

# Package ‘mCSEA’

May 6, 2024

**Type** Package

**Title** Methylated CpGs Set Enrichment Analysis

**Version** 1.24.0

**Author** Jordi Martorell-Marugán and Pedro Carmona-Sáez

**Maintainer** Jordi Martorell-Marugán <jmartorellm@gmail.com>

**Description** Identification of diferentially methylated regions (DMRs) in predefined regions (promoters, CpG islands...) from the human genome using Illumina's 450K or EPIC microarray data.  
Provides methods to rank CpG probes based on linear models and includes plotting functions.

**Depends** R (>= 3.5), mCSEAdata, Homo.sapiens

**Suggests** Biobase, BiocGenerics, BiocStyle, FlowSorted.Blood.450k, knitr, leukemiasEset, minfi, minfiData, rmarkdown, RUnit

**Imports** biomaRt, fgsea, GenomicFeatures, GenomicRanges, ggplot2, graphics, grDevices, Gviz, IRanges, limma, methods, parallel, S4Vectors, stats, SummarizedExperiment, utils

**VignetteBuilder** knitr

**biocViews** ImmunoOncology, DifferentialMethylation, DNAMethylation, Epigenetics, Genetics, GenomeAnnotation, MethylationArray, Microarray, MultipleComparison, TwoChannel

**License** GPL-2

**Encoding** UTF-8

**LazyData** true

**git\_url** <https://git.bioconductor.org/packages/mCSEA>

**git\_branch** RELEASE\_3\_19

**git\_last\_commit** 419e949

**git\_last\_commit\_date** 2024-04-30

**Repository** Bioconductor 3.19

**Date/Publication** 2024-05-05

## Contents

|                            |           |
|----------------------------|-----------|
| mCSEA-package . . . . .    | 2         |
| exprTest . . . . .         | 3         |
| mCSEAIIntegrate . . . . .  | 3         |
| mCSEAPlot . . . . .        | 4         |
| mCSEAPlotGSEA . . . . .    | 6         |
| mCSEATest . . . . .        | 7         |
| precomputedmCSEA . . . . . | 8         |
| rankProbes . . . . .       | 9         |
| <b>Index</b>               | <b>11</b> |

---

|               |  |
|---------------|--|
| mCSEA-package | <i>Methylated CpGs Set Enrichment Analysis</i> |
|---------------|--|

---

## Description

Identification of diferentially methylated regions (DMRs) in predefined regions (promoters, CpG islands...) from the human genome using Illumina's 450K or EPIC microarray data. Provides methods to rank CpG probes based on linear models and includes plotting functions.

## Author(s)

Jordi Martorell Marugán

Maintainer: Jordi Martorell Marugán<jordi.martorell@genyo.es>

## Examples

```
## Not run:
library(mCSEA)
data(mcseadata)
myRank <- rankProbes(betaTest, phenoTest, refGroup = "Control")
myResults <- mCSEATest(myRank, regionsTypes = "promoters", platform = "EPIC")

## End(Not run)
data(precomputedmCSEA)
head(myResults$promoters)
```

---

|          |                                |
|----------|--------------------------------|
| exprTest | <i>Expression data example</i> |
|----------|--------------------------------|

---

**Description**

exprTest is a subset of 100 genes' microarray expression data for 20 bone marrow samples: 10 from Acute Lymphoblastic Leukemia patients and 10 from healthy patients. It is useful to test mCSEAIIntegrate function.

**Usage**

```
data(exprTest)
```

**Format**

matrix

**Source**

Obtained from the leukemiasEset data package

---

|                 |   |
|-----------------|---|
| mCSEAIIntegrate | <i>Integrate methylation and expression</i> |
|-----------------|---|

---

**Description**

Uses mCSEA methylation analysis results and expression values to search for significant correlations between DMRs methylation and close genes expression.

**Usage**

```
mCSEAIIntegrate(mCSEAResults, exprData, regionType = c("promoters", "genes",
  "CGI", "custom"), geneIDs = "SYMBOL", dmrName = NULL, pcutoff = 0.05,
  minCor = 0.5, minP = 0.05, makePlot = TRUE, folder = ".", nproc = 1)
```

**Arguments**

|              |   |
|--------------|---|
| mCSEAResults | The object generated by mCSEATest function  |
| exprData     | A matrix or data frame with genes in rows and samples in columns. A SummarizedExperiment object can be used too |
| regionType   | The region types to be represented. Must be one or more of "promoters", "genes", "CGI" and "custom"             |
| geneIDs      | Gene identifiers used in exprData. One of "SYMBOL", "ENSEMBL", "ENTREZID", "GENEID", "REFSEQ" or "UNIGENE"      |

|          |   |
|----------|---|
| dmrName  | The DMR of interest to correlate with expression (e.g. gene name, CGI name...). If NULL (default), all DMRs with P-Value < pcutoff are selected |
| pcutoff  | P-Value threshold to select DMRs if dmrName = NULL  |
| minCor   | Correlation threshold to output the results   |
| minP     | Correlation P-Value threshold to output the results   |
| makePlot | If TRUE, generate correlation and save them in the folder specified by folder parameter   |
| folder   | Directory to save the correlation plots if makePlot = TRUE  |
| nproc    | Number of processors to be used   |

**Value**

A data.frame with the integration results.

**Author(s)**

Jordi Martorell Marugán, <jordi.martorell@genyo.es>

**See Also**

[rankProbes](#), [mCSEATest](#)

**Examples**

```
data(precomputedmCSEA)
data(exprTest)

resultsInt <- mCSEAIIntegrate(myResults, exprTest, "promoters", "ENSEMBL",
                             "GATA2", makePlot = FALSE)

resultsInt
```

---

mCSEAPlot

*Plot mCSEA results*


---

**Description**

Generate a graphic with the genomic context of the selected DMR, showing methylation status at each CpG site of different samples groups

**Usage**

```
mCSEAPlot(mCSEAResults, regionType, dmrName, extend = 1000,
          chromosome = TRUE, leadingEdge = TRUE, CGI = FALSE, genes = TRUE,
          transcriptAnnotation = "transcript", makePDF = TRUE,
          col = c("blue", "magenta", "green", "red", "black"))
```

**Arguments**

|                      |   |
|----------------------|---|
| mCSEAResults         | The object generated by mCSEATest function  |
| regionType           | The region type to be represented. Must be one of "promoters", "genes", "CGI" or "custom"                             |
| dmrName              | The DMR of interest to be represented (e.g. gene name, CGI name...)   |
| extend               | The number of base pairs to extend the plot around the DMR of interest (default = 1000 bp)                            |
| chromosome           | If TRUE, represent the chromosome and genome axis   |
| leadingEdge          | If TRUE, represent the bars indicating if each CpG belongs to the mCSEA leading edge (green) or not (red)             |
| CGI                  | If TRUE, represent the annotated CpG islands  |
| genes                | If TRUE, represent the annotated genes  |
| transcriptAnnotation | Labels showed at the genes track. Must be one of "transcript" (default), "symbol", "gene", "exon" or "feature"        |
| makePDF              | If TRUE, save the plot in pdf format in the working directory. Otherwise, draw the plot in the active graphics window |
| col                  | Vector with colors to plot methylation in different groups  |

**Value**

'NULL'

**Author(s)**

Jordi Martorell Marugán, <jordi.martorell@genyo.es>

**See Also**

[rankProbes](#), [mCSEATest](#), [mCSEAPlotGSEA](#)

**Examples**

```
library(mCSEAdata)
data(mcseadata)
## Not run:
myRank <- rankProbes(betaTest, phenoTest, refGroup = "Control")
set.seed(123)
myResults <- mCSEATest(myRank, betaTest, phenoTest,
  regionsTypes = "promoters", platform = "EPIC")

## End(Not run)
data(precomputedmCSEA)
mCSEAPlot(myResults, "promoters", "CLIC6",
  transcriptAnnotation = "symbol", makePDF = FALSE)
```

---

`mCSEAPlotGSEA`*Plot mCSEA results*

---

**Description**

Generate an enrichment plot

**Usage**

```
mCSEAPlotGSEA(rank, mCSEAResults, regionType, dmrName)
```

**Arguments**

|                           |   |
|---------------------------|---|
| <code>rank</code>         | A named numeric vector with the ranking statistic of each CpG site                        |
| <code>mCSEAResults</code> | The object generated by <code>mCSEATest</code> function                                   |
| <code>regionType</code>   | The region type to be represented. Must be one of "promoters", "genes", "CGI" or "custom" |
| <code>dmrName</code>      | The DMR of interest to be represented (e.g. gene name, CGI name...)                       |

**Value**

'NULL'

**Author(s)**

Jordi Martorell Marugán, <jordi.martorell@genyo.es>

**References**

Subramanian, A. et al (2005). *Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles*. PNAS 102, 15545-15550.

**See Also**

[rankProbes](#), [mCSEATest](#), [mCSEAPlot](#)

**Examples**

```
## Not run:
library(mCSEAdata)
data(mcseadata)
myRank <- rankProbes(betaTest, phenoTest, refGroup = "Control")
set.seed(123)
myResults <- mCSEATest(myRank, regionsTypes = "promoters",
platform = "EPIC")

## End(Not run)
data(precomputedmCSEA)
mCSEAPlotGSEA(myRank, myResults, "promoters", "CLIC6")
```

---

|           |                            |
|-----------|----------------------------|
| mCSEATest | <i>mCSEA core analysis</i> |
|-----------|----------------------------|

---

### Description

Perform a methylated CpG sites enrichment analysis in predefined genomic regions

### Usage

```
mCSEATest(rank, methData, pheno = NULL, column = 1,
  regionsTypes = c("promoters", "genes", "CGI"), customAnnotation = NULL,
  minCpGs = 5, nproc = 1, nperm = NULL, platform = "450k")
```

### Arguments

|                  |   |
|------------------|---|
| rank             | A named numeric vector with the ranking statistic of each CpG site  |
| methData         | A data frame or a matrix containing Illumina's CpG probes in rows and samples in columns. A SummarizedExperiment object can be used too                   |
| pheno            | A data frame or a matrix containing samples in rows and covariates in columns. If NULL (default), pheno is extracted from the SummarizedExperiment object |
| column           | The column name or number from pheno used to split the samples into groups (first column is used by default)  |
| regionsTypes     | A character or character vector indicating the predefined regions to be analyzed. NULL to skip this step and use customAnnotation.                        |
| customAnnotation | An optional list with the CpGs associated to each feature (default = NULL)  |
| minCpGs          | Minimum number of CpGs associated to a region. Regions below this threshold are not tested  |
| nproc            | Number of processors to use in GSEA step (default = 1)  |
| nperm            | (deprecated) Number of permutations to do in GSEA step in the previous implementation. Now, this parameter is ignored                                     |
| platform         | Platform used to get the methylation data (450k or EPIC)  |

### Value

A list with the results of each of the analyzed regions. For each region type, a data frame with the results and a list with the probes associated to each region are generated. In addition, this list also contains the input methData, pheno and platform objects

### Author(s)

Jordi Martorell Marugán, <jordi.martorell@genyo.es>

## References

Subramanian, A. et al (2005). *Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles* . PNAS 102, 15545-15550.

## See Also

[rankProbes](#), [mCSEAPlot](#), [mCSEAPlotGSEA](#)

## Examples

```
## Not run:
library(mCSEAdata)
data(mcseadata)
myRank <- rankProbes(betaTest, phenoTest, refGroup = "Control")
set.seed(123)
myResults <- mCSEATest(myRank, betaTest, phenoTest,
  regionsTypes = "promoters", platform = "EPIC")

## End(Not run)
data(precomputedmCSEA)
head(myResults[["promoters"]])
head(myResults[["promoters_association"]])
```

---

precomputedmCSEA

*Precomputed mCSEA results*

---

## Description

myRank is a result of rankProbes function. myResults is a result of mCSEATest function.

## Usage

```
data(precomputedmCSEA)
```

## Format

vector (myRank) and list data.frame (myResults)

## Source

Both objects were obtained with the example commands in the mCSEA vignette.



---

|            |                        |
|------------|------------------------|
| rankProbes | <i>Rank CpG probes</i> |
|------------|------------------------|

---

**Description**

Apply a linear model to Illumina's 450k or EPIC methylation data to get the t-value of each CpG probe

**Usage**

```
rankProbes(methData, pheno = NULL, paired = FALSE, explanatory = 1,  
           covariates = c(), pairColumn = c(), caseGroup = 1, refGroup = 2,  
           continuous = NULL, typeInput = "beta", typeAnalysis = "M")
```

**Arguments**

|              |   |
|--------------|---|
| methData     | A data frame or a matrix containing Illumina's CpG probes in rows and samples in columns. A SummarizedExperiment object can be used too                   |
| pheno        | A data frame or a matrix containing samples in rows and covariates in columns. If NULL (default), pheno is extracted from the SummarizedExperiment object |
| paired       | Perform a paired t-test (default = FALSE)   |
| explanatory  | The column name or position from pheno used to perform the comparison between groups (default = first column)   |
| covariates   | A list or character vector with column names from pheno used as data covariates in the linear model   |
| pairColumn   | Only for paired analysis. The column name or position from pheno used to connect the paired samples (default = NULL)                                      |
| caseGroup    | The group name or position from explanatory variable used as cases to perform the comparison (default = first group)                                      |
| refGroup     | The group name or position from explanatory variable used as reference to perform the comparison (default = second group)                                 |
| continuous   | A list or character vector with columns names from pheno which should be treated as continuous variables (default = none)                                 |
| typeInput    | Type of input methylation data. "beta" for Beta-values and "M" for M-values   |
| typeAnalysis | "M" to use M-values to rank the CpG probes (default). "beta" to use Beta-values instead   |

**Value**

A named vector containing the t-values from the linear model for each CpG probe

**Author(s)**

Jordi Martorell Marugán, <jordi.martorell@genyo.es>

**References**

Smyth, G. K. (2005). *Limma: linear models for microarray data*. Bioinformatics and Computational Biology Solutions using R and Bioconductor, 397-420.

**See Also**

[mCSEATest](#)

**Examples**

```
data(mcseadata)
myRank <- rankProbes(betaTest, phenoTest, refGroup = "Control")
head(myRank)
```

# Index

## \* datasets

exprTest, [3](#)

precomputedmCSEA, [8](#)

exprTest, [3](#)

mCSEA (mCSEA-package), [2](#)

mCSEA-package, [2](#)

mCSEAIIntegrate, [3](#)

mCSEAPlot, [4](#), [6](#), [8](#)

mCSEAPlotGSEA, [5](#), [6](#), [8](#)

mCSEATest, [4-6](#), [7](#), [10](#)

myRank (precomputedmCSEA), [8](#)

myResults (precomputedmCSEA), [8](#)

precomputedmCSEA, [8](#)

rankProbes, [4-6](#), [8](#), [9](#)