# Package 'MethTargetedNGS'

June 26, 2022

**Type** Package

**Title** Perform Methylation Analysis on Next Generation Sequencing Data

**Version** 1.29.0

**Date** 2015-02-06

**Depends** R (>= 3.1.2), stringr, seqinr, gplots, Biostrings

**SystemRequirements** HMMER3

**Author** Muhammad Ahmer Jamil with Contribution of Prof. Holger Frohlich and Priv.-Doz. Dr. Osman El-Maarri

**Maintainer** Muhammad Ahmer Jamil <engr.ahmerjamil@gmail.com>

**Description** Perform step by step methylation analysis of Next Generation Sequencing data.

**License** Artistic-2.0

**biocViews** ResearchField, Genetics, Sequencing, Alignment, SequenceMatching, DataImport

**git_url** https://git.bioconductor.org/packages/MethTargetedNGS

**git_branch** master

**git_last_commit** 77c2ac3

**git_last_commit_date** 2022-04-26

**Date/Publication** 2022-06-26

## R topics documented:

---

`MethTargetedNGS-package` *Methylation Analysis of Next Generation Sequencing data.*

---

## Description

This package helps in visualizing methylation in CpG sites in NGS data for given datasets (normal/tumor) and to identify differentially methylated CpG sites in normal/tumor. This package to help in perform profile hidden markov modelling of given sequences.

NOTE: For profile hidden markov model HMMER software is required

## Details

| | |
|---|---|
| Package: | MethTargetedNGS |
| Type: | Package |
| Version: | 1.0 |
| Date: | 2015-01-20 |
| License: | Artistic-2.0 |

Compare methylation status/pattern between samples.

`*compare_samples(healthy,tumor)`

Sequence alignment and create methylation pattern

`*methAlign(sequence_fasta, ref_seq)`

## Author(s)

Muhammad Ahmer Jamil, Prof. Holger Frohlich, Priv.-Doz. Dr. Osman El-Maarri

Maintainer: Muhammad Ahmer Jamil `<engr.ahmerjamil@gmail.com>`

---

`bconv` *Convert non-bisulfite sequences to bisulfite sequences.*

---

## Description

Bisulfite sequences are the bisulfite treated DNA sequences where all cytosines except cytosine from CpG sites are converted to thymie. This technique is used to determine pattern of methylation. This function convert all cytosine except cytosines from CpG sites to thymine.

### Usage

```
bconv(fasta_file, out_file = "output.fasta")
```

### Arguments

| | |
|---|---|
| fasta_file | : Input fasta file for conversion |
| out_file | : String value naming an output file. Default is output.fasta |

### Value

Fasta File

### Author(s)

Muhammad Ahmer Jamil, Prof. Holger Frohlich, Priv.-Doz. Dr. Osman El-Maarri

Maintainer: Muhammad Ahmer Jamil <engr.ahmerjamil@gmail.com>

### Examples

```
input =  system.file("extdata", "bisulfite.fasta", package = "MethTargetedNGS")
bconv(fasta_file = input, out_file = "output.fasta")
```

---

## compare_samples *Complete Methylation Analysis of Next Generation Sequencing Data*

---

### Description

This function perform complete methylation analysis of the data.

1. Visualize methylation pattern
2. Calculate methylation average
3. Calculate methylation entropy
4. Perform fisher exact test on the samples to identify significant CpG sites.

### Usage

```
compare_samples(healthy, tumor)
```

### Arguments

| | |
|---|---|
| healthy | : Output Matrix from methAlign |
| tumor | : Output Matrix from methAlign |

### Value

Generate a plot of Methylation Average, Methylation Entropy, Fisher Exact Test and Log Odd Ratio

## Note

This function needs time to process depending on the number of sequences in fasta file

## Author(s)

Muhammad Ahmer Jamil, Prof. Holger Frohlich, Priv.-Doz. Dr. Osman El-Maarri

Maintainer: Muhammad Ahmer Jamil <engr.ahmerjamil@gmail.com>

## See Also

methAlign, methAvg, methEntropy, odd_ratio, fishertest_cpg,

## Examples

```
healthy = system.file("extdata", "Healthy.fasta", package = "MethTargetedNGS")
tumor = system.file("extdata", "Tumor.fasta", package = "MethTargetedNGS")
reference =  system.file("extdata", "Reference.fasta", package = "MethTargetedNGS")

healthy = methAlign(healthy,reference)
tumor = methAlign(tumor,reference)
compare_samples(healthy,tumor)
```

---

# fishertest_cpg *Perform Fisher Exact Test on Methylation Data.*

---

## Description

Fisher exact test is a test to calculate the statistical significance using contingency table. It was used to find the statistically significant differences in the methylation status of one particular CpG site between healthy and tumor sample. Contingency matrix was created for each CpG site. P-value was corrected for multiple testing using Benjamini-Hochberg method to calculate False Discovery Rate (FDR)

## Usage

```
fishertest_cpg(healthy, tumor, plot = TRUE, main = "Fisher Exact Test")
```

## Arguments

| | |
|---|---|
| healthy | Matrix from methAlign. Also matrix where columns represents Cytosine of CpG sites and rows represents sequences. |
| tumor | Matrix from methAlign. Also matrix where columns represents Cytosine of CpG sites and rows represents sequences. |
| plot | Boolean. TRUE if need a plot after calculation. Default TRUE |
| main | Title of the plot. Default "Fisher Exact Test" |

## Value

Vector containing p-values.

## Author(s)

Muhammad Ahmer Jamil, Prof. Holger Frohlich, Priv.-Doz. Dr. Osman El-Maarri

Maintainer: Muhammad Ahmer Jamil <engr.ahmerjamil@gmail.com>

## See Also

methAlign, compare_samples

## Examples

```
healthy = system.file("extdata", "Healthy.fasta", package = "MethTargetedNGS")
tumor = system.file("extdata", "Tumor.fasta", package = "MethTargetedNGS")
reference =  system.file("extdata", "Reference.fasta", package = "MethTargetedNGS")

healthy = methAlign(healthy,reference)
tumor = methAlign(tumor,reference)
fisherexacttest <- fishertest_cpg(healthy,tumor)
```

---

# hmmbuild *Create Profile Hidden Markov Model of given aligned sequences*

---

## Description

This function creates profile hidden markov model of the given aligned sequences using HMMER algorithm.[1]

## Usage

```
hmmbuild(file_seq, file_out,pathHMMER="")
```

## Arguments

| | |
|---|---|
| file_seq | Multiple sequence aligned fasta file |
| file_out | Output hidden markov model file |
| pathHMMER | Path where HMMER software is installed. Note: Windows user must setup cygwin to use this feature and set path to HMMER binaries ( ~hmmer/binaries/) |

## Value

Create Profile Hidden Markov Model in local directory

## Note

Require HMMER software

Windows User: Please download HMMER from http://hmmer.janelia.org/

Setup cygwin from http://www.cygwin.com

Linux/Mac User: Download binaries or compile HMMER from http://hmmer.janelia.org/

## Author(s)

Muhammad Ahmer Jamil, Prof. Holger Frohlich, Priv.-Doz. Dr. Osman El-Maarri

Maintainer: Muhammad Ahmer Jamil <engr.ahmerjamil@gmail.com>

## References

[1]Finn, Robert D., Jody Clements, and Sean R. Eddy. "HMMER web server: interactive sequence similarity searching." Nucleic acids research (2011): gkr367.

## See Also

nhmmer

## Examples

```
msa = system.file("extdata", "msa.fasta", package = "MethTargetedNGS")
if (file.exists("/usr/bin/hmmbuild"))
  hmmbuild(file_seq=msa,file_out="hmm",pathHMMER = "/usr/bin")
```

---

`methAlign` *Align sequences with the reference sequence using pairwiseAlignment function from BioStrings*

---

## Description

This function allow users to align pool of sequences to the reference sequence.

## Usage

```
methAlign(sequence_fasta, ref_seq, sub_mat = FALSE, align_type = "local")
```

## Arguments

| | |
|---|---|
| `sequence_fasta` | String value naming an input fasta file. Single sequence or Multiple sequences in a single fasta file |
| `ref_seq` | String value naming an input fasta file. Single reference sequence is requried. If multiple sequences were passed only first sequence will be considered as reference. |
| `sub_mat` | Substitution matrix for the alignment. |

| | |
|---|---|
| align_type | type of alignment. One of "global", "local", "overlap", "global-local", and "local-global" where<br>"global" = align whole strings with end gap penalties,<br>"local" = align string fragments,<br>"overlap" = align whole strings without end gap penalties,<br>"global-local" = align whole strings in pattern with consecutive subsequence of subject,<br>"local-global" = align consecutive subsequence of pattern with whole strings in subject.<br>Default is "local" |

## Value

Methylation Matrix. Number of rows represents number of reads in sequence fasta file and number of columns represents number of CpG sites in reference fasta sequence. Only Cytosine of CpG site was observed in the table whether it is methylated or unmethylated.

## Note

This function need some time to process depending on the number of sequences in fasta file

## Author(s)

Muhammad Ahmer Jamil, Prof. Holger Frohlich, Priv.-Doz. Dr. Osman El-Maarri

Maintainer: Muhammad Ahmer Jamil <engr.ahmerjamil@gmail.com>

## See Also

compare_samples

## Examples

```
healthy = system.file("extdata", "Healthy.fasta", package = "MethTargetedNGS")
reference = system.file("extdata", "Reference.fasta", package = "MethTargetedNGS")
methAlign(healthy,reference)
```

---

# methAvg *Calculate Methylation Average of given methylation matrix*

---

## Description

Methylation average of a CpG site is the percentage of unmethylated cytosine or methylated cytosine in a particular CpG site. The methylation average of a particular CpG site was calculated by number of cytosine divided by sum of total number of methylated and unmethylated cytosine at particular CpG site in a group of reads.

average = NC/(NC + NT)

## Usage

```
methAvg(Sample, plot = FALSE)
```

## Arguments

| | |
|---|---|
| Sample | Matrix from methAlign. Also matrix where columns represents Cytosine of CpG sites and rows represents sequences. |
| plot | Boolean. TRUE if need a plot after calculation. Default FALSE |

## Value

Vector containing average methylation of given methylation matrix. Length of the vector represents the number of CpG sites in methylation matrix.

## Author(s)

Muhammad Ahmer Jamil, Prof. Holger Frohlich, Priv.-Doz. Dr. Osman El-Maarri

Maintainer: Muhammad Ahmer Jamil <engr.ahmerjamil@gmail.com>

## See Also

methAlign, compare_samples

## Examples

```
healthy = system.file("extdata", "Healthy.fasta", package = "MethTargetedNGS")
reference = system.file("extdata", "Reference.fasta", package = "MethTargetedNGS")
methP <- methAlign(healthy,reference)
avgMeth <- methAvg(methP,plot=TRUE)
```

---

# methEntropy *Calculate Methylation Entropy*

---

## Description

Entropy comparison between healthy and tumor samples can identify significant CpG sites which are contributing most in the tumor development either by hypomethylation or hypermethylation. Also such way can help in understanding the randomness in methylation status. Sliding window of 4 was used to calculate the entropy in the sample, which can analyze 16 different pattern for entropy calculation.

## Usage

```
methEntropy(x)
```

### Arguments

| | |
|---|---|
| x | Matrix from methAlign. Also matrix where columns represents Cytosine of CpG sites and rows represents sequences |

### Value

Matrix containing entropy for every sequence and group of 4 cpg sites.

### Note

This function needs time to process depending on the number of rows in matrix

### Author(s)

Muhammad Ahmer Jamil, Prof. Holger Frohlich, Priv.-Doz. Dr. Osman El-Maarri

Maintainer: Muhammad Ahmer Jamil <engr.ahmerjamil@gmail.com>

### References

Xie, H., Wang, M., de Andrade, A., Bonaldo, M.d.F., Galat, V., Arndt, K., Rajaram, V., Goldman, S., Tomita, T. and Soares, M.B. (2011) Genome-wide quantitative assessment of variation in DNA methylation patterns. Nucleic Acids Research, 39, 4099-4108.

### See Also

methAlign

### Examples

```
healthy = system.file("extdata", "Healthy.fasta", package = "MethTargetedNGS")
reference = system.file("extdata", "Reference.fasta", package = "MethTargetedNGS")
methP <- methAlign(healthy,reference)
entMeth <- methEntropy(methP)
plot(entMeth,type="l")
```

---

## methHeatmap *Generate Heatmap of the given methylation data.*

---

### Description

Heatmaps are the way of visualizing methylation statuses of a sample. This function allows user to visualize methylation statuses at each CpG site for every sequence available in pool.

### Usage

```
methHeatmap(Sample, yl = "", plot = TRUE, title = "")
```

### Arguments

| | |
|---|---|
| Sample | Matrix from methAlign. Also matrix where columns represents Cytosine of CpG sites and rows represents sequences. |
| yl | Ylabel for heatmap |
| plot | Boolean. If plot == FALSE, function will return a matrix of 1s and 0s. If plot == TRUE, function will create a heatmap as well as return a matrix of 1s and 0s |
| title | Title of the heatmap |

### Value

Heatmap

### Author(s)

Ahmer Jamil <engr.ahmerjamil@gmail.com>

### See Also

methAlign

### Examples

```
healthy = system.file("extdata", "Healthy.fasta", package = "MethTargetedNGS")
reference = system.file("extdata", "Reference.fasta", package = "MethTargetedNGS")
healthy = methAlign(healthy,reference)
hHeatmap = methHeatmap(healthy,plot=TRUE)
```

---

## nhmmer — *Calculate likelihood of the given profile hidden markov model against group of sequences*

---

### Description

This function calculates likelihood score of given pool of sequences against given profile hidden markov model using HMMER algorithm.[1]

### Usage

```
nhmmer(file_hmm, file_seq, pathHMMER="")
```

### Arguments

| | |
|---|---|
| file_hmm | HMM file from hmmbuild function |
| file_seq | Sequence fasta file for calculating likelihood |
| pathHMMER | Path where HMMER software is installed. Note: Windows user must setup cygwin to use this feature and set path to HMMER binaries ( ~hmmer/binaries/) |

## Value

Matrix containing likelihood scores

## Note

Require HMMER software

Windows User: Please download HMMER from http://hmmer.janelia.org/

Setup cygwin from http://www.cygwin.com

Linux/Mac User: Download binaries or compile HMMER from http://hmmer.janelia.org/

## Author(s)

Muhammad Ahmer Jamil, Prof. Holger Frohlich, Priv.-Doz. Dr. Osman El-Maarri

Maintainer: Muhammad Ahmer Jamil <engr.ahmerjamil@gmail.com>

## References

[1]Finn, Robert D., Jody Clements, and Sean R. Eddy. "HMMER web server: interactive sequence similarity searching." Nucleic acids research (2011): gkr367.

## See Also

hmmbuild

## Examples

```
msa = system.file("extdata", "msa.fasta", package = "MethTargetedNGS")
tumor = system.file("extdata", "Tumor.fasta", package = "MethTargetedNGS")
if (file.exists("/usr/bin/hmmbuild"))
{hmmbuild(file_seq=msa,file_out="hmm",pathHMMER = "/usr/bin")
res <- nhmmer("hmm",tumor,pathHMMER = "/usr/bin")
res}
```

---

# odd_ratio *Calculate log odd ratio of the given samples (healthy/tumor)*

---

## Description

Log Odd ratio defines the hypomethylation and hypermethylation of a sample in comparison to the other sample.

## Usage

```
odd_ratio(SampA, SampB, plot = TRUE, main = "Log Odd Ratio")
```

## Arguments

| | |
|---|---|
| SampA | Matrix from methAlign. Also matrix where columns represents Cytosine of CpG sites and rows represents sequences. |
| SampB | Matrix from methAlign. Also matrix where columns represents Cytosine of CpG sites and rows represents sequences. |
| plot | Boolean. TRUE if need a plot after calculation. Default TRUE |
| main | Title of the plot |

## Value

Vector containing log odd ratios.

## Author(s)

Muhammad Ahmer Jamil, Prof. Holger Frohlich, Priv.-Doz. Dr. Osman El-Maarri

Maintainer: Muhammad Ahmer Jamil <engr.ahmerjamil@gmail.com>

## See Also

methAlign

## Examples

```
healthy = system.file("extdata", "Healthy.fasta", package = "MethTargetedNGS")
tumor = system.file("extdata", "Tumor.fasta", package = "MethTargetedNGS")
reference =  system.file("extdata", "Reference.fasta", package = "MethTargetedNGS")

healthy = methAlign(healthy,reference)
tumor = methAlign(tumor,reference)
odd_ratio(healthy,tumor)
```

# Index