Package 'MLML2R'

October 17, 2019

Type Package

Title Maximum Likelihood Estimation of DNA Methylation and Hydroxymethylation Proportions

Version 0.3.3

Author Samara Kiihl [aut, cre],
Maria Jose Martinez-Garrido [aut],
Arce Domingo-Relloso [aut],
Jose Bermudez [aut],
Maria Tellez-Plaza [aut]

Maintainer Samara Kiihl <samara@ime.unicamp.br>

Description Maximum likelihood estimates (MLE) of the proportions of 5-mC and 5-hmC in the DNA using information from BS-conversion, TAB-conversion, and oxBS-conversion methods. One can use information from all three methods or any combination of two of them. Estimates are based on Binomial model by Qu et al. (2013) <doi:10.1093/bioinformatics/btt459> and Kiihl et al. (2019) <doi:10.1515/sagmb-2018-0031>.

License MIT + file LICENSE

Encoding UTF-8

VignetteBuilder knitr

LazyData true

biocViews Software, MethylationArray, Epigenetics, DNAMethylation, Microarray, TwoChannel, OneChannel

RoxygenNote 6.1.1

URL https://github.com/samarafk/MLML2R

Suggests minfi, microbenchmark, GEOquery, knitr, rmarkdown, IlluminaHumanMethylation450kmanifest

BugReports https://github.com/samarafk/MLML2R/issues

NeedsCompilation no

Repository CRAN

Date/Publication 2019-10-17 19:30:02 UTC

R topics documented:

`C_BS_sim` *C (unconverted) counts from BS*

Description

A matrix of simulated counts corresponding to 100 CpGs and 2 samples. True proportions of methylation, hydroxymethylation and unmethylated used in the simulation are .3, .2, and .5, respectively.

Usage

```
C_BS_sim
```

Format

100 x 2 matrix (CpGs in the rows and Samples in the columns).

`C_OxBS_sim` *C (unconverted) counts from oxBS*

Description

A matrix of simulated counts corresponding to 100 CpGs and 2 samples. True proportions of methylation, hydroxymethylation and unmethylated used in the simulation are .3, .2, and .5, respectively.

Usage

```
C_OxBS_sim
```

Format

100 x 2 matrix (CpGs in the rows and Samples in the columns).

C_TAB_sim — *C (unconverted) counts from TAB*

Description

A matrix of simulated counts corresponding to 100 CpGs and 2 samples. True proportions of methylation, hydroxymethylation and unmethylated used in the simulation are .3, .2, and .5, respectively.

Usage

```
C_TAB_sim
```

Format

100 x 2 matrix (CpGs in the rows and Samples in the columns).

MLML — *MLE (maximum likelihood estimates) of 5-mC and 5-hmC levels.*

Description

MLE (maximum likelihood estimates) of 5-mC and 5-hmC levels.

Usage

```
MLML(U.matrix = NULL, T.matrix = NULL, G.matrix = NULL,
  H.matrix = NULL, L.matrix = NULL, M.matrix = NULL,
  iterative = FALSE, tol = 1e-05)
```

Arguments

U.matrix	Converted cytosines (T counts or U channel) from standard BS-conversion (True 5-C).
T.matrix	Unconverted cytosines (C counts or M channel) from standard BS-conversion (reflecting 5-mC+5-hmC).
G.matrix	Converted cytosines (T counts or U channel) from TAB-conversion (reflecting 5-C + 5-mC).
H.matrix	Unconverted cytosines (C counts or M channel) from TAB-conversion(reflecting True 5-hmC).
L.matrix	Converted cytosines (T counts or U channel) from oxBS-conversion (reflecting 5-C + 5-hmC).
M.matrix	Unconverted cytosines (C counts or M channel) from oxBS-conversion (reflecting True 5-mC).

iterative	logical. If iterative=TRUE EM-algorithm is used. For the combination of two methods, iterative=FALSE returns the exact constrained MLE using the pool-adjacent-violators algorithm (PAVA). When all three methods are combined, iterative=FALSE returns the constrained MLE using Lagrange multiplier.
tol	convergence tolerance; considered only if iterative=TRUE

Details

The function returns MLE estimates (binomial model assumed). The function assumes that the order of the rows and columns in the input matrices are consistent. In addition, all the input matrices must have the same dimension. Usually, rows represent CpG loci and columns are the samples.

Value

The returned value is a list with the following components.

mC	maximum likelihood estimate for the proportion of methylation.
hmC	maximum likelihood estimate for the proportion of hydroxymethylation.
C	maximum likelihood estimate for the proportion of unmethylation.
methods	the conversion methods used to produce the MLE

Author(s)

Samara Kiihl samara@ime.unicamp.br; Maria Jose Garrido; Arce Domingo-Relloso; Jose Bermudez; Maria Tellez-Plaza.

References

Kiihl SF, Martinez-Garrido MJ, Domingo-Relloso A, Bermudez J, Tellez-Plaza M. MLML2R: an R package for maximum likelihood estimation of DNA methylation and hydroxymethylation proportions. Statistical Applications in Genetics and Molecular Biology. 2019;18(1). doi:10.1515/sagmb-2018-0031.

Qu J, Zhou M, Song Q, Hong EE, Smith AD. MLML: consistent simultaneous estimates of DNA methylation and hydroxymethylation. Bioinformatics. 2013;29(20):2645-2646. doi:10.1093/bioinformatics/btt459.

Ayer M, Brunk HD, Ewing GM, Reid WT, Silverman E. An Empirical Distribution Function for Sampling with Incomplete Information. Ann. Math. Statist. 1955, 26(4), 641-647. doi:10.1214/aoms/1177728423.

Zongli Xu, Jack A. Taylor, Yuet-Kin Leung, Shuk-Mei Ho, Liang Niu; oxBS-MLE: an efficient method to estimate 5-methylcytosine and 5-hydroxymethylcytosine in paired bisulfite and oxidative bisulfite treated DNA, Bioinformatics, 2016;32(23):3667-3669.

Examples

```
# load the example datasets from BS, oxBS and TAB methods
data(C_BS_sim)
data(C_OxBS_sim)
data(T_BS_sim)
data(T_OxBS_sim)
data(C_TAB_sim)
```

```
data(T_TAB_sim)

# obtain MLE via EM-algorithm for BS+oxBS:
results_em <- MLML(T.matrix = C_BS_sim , U.matrix = T_BS_sim,
L.matrix = T_OxBS_sim, M.matrix = C_OxBS_sim,iterative=TRUE)

# obtain constrained exact MLE for BS+oxBS:
results_exact <- MLML(T.matrix = C_BS_sim , U.matrix = T_BS_sim,
L.matrix = T_OxBS_sim, M.matrix = C_OxBS_sim)

# obtain MLE via EM-algorithm for BS+TAB:
results_em <- MLML(T.matrix = C_BS_sim , U.matrix = T_BS_sim,
G.matrix = T_TAB_sim, H.matrix = C_TAB_sim,iterative=TRUE)

# obtain constrained exact MLE for BS+TAB:
results_exact <- MLML(T.matrix = C_BS_sim , U.matrix = T_BS_sim,
G.matrix = T_TAB_sim, H.matrix = C_TAB_sim)

# obtain MLE via EM-algorithm for oxBS+TAB:
results_em <- MLML(L.matrix = T_OxBS_sim, M.matrix = C_OxBS_sim,
G.matrix = T_TAB_sim, H.matrix = C_TAB_sim,iterative=TRUE)

# obtain constrained exact MLE for oxBS+TAB:
results_exact <- MLML(L.matrix = T_OxBS_sim, M.matrix = C_OxBS_sim,
G.matrix = T_TAB_sim, H.matrix = C_TAB_sim)

# obtain MLE via EM-algorithm for BS+oxBS+TAB:
results_em <- MLML(T.matrix = C_BS_sim , U.matrix = T_BS_sim,
L.matrix = T_OxBS_sim, M.matrix = C_OxBS_sim,
G.matrix = T_TAB_sim, H.matrix = C_TAB_sim,iterative=TRUE)

#' # obtain MLE via Lagrange multiplier for BS+oxBS+TAB:
results_exact <- MLML(T.matrix = C_BS_sim , U.matrix = T_BS_sim,
L.matrix = T_OxBS_sim, M.matrix = C_OxBS_sim,
G.matrix = T_TAB_sim, H.matrix = C_TAB_sim)


# Example of datasets with zero counts and missing values:

C_BS_sim[1,1] <- 0
C_OxBS_sim[1,1] <- 0
C_TAB_sim[1,1] <- 0
T_BS_sim[1,1] <- 0
T_OxBS_sim[1,1] <- 0
T_TAB_sim[1,1] <- 0

C_BS_sim[2,2] <- NA
C_OxBS_sim[2,2] <- NA
C_TAB_sim[2,2] <- NA
T_BS_sim[2,2] <- NA
T_OxBS_sim[2,2] <- NA
T_TAB_sim[2,2] <- NA
```

`T_BS_sim` *T (converted) counts from BS*

Description

A matrix of simulated counts corresponding to 100 CpGs and 2 samples. True proportions of methylation, hydroxymethylation and unmethylated used in the simulation are .3, .2, and .5, respectively.

Usage

```
T_BS_sim
```

Format

100 x 2 matrix (CpGs in the rows and Samples in the columns).

`T_OxBS_sim` *T (converted) counts from oxBS*

Description

A matrix of simulated counts corresponding to 100 CpGs and 2 samples. True proportions of methylation, hydroxymethylation and unmethylated used in the simulation are .3, .2, and .5, respectively.

Usage

```
T_OxBS_sim
```

Format

100 x 2 matrix (CpGs in the rows and Samples in the columns).

T_TAB_sim *T (converted) counts from TAB*

Description

A matrix of simulated counts corresponding to 100 CpGs and 2 samples. True proportions of methylation, hydroxymethylation and unmethylated used in the simulation are .3, .2, and .5, respectively.

Usage

```
T_TAB_sim
```

Format

100 x 2 matrix (CpGs in the rows and Samples in the columns).

Index