2 must be stored in groups.
metilene.output	metilene input file path. This file should stored as a sep(e.g. TAB) separated file with two key columns and several value columns: The columns are (in order): - chr: Chromosome - start: The positions of the start sites of the corresponding region - end: The positions of the end sites of the corresponding region - q-value: The adjusted p-value based on BH method in MWU-test - methyl.diff: The difference between the group means of methylation level - CpGs: The number of CpG sites within the corresponding region - p : p-value based on MWU-test - p2: p-value based on 2D KS-test - m1: The absolute mean methylation level for the corresponding segment of group 1 - m2: The absolute mean methylation level for the corresponding segment of group 2
adjust.methods	is the adjust methods of e-value. It can be 'bonferroni', 'hochberg', 'holm', 'hommel', 'BH', 'BY'
sep	seperator, default is the TAB key.
bheader	a logical value indicating whether the metilene.output file contains the names of the variables as its first line. By default, bheader = FALSE.

Value

a dataframe, the columns are (in order):

- chr: Chromosome

- start: The positions of the start sites of the corresponding region

- end: The positions of the end sites of the corresponding region

- q-value: The adjusted p-value based on BH method in MWU-test

- methyl.diff: The difference between the group means of methylation level

- CpGs: The number of CpG sites within the corresponding region
- p : p-value based on MWU-test
- p2: p-value based on 2D KS-test
- m1: The absolute mean methylation level for the corresponding segment of group 1
- m2: The absolute mean methylation level for the corresponding segment of group 2
- e_value: The e-value of the corresponding region

Examples

```
#### metilene example ####'
data(demo_metilene_input)
data(demo_metilene_out)
```

`metevalue.metilene.chk` *Check the Metilene data format*

Description

Check the Metilene data format

Usage

```
metevalue.metilene.chk(
  input_filename_a,
  input_filename_b,
  sep = "\t",
  bheader = FALSE
)
```

Arguments

`input_filename_a`
: metilene input file path. This file is a sep (e.g. TAB) separated file with two key columns and several value columns: For exampe:
chrom pos g1 g1 g1 g1 g1 g1 g1 g1 g2 g2 g2 g2 g2 g2 g2 g2
chrom and pos are keys; g1 g1 g2 g2 are samples in two experiment groups.

`input_filename_b`
: metilene input file path. This file should stored as a sep(e.g. TAB) separated file with two key columns and several value columns: The columns are (in order):
- chr: Chromosome
- start: The position of the start sites of the corresponding region
- end: The position of the end sites of the corresponding region
- q-value: The adjusted p-value based on BH method in MWU-test
- methyl.diff: The difference between the group means of methylation level

	- CpGs: The number of CpG sites within the corresponding region - p : p-value based on MWU-test - p2: p-value based on 2D KS-test - m1: The absolute mean methylation level for the corresponding segment of group 1 - m2: The absolute mean methylation level for the corresponding segment of group 2
sep	separator, default is the TAB key.
bheader	a logical value indicating whether the input_filename_b file contains the names of the variables as its first line. By default, bheader = FALSE.

Value

list(file_a, file_b, file_a_b) returns a list with three pr-handled data.frames corresponding to the input_filename_a, input_filename_b file and a A JOIN B file.

Examples

```
data("demo_metilene_out")
data("demo_metilene_input")
```

`varevalue.metilene` *Evalue of the Metilene data*

Description

Perform the Evaluation for the Metilene data. The data file could be pre-handled by the evalue.metilene.chk function.

Usage

```
varevalue.metilene(a, b, a_b, adjust.methods = "BH")
```

Arguments

a	A data.frame object, the columns should be (in order): chrom pos g1 g1 g1 g1 g1 g1 g1 g1 g2 g2 g2 g2 g2 g2 g2 g2 i.e two key columns (chrom, pos) with several value columns in groups.
b	A data.frame object stores the data, the columns are (in order): - chr: Chromosome - start: The positions of the start sites of the corresponding region - end: The positions of the end sites of the corresponding region - q-value: The adjusted p-value based on BH method in MWU-test - methyl.diff: The difference between the group means of methylation level - CpGs: The number of CpG sites within the corresponding region

- p : p-value based on MWU-test
- p2: p-value based on 2D KS-test
- m1: The absolute mean methylation level for the corresponding segment of group 1
- m2: The absolute mean methylation level for the corresponding segment of group 2

`a_b` A data.frame object of a join b with particular data clean processes. Check the function [evalue.methylKit.chk()] for more details.

`adjust.methods` is the adjust methods of e-value. It can be 'bonferroni', 'hochberg', 'holm', 'hommel', 'BH', 'BY'. The default value is 'BH'.

Value

a dataframe, the columns are (in order):

- chr: Chromosome
- start: The positions of the start sites of the corresponding region
- end: The positions of the end sites of the corresponding region
- q-value: The adjusted p-value based on BH method in MWU-test
- methyl.diff: The difference between the group means of methylation level
- CpGs: The number of CpG sites within the corresponding region
- p : p-value based on MWU-test
- p2: p-value based on 2D KS-test
- m1: The absolute mean methylation level for the corresponding segment of group 1
- m2: The absolute mean methylation level for the corresponding segment of group 2
- e_value: The e-value of the corresponding region

Examples

```
#### methylKit example ####
data(demo_methylkit_methyrate)
data(demo_methylkit_met_all)
example_tempfiles = tempfile(c("rate_combine", "methylKit_DMR_raw"))
tempdir()
write.table(demo_methylkit_methyrate, file=example_tempfiles[1],
      row.names=FALSE, col.names=TRUE, quote=FALSE, sep='\t')
write.table (demo_methylkit_met_all, file=example_tempfiles[2],
      sep ="\t", row.names =FALSE, col.names =TRUE, quote =FALSE)
result = metevalue.methylKit(example_tempfiles[1], example_tempfiles[2],
      bheader = TRUE)
str(result)
```

Index