Package: scmeth (via r-universe)

June 20, 2024

Type Package

Title Functions to conduct quality control analysis in methylation data

Version 1.25.0

Author Divy Kangeyan <divyswar01@g.harvard.edu>

Maintainer Divy Kangeyan <divyswar01@g.harvard.edu>

Depends R (>= 3.5.0)

Imports knitr, rmarkdown, bsseq, AnnotationHub, GenomicRanges, reshape2, stats, utils, BSgenome, DelayedArray (>= 0.5.15), annotatr, SummarizedExperiment (>= 1.5.6), GenomeInfoDb, Biostrings, DT, HDF5Array (>= 1.7.5)

Suggests BSgenome.Mmusculus.UCSC.mm10, BSgenome.Hsapiens.NCBI.GRCh38, TxDb.Hsapiens.UCSC.hg38.knownGene, org.Hs.eg.db, Biobase, BiocGenerics, ggplot2, ggthemes

Description Functions to analyze methylation data can be found here. Some functions are relevant for single cell methylation data but most other functions can be used for any methylation data. Highlight of this workflow is the comprehensive quality control report.

biocViews DNAMethylation, QualityControl, Preprocessing, SingleCell, ImmunoOncology

BugReports https://github.com/aryeelab/scmeth/issues

License GPL-2

LazyData TRUE

RoxygenNote 6.1.1

VignetteBuilder knitr

Repository https://bioc.r-universe.dev

RemoteUrl https://github.com/bioc/scmeth

RemoteRef HEAD

RemoteSha 363cbc55bbfefb46fd825b92e74cc9a3ed0e01eb

Contents

`bsConversionPlot` *Bisulfite conversion rate visualization*

Description

Plot the bisulfite conversion rate for each sample based on the pheno data in the bs object

Usage

```
bsConversionPlot(bs)
```

Arguments

`bs`	bsseq object

Value

Plot showing bisulfite conversion rate for each sample

Examples

```
directory <- system.file("extdata/bismark_data", package='scmeth')
bs <- HDF5Array::loadHDF5SummarizedExperiment(directory)
bsConversionPlot(bs)
```

chromosomeCoverage *CpG covergae in each chromosome*

Description

Provides Coverage metrics for each sample by the chromosome

Usage

```
chromosomeCoverage(bs)
```

Arguments

bs	bsseq object

Value

matrix of chromsome covergae with column and rows indicating the samples and the chromosome respectively

Examples

```
directory <- system.file("extdata/bismark_data",package='scmeth')
bs <- HDF5Array::loadHDF5SummarizedExperiment(directory)
chromosomeCoverage(bs)
```

coverage *Coverage for single cells*

Description

Provides Coverage for each cell in a library pool

Usage

```
coverage(bs, subSample = 1e+06, offset = 50000)
```

Arguments

bs	bsseq object
subSample	number of CpGs to subsample. Default value is 1000000.
offset	how many CpGs to offset when subsampling Default value is set to be 50000, i.e. first 50000 CpGs will be ignored in subsampling.

Value

vector of coverage for the cells in bs object

Examples

```
directory <- system.file("extdata/bismark_data", package='scmeth')
bs <- HDF5Array::loadHDF5SummarizedExperiment(directory)
coverage(bs)
```

`cpgDensity` *Provides Coverage by the CpG density. CpG Density is defined as the number of CpGs observed in certain base pair long region.*

Description

Provides Coverage by the CpG density. CpG Density is defined as the number of CpGs observed in certain base pair long region.

Usage

```
cpgDensity(bs, organism, windowLength = 1000, small = FALSE)
```

Arguments

`bs`	bsseq object
`organism`	scientific name of the organism of interest, e.g. Mmusculus or Hsapiens
`windowLength`	Length of the window to calculate the density
`small`	Indicator for a small dataset, cpg density is calculated more memory efficiently for large dataset but for small dataset a different quicker method is used Default value for window length is 1000 basepairs.

Value

Data frame with sample name and coverage in repeat masker regions

Examples

```
library(BSgenome.Hsapiens.NCBI.GRCh38)
directory <- system.file("extdata/bismark_data", package='scmeth')
bs <- HDF5Array::loadHDF5SummarizedExperiment(directory)
cpgDensity(bs, Hsapiens, 1000, small=TRUE)
```

`cpgDiscretization` *Discretize the CpG methylation values to align with single cell analysis*

Description

In single cell analysis overwhelmingly large number of CpGs have binary methylation Due to errors in sequencing and amplification many CpGs tend to have non-binary methylation. Hence this function catergorizes the non-binary CpGs as methylated if the methyation is above 0.8 and unmethylated if the methylation is below 0.2

Usage

```
cpgDiscretization(bs, subSample = 1e+06, offset = 50000,
  coverageVec = NULL)
```

Arguments

`bs`	bsseq object
`subSample`	number of CpGs to subsample. Default value is 1000000.
`offset`	how many CpGs to offset when subsampling Default value is set to be 50000, i.e. first 50000 CpGs will be ignored in subsampling.
`coverageVec`	If coverage vector is already calculated provide it to speed up the process

Value

meth discretized methylation matrix

discard total number of removed CpGs from each sample

Percentage of CpGs discarded compared to the total number of CpGs

Examples

```
directory <- system.file("extdata/bismark_data", package='scmeth')
bs <- HDF5Array::loadHDF5SummarizedExperiment(directory)
cpgDiscretization(bs)
```

downsample *Downsample analysis*

Description

Downsample the CpG coverage matrix for saturation analysis

Usage

```
downsample(bs, dsRates = c(0.01, 0.02, 0.05, seq(0.1, 0.9, 0.1)),
  subSample = 1e+06, offset = 50000)
```

Arguments

bs	bsseq object
dsRates	downsampling rate. i.e. the probabaility of sampling a single CpG default is list of probabilities ranging from 0.01 to 1 0.01 0.02 0.05 0.10 0.20 0.30 0.40 0.50 0.60 0.70 0.80 0.90 For more continuous saturation curve dsRates can be changed to add more sampling rates
subSample	number of CpGs to subsample Default value is 1000000.
offset	how many CpGs to offset when subsampling Default value is set to be 50000, i.e. first 50000 CpGs will be ignored in subsampling.

Value

Data frame with the CpG coverage for each sample at each sampling rate

Examples

```
directory <- system.file("extdata/bismark_data", package='scmeth')
bs <- HDF5Array::loadHDF5SummarizedExperiment(directory)
scmeth::downsample(bs)
```

featureCoverage *Coverage based on the genomic feature*

Description

Provides Coverage metrics for the sample by each genomic features provided by the user

Usage

```
featureCoverage(bs, features, genomebuild)
```

Arguments

bs	bsseq object
features	list of genomic features, e.g. genes_exons, genes_introns, cpg_islands, cpg_shelves Names are based on the annotatr packages, so all the features provided by the annotatr package will be supported in this function
genomebuild	reference alignment, i.e. mm10 or hg38

Value

a data frame with genomic feature names and the number of CpG covered in each feature

Examples

```
directory <- system.file("extdata/bismark_data", package='scmeth')
bs <- HDF5Array::loadHDF5SummarizedExperiment(directory)
featureCoverage(bs, c('cpg_islands', 'cpg_shores'), 'hg38')
```

---

mbiasplot *Methylation bias plot*

Description

Plot the methylation at each position of the read to observe any biases in the methylation based on the read position

Usage

```
mbiasplot(dir = NULL, mbiasFiles = NULL)
```

Arguments

dir	directory name with mbias files
mbiasFiles	list of mbias files

Value

Returns a list containing the methylation across the read position in original top and original bottom strand both in forward and reverse reads for multiple samples

Examples

```
mbiasFile <- '2017-04-21_HG23KBCXY_2_AGGCAGAA_TATCTC_pe.M-bias.txt'
mbiasplot(mbiasFiles=system.file("extdata", mbiasFile, package='scmeth'))
```

methylationDist *Provide graphics for methylation distribution*

Description

Plot the methylation distribution for the cells in bsseq object

Usage

```
methylationDist(bs)
```

Arguments

bs	bsseq object

Value

mean methylation for each sample

Examples

```
directory <- system.file("extdata/bismark_data", package='scmeth')
bs <- HDF5Array::loadHDF5SummarizedExperiment(directory)
methylationDist(bs)
```

readmetrics *Provide graphics for read information*

Description

Plot the mapped and unmapped reads

Usage

```
readmetrics(bs)
```

Arguments

bs	bsseq object

Value

Plot showing the mapped and unmapped read information for each cell

Examples

```
directory <- system.file("extdata/bismark_data", package='scmeth')
bs <- HDF5Array::loadHDF5SummarizedExperiment(directory)
readmetrics(bs)
```

repMask *Provides Coverage metrics in the repeat masker region*

Description

Provides Coverage metrics in the repeat masker region

Usage

```
repMask(bs, organism, genome)
```

Arguments

bs	bsseq object
organism	scientific name of the organism of interest, e.g. Mmusculus or Hsapiens
genome	reference alignment, i.e. mm10 or hg38

Value

Data frame with sample name and coverage in repeat masker regions

Examples

```
library(BSgenome.Mmusculus.UCSC.mm10)
library(AnnotationHub)
load(system.file("extdata", 'bsObject.rda', package='scmeth'))
repMask(bs, Mmusculus, 'mm10')
```

report *Generates an inclusive report on methylation analysis*

Description

This function uses most of the functions in this package to generate a report for the user

Usage

```
report(bsObj, outdirectory, organism, genome, mbiasDir = NULL,
  subSample = 1e+06, offset = 50000, small = FALSE)
```

Arguments

bsObj	bsseq object
outdirectory	name of the output directory where the report will be saved
organism	scientific name of the organism of interest, e.g. Mmusculus or Hsapiens
genome	reference alignment, e.g. mm10 or hg38 the report will have graphics on read information
mbiasDir	Optional argument to provide directory name that has the mbias files or the list of mbias files
subSample	number of CpGs to subsample Default value is 1000000.
offset	how many CpGs to offset when subsampling Default value is set to be 50000, i.e. first 50000 CpGs will be ignored in subsampling.
small	Indicator for a small dataset, cpg density is calculated more

Value

Report will be an html file

Examples

```
library(BSgenome.Hsapiens.NCBI.GRCh38)
directory <- system.file("extdata/bismark_data", package='scmeth')
bs <- HDF5Array::loadHDF5SummarizedExperiment(directory)
mbiasDirectory=system.file("extdata", package='scmeth')
outDir <- system.file(package='scmeth')
report(bs, outDir, Hsapiens, 'hg38', mbiasDir=mbiasDirectory, small=TRUE)
```

---

scmeth *scmeth: a package to conduct quality control analysis for methylation data. Most functions can be applied to both bulk and single-cell methylation while other functions are specific to single-cell methylation data. scmeth is especially customized to use the output from the FireCloud implementation of methylation pipeline to produce comprehensive quality control report*

Description

scmeth: a package to conduct quality control analysis for methylation data. Most functions can be applied to both bulk and single-cell methylation while other functions are specific to single-cell methylation data. scmeth is especially customized to use the output from the FireCloud implementation of methylation pipeline to produce comprehensive quality control report

Index