

# Package ‘methylCC’

October 16, 2025

**Title** Estimate the cell composition of whole blood in DNA methylation samples

**Version** 1.22.0

**Imports** Biobase, GenomicRanges, IRanges, S4Vectors, dplyr, magrittr, minfi, bsseq, quadprog, plyranges, stats, utils, bumphunter, genefilter, methods, IlluminaHumanMethylation450kmanifest, IlluminaHumanMethylation450kanno.ilmn12.hg19

**Depends** R (>= 3.6), FlowSorted.Blood.450k

**Suggests** rmarkdown, knitr, testthat (>= 2.1.0), BiocGenerics, BiocStyle, tidyr, ggplot2

**Description** A tool to estimate the cell composition of DNA methylation whole blood sample measured on any platform technology (microarray and sequencing).

**biocViews** Microarray, Sequencing, DNAMethylation, MethylationArray, MethylSeq, WholeGenome

**VignetteBuilder** knitr

**RoxygenNote** 6.1.1

**Encoding** UTF-8

**License** GPL-3

**BugReports** <https://github.com/stephaniehicks/methylCC/>

**URL** <https://github.com/stephaniehicks/methylCC/>

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

**git\_branch** RELEASE\_3\_21

**git\_last\_commit** 9d50b97

**git\_last\_commit\_date** 2025-04-15

**Repository** Bioconductor 3.21

**Date/Publication** 2025-10-15

**Author** Stephanie C. Hicks [aut, cre] (ORCID:

<<https://orcid.org/0000-0002-7858-0231>>),

Rafael Irizarry [aut] (ORCID: <<https://orcid.org/0000-0002-3944-4309>>)

**Maintainer** Stephanie C. Hicks <[shicks19@jhu.edu](mailto:shicks19@jhu.edu)>

Contents

|                                     |           |
|-------------------------------------|-----------|
| .extract_raw_data . . . . .         | 2         |
| .find_dmrs . . . . .                | 3         |
| .initializeMLEs . . . . .           | 4         |
| .initialize_theta . . . . .         | 5         |
| .methylcc_engine . . . . .          | 5         |
| .methylcc_estep . . . . .           | 6         |
| .methylcc_mstep . . . . .           | 7         |
| .pick_target_positions . . . . .    | 7         |
| .preprocess_estimatecc . . . . .    | 8         |
| .splitit . . . . .                  | 8         |
| .WFun . . . . .                     | 9         |
| cell_counts . . . . .               | 9         |
| estimatecc . . . . .                | 10        |
| estimatecc-class . . . . .          | 12        |
| FlowSorted.Blood.450k.sub . . . . . | 12        |
| offMethRegions . . . . .            | 13        |
| onMethRegions . . . . .             | 13        |
| <b>Index</b>                        | <b>14</b> |

---

|                   |                         |
|-------------------|-------------------------|
| .extract_raw_data | <i>Extract raw data</i> |
|-------------------|-------------------------|

---

Description

Extract the methylation values and GRanges objects

Usage

.extract\_raw\_data(object)

Arguments

object                    an object can be a RGChannelSet, GenomicMethylSet or BSseq object

Value

A list preprocessed objects from the RGChannelSet, GenomicMethylSet or BSseq objects to be used in .preprocess\_estimatecc().

---

.find\_dmrs*Finding differentially methylated regions*

---

**Description**

This function uses the FlowSorted.Blood.450k whole blood reference methylomes with six cell types to identify differentially methylated regions.

**Usage**

```
.find_dmrs(verbose = TRUE, gr_target = NULL, include_cpgs = FALSE,
  include_dmrs = TRUE, num_cpgs = 50, num_regions = 50,
  bumhunter_beta_cutoff = 0.2, dmr_up_cutoff = 0.5,
  dmr_down_cutoff = 0.4, dmr_pval_cutoff = 1e-11,
  cpg_pval_cutoff = 1e-08, cpg_up_dm_cutoff = 0,
  cpg_down_dm_cutoff = 0, pairwise_comparison = FALSE,
  mset_train_flow_sort = NULL)
```

**Arguments**

|                       |                                                                                                                                                                                                                                                                                     |
|-----------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| verbose               | TRUE/FALSE argument specifying if verbose messages should be returned or not. Default is TRUE.                                                                                                                                                                                      |
| gr_target             | Default is NULL. However, the user can provide a GRanges object from the object in estimatecc. Before starting the procedure to find differentially methylated regions, the intersection of the gr_target and GRanges object from the reference methylomes (FlowSorted.Blood.450k). |
| include_cpgs          | TRUE/FALSE. Should individual CpGs be returned. Default is FALSE.                                                                                                                                                                                                                   |
| include_dmrs          | TRUE/FALSE. Should differentially methylated regions be returned. Default is TRUE. User can turn this to FALSE and search for only CpGs.                                                                                                                                            |
| num_cpgs              | The max number of CpGs to return for each cell type. Default is 50.                                                                                                                                                                                                                 |
| num_regions           | The max number of DMRs to return for each cell type. Default is 50.                                                                                                                                                                                                                 |
| bumhunter_beta_cutoff | The cutoff threshold in bumhunter() in the bumhunter package.                                                                                                                                                                                                                       |
| dmr_up_cutoff         | A cutoff threshold for identifying DMRs that are methylated in one cell type, but not in the other cell types.                                                                                                                                                                      |
| dmr_down_cutoff       | A cutoff threshold for identifying DMRs that are not methylated in one cell type, but methylated in the other cell types.                                                                                                                                                           |
| dmr_pval_cutoff       | A cutoff threshold for the p-values when identifying DMRs that are methylated in one cell type, but not in the other cell types (or vice versa).                                                                                                                                    |
| cpg_pval_cutoff       | A cutoff threshold for the p-values when identifying differentially methylated CpGs that are methylated in one cell type, but not in the other cell types (or vice versa).                                                                                                          |

|                                   |                                                                                                                                                                                                                |
|-----------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <code>cpg_up_dm_cutoff</code>     | A cutoff threshold for identifying differentially methylated CpGs that are methylated in one cell type, but not in the other cell types.                                                                       |
| <code>cpg_down_dm_cutoff</code>   | A cutoff threshold for identifying differentially methylated CpGs that are not methylated in one cell type, but are methylated in the other cell types.                                                        |
| <code>pairwise_comparison</code>  | TRUE/FALSE of whether all pairwise comparisons (e.g. methylated in Granulocytes and Monocytes, but not methylated in other cell types). Default if FALSE.                                                      |
| <code>mset_train_flow_sort</code> | Default is NULL. However, a user can provide a <code>MethylSet</code> object after processing the <code>FlowSorted.Blood.450k</code> dataset. The default normalization is <code>preprocessIllumina()</code> . |

**Value**

A list of data frames and `GRanges` objects.

---

|                              |                        |
|------------------------------|------------------------|
| <code>.initializeMLEs</code> | <i>.initializeMLEs</i> |
|------------------------------|------------------------|

---

**Description**

Helper functions to initialize MLEs in `estimatecc()`.

**Usage**

```
.initializeMLEs(init_param_method, n, K, Ys, Zs, a0init, a1init, sig0init,
  sig1init, tauinit)
```

**Arguments**

|                                |                                                                                                                                                                                                                                               |
|--------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <code>init_param_method</code> | method to initialize parameter estimates. Choose between "random" (randomly sample) or "known_regions" (uses unmethylated and methylated regions that were identified based on Reinus et al. (2012) cell sorted data.). Defaults to "random". |
| <code>n</code>                 | Number of samples                                                                                                                                                                                                                             |
| <code>K</code>                 | Number of cell types                                                                                                                                                                                                                          |
| <code>Ys</code>                | observed methylation levels in samples provided by user of dimension $R \times n$                                                                                                                                                             |
| <code>Zs</code>                | Cell type specific regions of dimension $R \times K$                                                                                                                                                                                          |
| <code>a0init</code>            | Default NULL. Initial mean methylation level in unmethylated regions                                                                                                                                                                          |
| <code>a1init</code>            | Default NULL. Initial mean methylation level in methylated regions                                                                                                                                                                            |
| <code>sig0init</code>          | Default NULL. Initial var methylation level in unmethylated regions                                                                                                                                                                           |
| <code>sig1init</code>          | Default NULL. Initial var methylation level in methylated regions                                                                                                                                                                             |
| <code>tauint</code>            | Default NULL. Initial var for measurement error                                                                                                                                                                                               |

**Value**

A list of MLE estimates to be used in `estimatecc()`.

---

|                          |                          |
|--------------------------|--------------------------|
| <i>.initialize_theta</i> | <i>.initialize_theta</i> |
|--------------------------|--------------------------|

---

**Description**

Creates a container with initial theta parameter estimates

**Usage**

```
.initialize_theta(n, K, alpha0 = NULL, alpha1 = NULL, sig0 = NULL,  
  sig1 = NULL, tau = NULL)
```

**Arguments**

|        |                                                                      |
|--------|----------------------------------------------------------------------|
| n      | Number of samples                                                    |
| K      | Number of cell types                                                 |
| alpha0 | Default NULL. Initial mean methylation level in unmethylated regions |
| alpha1 | Default NULL. Initial mean methylation level in methylated regions   |
| sig0   | Default NULL. Initial var methylation level in unmethylated regions  |
| sig1   | Default NULL. Initial var methylation level in methylated regions    |
| tau    | Default NULL. Initial var for measurement error                      |

**Value**

A data frame with initial parameter estimates to be used in `.initializeMLEs()`.

---

|                         |                         |
|-------------------------|-------------------------|
| <i>.methylcc_engine</i> | <i>.methylcc_engine</i> |
|-------------------------|-------------------------|

---

**Description**

Helper function for `estimatecc`

**Usage**

```
.methylcc_engine(Ys, Zs, current_pi_mle, current_theta, epsilon, max_iter)
```

**Arguments**

|                             |                                                                                   |
|-----------------------------|-----------------------------------------------------------------------------------|
| <code>Ys</code>             | observed methylation levels in samples provided by user of dimension $R \times n$ |
| <code>Zs</code>             | Cell type specific regions of dimension $R \times K$                              |
| <code>current_pi_mle</code> | cell composition MLE estimates of dimension $K \times n$                          |
| <code>current_theta</code>  | other parameter estimates in EM algorithm                                         |
| <code>epsilon</code>        | Add here.                                                                         |
| <code>max_iter</code>       | Add here.                                                                         |

**Value**

A list of MLE estimates that is used in `estimatecc()`.

---

|                              |                         |
|------------------------------|-------------------------|
| <code>.methylcc_estep</code> | <i>Expectation step</i> |
|------------------------------|-------------------------|

---

**Description**

Expectation step in EM algorithm for methylCC

**Usage**

```
.methylcc_estep(Ys, Zs, current_pi_mle, current_theta, meth_status = 0)
```

**Arguments**

|                             |                                                                                                                                               |
|-----------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------|
| <code>Ys</code>             | observed methylation levels in samples provided by user of dimension $R \times n$                                                             |
| <code>Zs</code>             | Cell type specific regions of dimension $R \times K$                                                                                          |
| <code>current_pi_mle</code> | cell composition MLE estimates of dimension $K \times n$                                                                                      |
| <code>current_theta</code>  | other parameter estimates in EM algorithm                                                                                                     |
| <code>meth_status</code>    | Indicator function corresponding to regions that are unmethylated ( <code>meth_status=0</code> ) or methylated ( <code>meth_status=1</code> ) |

**Value**

List of expected value of the first two moments of the random effects (or the E-Step in the EM algorithm) used in `.methylcc_engine()`

---

|                              |                          |
|------------------------------|--------------------------|
| <code>.methylcc_mstep</code> | <i>Maximization step</i> |
|------------------------------|--------------------------|

---

**Description**

Maximization step in EM Algorithm for methylCC

**Usage**

```
.methylcc_mstep(Ys, Zs, current_pi_mle, current_theta, estep0, estep1)
```

**Arguments**

|                             |                                                                            |
|-----------------------------|----------------------------------------------------------------------------|
| <code>Ys</code>             | observed methylation levels in samples provided by user of dimension R x n |
| <code>Zs</code>             | Cell type specific regions of dimension R x K                              |
| <code>current_pi_mle</code> | cell composition MLE estimates of dimension K x n                          |
| <code>current_theta</code>  | other parameter estimates in EM algorithm                                  |
| <code>estep0</code>         | Results from expectation step for unmethylated regions                     |
| <code>estep1</code>         | Results from expectation step for methylated regions                       |

**Value**

A list of the updated MLEs (or the M-Step in the EM algorithm) used in `.methylcc_engine()`

---

|                                     |                              |
|-------------------------------------|------------------------------|
| <code>.pick_target_positions</code> | <i>Pick target positions</i> |
|-------------------------------------|------------------------------|

---

**Description**

Pick probes from target data using the indices in `dmp_regions`

**Usage**

```
.pick_target_positions(target_granges, target_object = NULL,  
  target_cvg = NULL, dmp_regions)
```

**Arguments**

|                             |                                                                                                                                                                                                       |
|-----------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <code>target_granges</code> | add more here.                                                                                                                                                                                        |
| <code>target_object</code>  | an optional argument which contains the meta-data for <code>target_granges</code> . If <code>target_granges</code> already contains the meta-data, do not need to supply <code>target_object</code> . |
| <code>target_cvg</code>     | coverage reads for the target object                                                                                                                                                                  |
| <code>dmp_regions</code>    | differentially methylated regions                                                                                                                                                                     |

**Value**

A list of GRanges objects to be used in `.preprocess_estimatecc()`

---

```
.preprocess_estimatecc
```

```
.preprocess_estimatecc
```

---

**Description**

This function preprocesses the data before the `estimatecc()` function

**Usage**

```
.preprocess_estimatecc(object, verbose = TRUE,
  init_param_method = "random",
  celltype_specific_dmrs = celltype_specific_dmrs)
```

**Arguments**

|                                     |                                                                                                                                                                                                                                               |
|-------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <code>object</code>                 | an object can be a <code>RGChannelSet</code> , <code>GenomicMethylSet</code> or <code>BSseq</code> object                                                                                                                                     |
| <code>verbose</code>                | TRUE/FALSE argument specifying if verbose messages should be returned or not. Default is TRUE.                                                                                                                                                |
| <code>init_param_method</code>      | method to initialize parameter estimates. Choose between "random" (randomly sample) or "known_regions" (uses unmethylated and methylated regions that were identified based on Reinus et al. (2012) cell sorted data.). Defaults to "random". |
| <code>celltype_specific_dmrs</code> | cell type specific differentially methylated regions (DMRs).                                                                                                                                                                                  |

**Value**

A list of object to be used in `estimatecc`

---

```
.splitit
```

```
.splitit
```

---

**Description**

helper function to split along a variable

**Usage**

```
.splitit(x)
```



### Arguments

x                      a vector

### Value

A list to be used in find\_dmrs()

---

|       |                                                                                |
|-------|--------------------------------------------------------------------------------|
| .WFun | <i>Helper function to take the product of Z and cell composition estimates</i> |
|-------|--------------------------------------------------------------------------------|

---

### Description

Helper function which is the product of Z and pi\_mle

### Usage

```
.WFun(Zs, pi_mle)
```

### Arguments

|        |                                               |
|--------|-----------------------------------------------|
| Zs     | Cell type specific regions of dimension R x K |
| pi_mle | cell composition MLE estimates                |

### Value

A list of output after taking the product of Z and cell composition mle estimates to be used in .methylcc\_estep().

---

|             |                                                                     |
|-------------|---------------------------------------------------------------------|
| cell_counts | <i>Generic function that returns the cell composition estimates</i> |
|-------------|---------------------------------------------------------------------|

---

### Description

Given a estimatecc object, this function returns the cell composition estimates

Accessors for the 'cell\_counts' slot of a estimatecc object.

### Usage

```
cell_counts(object)

## S4 method for signature 'estimatecc'
cell_counts(object)
```

### Arguments

object                      an object of class estimatecc.

**Value**

Returns the cell composition estimates

**Examples**

```
# This is a reduced version of the FlowSorted.Blood.450k
# dataset available by using BiocManager::install("FlowSorted.Blood.450k"),
# but for purposes of the example, we use the smaller version
# and we set \code{demo=TRUE}. For any case outside of this example for
# the package, you should set \code{demo=FALSE} (the default).

dir <- system.file("data", package="methylCC")
files <- file.path(dir, "FlowSorted.Blood.450k.sub.RData")
if(file.exists(files)){
  load(file = files)

  set.seed(12345)
  est <- estimatecc(object = FlowSorted.Blood.450k.sub, demo = TRUE)
  cell_counts(est)
}
```

---

estimatecc

---

*Estimate cell composition from DNAm data*


---

**Description**

Estimate cell composition from DNAm data

**Usage**

```
estimatecc(object, find_dmrs_object = NULL, verbose = TRUE,
  epsilon = 0.01, max_iter = 100, take_intersection = FALSE,
  include_cpgs = FALSE, include_dmrs = TRUE,
  init_param_method = "random", a0init = NULL, a1init = NULL,
  sig0init = NULL, sig1init = NULL, tauinit = NULL, demo = FALSE)
```

**Arguments**

|                  |                                                                                                                                                                              |
|------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| object           | an object can be a RGChannelSet, GenomicMethylSet or BSseq object                                                                                                            |
| find_dmrs_object | If the user would like to supply different differentially methylated regions, they can use the output from the find_dmrs function to supply different regions to estimatecc. |
| verbose          | TRUE/FALSE argument specifying if verbose messages should be returned or not. Default is TRUE.                                                                               |
| epsilon          | Threshold for EM algorithm to check for convergence. Default is 0.01.                                                                                                        |

|                   |                                                                                                                                                                                                                                               |
|-------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| max_iter          | Maximum number of iterations for EM algorithm. Default is 100 iterations.                                                                                                                                                                     |
| take_intersection | TRUE/FALSE asking if only the CpGs included in object should be used to find DMRs. Default is FALSE.                                                                                                                                          |
| include_cpgs      | TRUE/FALSE. Should individual CpGs be returned. Default is FALSE.                                                                                                                                                                             |
| include_dmrs      | TRUE/FALSE. Should differentially methylated regions be returned. Default is TRUE.                                                                                                                                                            |
| init_param_method | method to initialize parameter estimates. Choose between "random" (randomly sample) or "known_regions" (uses unmethylated and methylated regions that were identified based on Reinus et al. (2012) cell sorted data.). Defaults to "random". |
| a0init            | Default NULL. Initial mean methylation level in unmethylated regions                                                                                                                                                                          |
| a1init            | Default NULL. Initial mean methylation level in methylated regions                                                                                                                                                                            |
| sig0init          | Default NULL. Initial var methylation level in unmethylated regions                                                                                                                                                                           |
| sig1init          | Default NULL. Initial var methylation level in methylated regions                                                                                                                                                                             |
| tauinit           | Default NULL. Initial var for measurement error                                                                                                                                                                                               |
| demo              | TRUE/FALSE. Should the function be used in demo mode to shorten examples in package. Defaults to FALSE.                                                                                                                                       |

## Value

A object of the class `estimatecc` that contains information about the cell composition estimation (in the `summary` slot) and the cell composition estimates themselves (in the `cell_counts` slot).

## Examples

```
# This is a reduced version of the FlowSorted.Blood.450k
# dataset available by using BiocManager::install("FlowSorted.Blood.450k"),
# but for purposes of the example, we use the smaller version
# and we set \code{demo=TRUE}. For any case outside of this example for
# the package, you should set \code{demo=FALSE} (the default).

dir <- system.file("data", package="methylCC")
files <- file.path(dir, "FlowSorted.Blood.450k.sub.RData")
if(file.exists(files)){
  load(file = files)

  set.seed(12345)
  est <- estimatecc(object = FlowSorted.Blood.450k.sub, demo = TRUE)
  cell_counts(est)
}
```

---

|                  |                             |
|------------------|-----------------------------|
| estimatecc-class | <i>the estimatecc class</i> |
|------------------|-----------------------------|

---

## Description

Objects of this class store all the values needed information to work with a estimatecc object

## Value

summary returns the summary information about the cell composition estimate procedure and cell\_counts  
returns the cell composition estimates

## Slots

summary information about the samples and regions used to estimate cell composition  
cell\_counts cell composition estimates

## Examples

```
# This is a reduced version of the FlowSorted.Blood.450k
# dataset available by using BiocManager::install("FlowSorted.Blood.450k"),
# but for purposes of the example, we use the smaller version
# and we set \code{demo=TRUE}. For any case outside of this example for
# the package, you should set \code{demo=FALSE} (the default).

dir <- system.file("data", package="methylCC")
files <- file.path(dir, "FlowSorted.Blood.450k.sub.RData")
if(file.exists(files)){
  load(file = files)

  set.seed(12345)
  est <- estimatecc(object = FlowSorted.Blood.450k.sub, demo = TRUE)
  cell_counts(est)
}
```

---

FlowSorted.Blood.450k.sub

*A reduced size of the FlowSorted.Blood.450k dataset*

---

## Description

A reduced size of the FlowSorted.Blood.450k dataset

The object was created using the script in /inst and located in the /data folder.

## Format

A RGset object with 2e5 rows (probes) and 6 columns (whole blood samples).

---

|                |                                               |
|----------------|-----------------------------------------------|
| offMethRegions | <i>Unmethylated regions for all celltypes</i> |
|----------------|-----------------------------------------------|

---

**Description**

This is the script used to create the offMethRegions data set. The purpose is use in the estimate\_cc() function.

The object was created using the script in /inst and located in the /data folder.

**Format**

add more here.

---

|               |                                             |
|---------------|---------------------------------------------|
| onMethRegions | <i>Methylated regions for all celltypes</i> |
|---------------|---------------------------------------------|

---

**Description**

This is the script used to create the onMethRegions data set. The purpose is use in the estimate\_cc() function.

The object was created using the script in /inst and located in the /data folder.

**Format**

add more here.

# Index

`.WFun`, [9](#)  
`.extract_raw_data`, [2](#)  
`.find_dmrs`, [3](#)  
`.initializeMLEs`, [4](#)  
`.initialize_theta`, [5](#)  
`.methylcc_engine`, [5](#)  
`.methylcc_estep`, [6](#)  
`.methylcc_mstep`, [7](#)  
`.pick_target_positions`, [7](#)  
`.preprocess_estimatecc`, [8](#)  
`.splitit`, [8](#)  
  
`cell_counts`, [9](#)  
`cell_counts,estimatecc-method`  
    (`cell_counts`), [9](#)  
  
`estimatecc`, [10](#)  
`estimatecc-class`, [12](#)  
  
`FlowSorted.Blood.450k.sub`, [12](#)  
  
`offMethRegions`, [13](#)  
`onMethRegions`, [13](#)