

# Package ‘methodical’

May 16, 2026

**Title** Discovering genomic regions where methylation is strongly associated with transcriptional activity

**Version** 1.9.0

## Description

DNA methylation is generally considered to be associated with transcriptional silencing. However, comprehensive, genome-wide investigation of this relationship requires the evaluation of potentially millions of correlation values between the methylation of individual genomic loci and expression of associated transcripts in a relatively large numbers of samples. Methodical makes this process quick and easy while keeping a low memory footprint. It also provides a novel method for identifying regions where a number of methylation sites are consistently strongly associated with transcriptional expression. In addition, Methodical enables housing DNA methylation data from diverse sources (e.g. WGBS, RRBS and methylation arrays) with a common framework, lifting over DNA methylation data between different genome builds and creating base-resolution plots of the association between DNA methylation and transcriptional activity at transcriptional start sites.

**License** GPL (>= 3)

**BugReports** <https://github.com/richardheery/methodical/issues>

**biocViews** DNAMethylation, MethylationArray, Transcription, GenomeWideAssociation, Software

**Encoding** UTF-8

**Roxygen** list(markdown = TRUE)

**RoxygenNote** 7.3.3

**Depends** GenomicRanges, ggplot2, R (>= 4.0), SummarizedExperiment

**LazyData** false

**Imports** AnnotationHub, BiocCheck, BiocManager, BiocParallel, BiocStyle, Biostrings, BSgenome, bsseq, cowplot, data.table, DelayedArray, devtools, dplyr, ExperimentHub, foreach, GenomeInfoDb, HDF5Array, IRanges, knitr, MatrixGenerics, R.utils, rcmdcheck, RcppRoll, remotes, rhdf5, rtracklayer, S4Vectors, scales, tibble, tidyr, tools, usethis

**Suggests** DESeq2, rmarkdown, TumourMethData

**VignetteBuilder** knitr

**URL** <https://github.com/richardheery/methodical>

**Config/Bioconductor/UnsupportedPlatforms** windows

**git\_url** <https://git.bioconductor.org/packages/methodical>  
**git\_branch** devel  
**git\_last\_commit** 1387d02  
**git\_last\_commit\_date** 2026-05-01  
**Repository** Bioconductor 3.24  
**Date/Publication** 2026-05-15  
**Author** Richard Heery [aut, cre] (ORCID:  
<https://orcid.org/0000-0001-8067-3114>)  
**Maintainer** Richard Heery <[richardheery@gmail.com](mailto:richardheery@gmail.com)>

## Contents

methodical-package . . . . .	3
.check_input_files . . . . .	4
.chunk_regions . . . . .	4
.collapse_strands . . . . .	5
.create_meth_rse_from_hdf5 . . . . .	5
.find_tmrs_single . . . . .	6
.make_meth_rse_setup . . . . .	7
.set_meth_df_columns . . . . .	8
.split_meth_file . . . . .	8
.split_meth_files_into_chunks . . . . .	9
.summarize_chunk_methylation . . . . .	10
.test_tmrs . . . . .	10
.tss_correlations . . . . .	11
.tss_iterator . . . . .	11
.write_chunks_to_hdf5 . . . . .	12
annotatePlot . . . . .	13
calculateMethSiteTranscriptCors . . . . .	14
calculateRegionMethylationTranscriptCors . . . . .	16
calculateSmoothedMethodicalScores . . . . .	18
chr11_subset_hg38_cpgs . . . . .	19
chr1_subset_hg38 . . . . .	19
chr4_subset_a_thal . . . . .	20
convert_rse_to_bsseq . . . . .	20
correct_correlation_pvalues . . . . .	21
expand_granges . . . . .	21
export_bedGraph_from_rse . . . . .	22
extractGRangesMethSiteValues . . . . .	22
extractMethSitesFromGenome . . . . .	23
findTMRs . . . . .	24
hg19_chr18_cpgs . . . . .	25
hg38_cpg_islands . . . . .	25
infinium_450k_probe_granges_hg19 . . . . .	26
liftoverMethRSE . . . . .	26
makeMethRSEFromInputFiles . . . . .	27
maskRangesInRSE . . . . .	30
methrixToRSE . . . . .	31
plotMethodicalScores . . . . .	31

plotMethSiteCorCoefs . . . . .	33
plotRegionValues . . . . .	34
plotTMRs . . . . .	35
rangesRelativeToTSS . . . . .	36
rapidCorTest . . . . .	37
sampleMethSites . . . . .	38
strandedDistance . . . . .	39
summarizeRegionMethylation . . . . .	40
tubb6_correlation_plot . . . . .	41
tubb6_cpg_meth_transcript_cors . . . . .	42
tubb6_meth_rse . . . . .	42
tubb6_tmrs . . . . .	43
tubb6_transcript_counts . . . . .	43
tubb6_tss . . . . .	44
<b>Index</b>	<b>45</b>

---

methodical-package	<i>methodical: A one-stop shop for dealing with big DNA methylation datasets</i>
--------------------	--

---

## Description

DNA methylation is generally considered to be associated with transcriptional silencing. However, comprehensive, genome-wide investigation of this relationship requires the evaluation of potentially millions of correlation values between the methylation of individual genomic loci and expression of associated transcripts in a relatively large numbers of samples. Methodical makes this process quick and easy while keeping a low memory footprint. It also provides a novel method for identifying regions where a number of methylation sites are consistently strongly associated with transcriptional expression. In addition, Methodical enables housing DNA methylation data from diverse sources (e.g. WGBS, RRBS and methylation arrays) with a common framework, lifting over DNA methylation data between different genome builds and creating base-resolution plots of the association between DNA methylation and transcriptional activity at transcriptional start sites.

## Author(s)

Richard Heery

## See Also

Useful links:

- <https://github.com/richardheery/methodical>
- Report bugs at <https://github.com/richardheery/methodical/issues>

---

<code>.check_input_files</code>	<i>Check input files have the correct number of columns specified and that columns seem to be of correct type</i>
---------------------------------	---

---

### Description

Check input files have the correct number of columns specified and that columns seem to be of correct type

### Usage

```
.check_input_files(input_files, meth_files_columns)
```

### Arguments

<code>input_files</code>	A vector of input filepaths.
<code>meth_files_columns</code>	A list specifying the columns in <code>meth_files</code> .

---

<code>.chunk_regions</code>	<i>Split genomic regions into balanced chunks based on the number of methylation sites that they cover</i>
-----------------------------	--

---

### Description

Split genomic regions into balanced chunks based on the number of methylation sites that they cover

### Usage

```
.chunk_regions(  
  meth_rse,  
  genomic_regions,  
  max_sites_per_chunk = NULL,  
  ncores = 1  
)
```

### Arguments

<code>meth_rse</code>	A RangedSummarizedExperiment with methylation values.
<code>genomic_regions</code>	A GRanges object.
<code>max_sites_per_chunk</code>	The maximum number of methylation sites to load into memory at once for each chunk.
<code>ncores</code>	The number of cores that will be used.

### Value

A GRangesList where each GRanges object overlaps approximately the number of methylation sites given by `max_sites_per_chunk`

---

.collapse\_strands      *Combine values for stranded data*

---

### Description

Combine values for stranded data

### Usage

```
.collapse_strands(meth_df, sequence_context)
```

### Arguments

meth\_df            A data.frame with methylation data.  
sequence\_context    A single character string or DNASTring with the sequence context of the methylation sites e.g. CG or CHG.

---

.create\_meth\_rse\_from\_hdf5  
*Create a RangedSummarizedExperiment for methylation values already deposited in HDF5*

---

### Description

Create a RangedSummarizedExperiment for methylation values already deposited in HDF5

### Usage

```
.create_meth_rse_from_hdf5(  
  hdf5_filepath,  
  hdf5_dir,  
  meth_sites,  
  sample_metadata  
)
```

### Arguments

hdf5\_filepath    Path to HDF5 file  
hdf5\_dir        The path to the HDF5 directory.  
meth\_sites      A sorted GRanges object with the locations of the methylation sites of interest.  
sample\_metadata    A data.frame with sample metadata

### Value

A RangedSummarizedExperiment with methylation values

---

`.find_tmrs_single`      *Find TSS-Proximal Methylation-Controlled Regulatory Sites (TMRs)*

---

### Description

Find TSS-Proximal Methylation-Controlled Regulatory Sites (TMRs)

### Usage

```
.find_tmrs_single(
  correlation_df,
  offset_length = 10,
  p_value_threshold = 0.05,
  smoothing_factor = 0.75,
  min_gapwidth = 150,
  min_meth_sites = 5
)
```

### Arguments

`correlation_df` A data.frame with correlation values between methylation sites and a transcript or a path to an RDS file containing such a data.frame as returned by `calculateMethSiteTranscriptCors`.

`offset_length` Number of methylation sites added upstream and downstream of a central methylation site to form a window, resulting in a window size of  $2 * \text{offset\_length} + 1$ . Default value is 10.

`p_value_threshold` The p\_value cutoff to use. Default value is 0.05.

`smoothing_factor` Smoothing factor for exponential moving average. Should be a value between 0 and 1 with higher values resulting in a greater degree of smoothing. Default is 0.75.

`min_gapwidth` Merge TMRs with the same direction separated by less than this number of base pairs. Default value is 150.

`min_meth_sites` Minimum number of methylation sites that TMRs can contain. Default value is 5.

### Value

A GRanges object with the location of TMRs.

### Examples

```
# Load methylation-transcript correlation results for TUBB6 gene
data("tubb6_cpg_meth_transcript_cors", package = "methodical")

# Find TMRs for
tubb6_tmrs <- methodical:::.find_tmrs_single(correlation_df = tubb6_cpg_meth_transcript_cors)
print(tubb6_tmrs)
```

---

.make\_meth\_rse\_setup *Perform setup for makeMethRSEFromInputFiles or makeMethRSEFromArrayFiles*

---

### Description

Perform setup for makeMethRSEFromInputFiles or makeMethRSEFromArrayFiles

### Usage

```
.make_meth_rse_setup(  
  meth_files,  
  meth_sites,  
  sample_metadata,  
  hdf5_dir,  
  overwrite,  
  chunkdim,  
  temporary_dir,  
  ...  
)
```

### Arguments

meth_files	A vector of paths to files with methylation values. Automatically detects if meth_files contain a header if every field in the first line is a character.
meth_sites	A GRanges object with the locations of the methylation sites of interest. Should contain separate ranges for each strand if meth_files are stranded (i.e. separate ranges for the C and G positions of CpG sites). Any positions in meth_files that are not in meth_sites are ignored.
sample_metadata	A data.frame with sample metadata to be used as colData for the RangedSummarizedExperiment.
hdf5_dir	Directory to save HDF5 file. Is created if it doesn't exist. HDF5 file is called assays.h5.
overwrite	TRUE or FALSE indicating whether to allow overwriting if hdf5_dir already exists.
chunkdim	The dimensions of the chunks for the HDF5 file.
temporary_dir	Name to give a temporary directory to store intermediate files. A directory with this name cannot already exist.
...	Additional arguments to be passed to HDF5Array::HDF5RealizationSink.

### Value

A list describing the setup to be used for makeMethRSEFromInputFiles or makeMethRSEFromArrayFiles.

---

`.set_meth_df_columns`    *Process a data.frame with methylation data so that it contains the correct columns*

---

**Description**

Process a data.frame with methylation data so that it contains the correct columns

**Usage**

```
.set_meth_df_columns(meth_df, zero_based)
```

**Arguments**

<code>meth_df</code>	A data.frame with methylation data.
<code>zero_based</code>	TRUE or FALSE indicating if files are zero-based.

---

`.split_meth_file`    *Split data from a single input methylation file into chunks*

---

**Description**

Split data from a single input methylation file into chunks

**Usage**

```
.split_meth_file(  
  meth_file,  
  meth_files_columns,  
  grid_column,  
  file_count,  
  parameters  
)
```

**Arguments**

<code>meth_file</code>	Path to an input methylation file.
<code>meth_files_columns</code>	A list specifying the columns in meth_files.
<code>grid_column</code>	The current grid column being processed.
<code>file_count</code>	The number of the current file being processed.
<code>parameters</code>	A list of parameters for processing meth_file.

**Value**

Invisibly returns NULL.

---

`.split_meth_files_into_chunks`

*Split data from input methylation files into chunks*

---

### **Description**

Split data from input methylation files into chunks

### **Usage**

```
.split_meth_files_into_chunks(  
  meth_files,  
  meth_files_columns,  
  file_grid_columns,  
  meth_sites_df,  
  collapse_strands,  
  sequence_context,  
  meth_site_groups,  
  temp_chunk_dirs,  
  zero_based,  
  decimal_places,  
  BPPARAM  
)
```

### **Arguments**

<code>meth_files</code>	Paths to input methylation files.
<code>meth_files_columns</code>	A list specifying the columns in <code>meth_files</code> .
<code>file_grid_columns</code>	The grid column number for each file.
<code>meth_sites_df</code>	A <code>data.table</code> with the positions of methylation sites.
<code>collapse_strands</code>	TRUE or FALSE indicating whether or not to collapse data on + and - strands.
<code>sequence_context</code>	A single character string or <code>DNAStr</code> with the sequence context of the methylation sites e.g. CG or CHG.
<code>meth_site_groups</code>	A list with the indices of the methylation sites in each group.
<code>temp_chunk_dirs</code>	A vector giving the temporary directory associated with each chunk.
<code>zero_based</code>	TRUE or FALSE indicating if files are zero-based.
<code>decimal_places</code>	Integer indicating the number of decimal places to round beta values to.
<code>BPPARAM</code>	A <code>BiocParallelParam</code> object.

### **Value**

Invisibly returns `NULL`.

---

```
.summarize_chunk_methylation
```

*Summarize methylation values for regions in a chunk*

---

### Description

Summarize methylation values for regions in a chunk

### Usage

```
.summarize_chunk_methylation(
  chunk_regions,
  meth_rse,
  assay,
  col_summary_function,
  na.rm,
  ...
)
```

### Arguments

chunk_regions	Chunk with genomic regions of interest.
meth_rse	A RangedSummarizedExperiment with methylation values.
assay	The assay from meth_rse to extract values from. Should be either an index or the name of an assay.
col_summary_function	A function that summarizes column values.
na.rm	TRUE or FALSE indicating whether to remove NA values when calculating summaries.
...	Additional arguments to be passed to col_summary_function.

### Value

A function which returns a list with the summarized methylation values for regions in each sample.

---

```
.test_tmrs
```

*Find TMRs where smoothed methodical scores exceed thresholds*

---

### Description

Find TMRs where smoothed methodical scores exceed thresholds

### Usage

```
.test_tmrs(
  meth_sites_gr,
  smoothed_methodical_scores,
  p_value_threshold = 0.05,
  tss_gr = NULL,
  transcript_id = NULL
)
```

### Arguments

- meth\_sites\_gr A GRanges object with the location of methylation sites.
- smoothed\_methodical\_scores A numeric vector with the smoothed methodical scores associated with each methylation site.
- p\_value\_threshold The p\_value cutoff to use. Default value is 0.05.
- tss\_gr An optional GRanges object giving the location of the TSS meth\_sites\_gr is associated with.
- transcript\_id Name of the transcript associated with the TSS.

### Value

A GRanges object with the location of TMRs.

---

.tss\_correlations *Calculate meth site-transcript correlations for given TSS*

---

### Description

Calculate meth site-transcript correlations for given TSS

### Usage

```
.tss_correlations(correlation_objects)
```

### Arguments

- correlation\_objects A list with a table of methylation values, expression values for transcripts, a GRangesList for the transcript and the name of the transcript.

### Value

A data.frame with the correlation values

---

.tss\_iterator *Create an iterator function for use with biterate*

---

### Description

Create an iterator function for use with biterate

**Usage**

```
.tss_iterator(
  meth_values_chunk,
  tss_region_indices_list,
  transcript_values_list,
  tss_gr_chunk_list,
  cor_method,
  add_distance_to_region,
  min_number_complete_pairs,
  results_dir
)
```

**Arguments**

`meth_values_chunk` A table with methylation values for current chunk

`tss_region_indices_list` A list with the indices for methylation sites associated with each TSS.

`transcript_values_list` A list with expression values for transcripts.

`tss_gr_chunk_list` A list of GRanges with the TSS for the current chunk.

`cor_method` Correlation method to use.

`add_distance_to_region` TRUE or FALSE indicating whether to add distance to TSS.

`min_number_complete_pairs` The minimum number of complete pairs required to return a p-value for a correlation.

`results_dir` Location of results directory.

**Value**

An iterator function which returns a list with the parameters necessary for `.tss_correlations`.

---

*.write\_chunks\_to\_hdf5* Write chunks of data to a HDF5 sink

---

**Description**

Write chunks of data to a HDF5 sink

**Usage**

```
.write_chunks_to_hdf5(
  temp_chunk_dirs,
  files_in_chunks,
  beta_sink,
  Cov_sink,
  hdf5_grid
)
```

**Arguments**

- temp\_chunk\_dirs            A vector giving the temporary directory associated with each chunk.
- files\_in\_chunks            A list of files associated with each chunk in the order they should be placed.
- beta\_sink                  A HDF5RealizationSink for methylation proportions.
- Cov\_sink                    A HDF5RealizationSink for coverage.
- hdf5\_grid                  A RegularArrayGrid.

**Value**

Invisibly returns TRUE.

---

annotatePlot	<i>Create a plot with genomic annotation for a plot of values at methylation sites.</i>
--------------	---

---

**Description**

Works with plots returned by plotRegionValues(), plotMethSiteCorCoefs() or plotMethodicalScores. Can combine the meth site values plot and genomic annotation together into a single plot or return the annotation plot separately.

**Usage**

```

annotatePlot(
  meth_site_plot,
  annotation_grl,
  reference_tss = FALSE,
  grl_colours = NULL,
  annotation_line_size = 5,
  ylab = "Genome Annotation",
  annotation_plot_proportion = 0.5,
  keep_meth_site_plot_legend = FALSE,
  annotation_plot_only = FALSE
)

```

**Arguments**

- meth\_site\_plot    A plot of methylation site values (generally methylation level or correlation of methylation with transcription) around a TSS
- annotation\_grl    A GRangesList object (or list coercible to a GRangesList) where each component GRanges gives the locations of different classes of regions to display. Each class of region will be given a separate colour in the plot, with regions ordered by the order of names(annotation\_grl).
- reference\_tss     TRUE or FALSE indicating whether to show distances on the X-axis relative to the TSS stored as an attribute tss\_range of meth\_site\_plot. Alternatively, can provide a GRanges object with a single range for such a TSS site. In either case, will show the distance of methylation sites to the start of this region with

- methylation sites upstream relative to the reference\_tss shown first. If FALSE (the default), the x-axis will instead show the start site coordinate of the methylation site. relative to the reference\_tss shown first. If not, the x-axis will show the start site coordinate of the methylation site.
- `grl_colours` An optional vector of colours used to display each of the GRanges making up `annotation_grl`. Must have same length as `annotation_grl`.
- `annotation_line_size` Linewidth for annotation plot. Default is 5.
- `ylab` The title to give the Y axis in the annotation plot. Default is "Genome Annotation".
- `annotation_plot_proportion` A value giving the proportion of the height of the plot devoted to the annotation. Default is 0.5.
- `keep_meth_site_plot_legend` TRUE or FALSE indicating whether to retain the legend of `meth_site_plot`, if it has one. Default value is FALSE.
- `annotation_plot_only` TRUE or FALSE indicating whether to return only the annotation plot. Default is to combine `meth_site_plot` with the annotation.

### Value

A ggplot object

### Examples

```
# Get CpG islands from UCSC
data("hg38_cpg_islands", package = "methodical")
hg38_cpg_islands <- GRangesList(split(hg38_cpg_islands, hg38_cpg_islands$type))

# Load plot with CpG methylation correlation values for TUBB6
data("tubb6_correlation_plot", package = "methodical")

# Add positions of CpG islands to tubb6_correlation_plot
methodical::annotatePlot(tubb6_correlation_plot, annotation_grl = hg38_cpg_islands, annotation_plot_proporti
```

---

calculateMethSiteTranscriptCors

*Calculate correlation between expression of transcripts and methylation of sites surrounding their TSS*

---

### Description

Calculate correlation between expression of transcripts and methylation of sites surrounding their TSS

**Usage**

```
calculateMethSiteTranscriptCors(
  meth_rse,
  assay_number = 1,
  transcript_expression_table,
  samples_subset = NULL,
  tss_gr,
  tss_associated_gr,
  cor_method = "pearson",
  min_number_complete_pairs = 30,
  add_distance_to_region = TRUE,
  max_sites_per_chunk = NULL,
  BPPARAM = BiocParallel::SerialParam(),
  results_dir = NULL
)
```

**Arguments**

**meth\_rse** A RangedSummarizedExperiment for methylation sites.

**assay\_number** The assay from meth\_rse to extract values from. Default is the first assay.

**transcript\_expression\_table**  
A matrix or data.frame with the expression values for transcripts, where row names are transcript names and columns sample names. There should be a row corresponding to each transcript associated with each range in tss\_gr. Names of samples must match those in meth\_rse unless samples\_subset provided.

**samples\_subset** Sample names used to subset meth\_rse and transcript\_expression\_table. Provided samples must be found in both meth\_rse and transcript\_expression\_table. Default is to use all samples in meth\_rse and transcript\_expression\_table.

**tss\_gr** A GRanges object with the locations of transcription start sites. Names of regions cannot contain any duplicates and should and match those of tss\_associated\_gr and be present in transcript\_expression table.

**tss\_associated\_gr**  
A GRanges object with the locations of regions associated with each transcription start site. Names of regions cannot contain any duplicates and should and match those of tss\_gr and be present in transcript\_expression table.

**cor\_method** A character string indicating which correlation coefficient is to be computed. One of either "pearson" or "spearman" or their abbreviations.

**min\_number\_complete\_pairs**  
The minimum number of complete pairs required to return a p-value for a correlation. Correlations with less than this number are given a p-value of NaN. Default value is 30.

**add\_distance\_to\_region**  
TRUE or FALSE indicating whether to add the distance of methylation sites to the TSS. Default value is TRUE. Setting to FALSE will roughly half the total running time.

**max\_sites\_per\_chunk**  
The approximate maximum number of methylation sites to try to load into memory at once. The actual number loaded may vary depending on the number of methylation sites overlapping each region, but so long as the size of any individual regions is not enormous ( $\geq$  several MB), it should vary only very

slightly. Some experimentation may be needed to choose an optimal value as low values will result in increased running time, while high values will result in a large memory footprint without much improvement in running time. Default is  $\text{floor}(62500000/\text{ncol}(\text{meth\_rse}))$ , resulting in each chunk requiring approximately 500 MB of RAM.

**BPPARAM** A BiocParallelParam object for parallel processing. Defaults to `BiocParallel::SerialParam()`.

**results\_dir** An optional path to a directory to save results as RDS files. There will be one RDS file for each transcript. If not provided, returns the results as a list.

### Value

If `results_dir` is `NULL`, a list of data.frames with the correlation of methylation sites surrounding a specified genomic region with a given feature, p-values and adjusted q-values for the correlations. Distance of the methylation sites upstream or downstream to the center of the region is also provided. If `results_dir` is provided, instead returns a list with the paths to the RDS files with the results.

### Examples

```
# Load TUBB6 TSS GRanges, RangedSummarizedExperiment with methylation values for CpGs around TUBB6 TSS and TUBB6
data(tubb6_tss, package = "methodical")
data(tubb6_meth_rse, package = "methodical")
tubb6_meth_rse <- eval(tubb6_meth_rse)
data(tubb6_transcript_counts, package = "methodical")

# Calculate correlation values between methylation values and transcript values for TUBB6
tubb6_cpg_meth_transcript_cors <- methodical::calculateMethSiteTranscriptCors(meth_rse = tubb6_meth_rse,
  transcript_expression_table = tubb6_transcript_counts, tss_gr = tubb6_tss,
  tss_associated_gr = methodical::expand_granges(tubb6_tss, upstream = 5000, downstream = 5000))
head(tubb6_cpg_meth_transcript_cors$ENST00000591909)
```

---

calculateRegionMethylationTranscriptCors

*Calculate the correlation values between the methylation of genomic regions and the expression of associated transcripts*

---

### Description

Calculate the correlation values between the methylation of genomic regions and the expression of associated transcripts

### Usage

```
calculateRegionMethylationTranscriptCors(
  meth_rse,
  assay = 1,
  transcript_expression_table,
  samples_subset = NULL,
  genomic_regions,
  genomic_region_names = NULL,
  genomic_region_transcripts = NULL,
```

```

    genomic_region_methylation = NULL,
    cor_method = "pearson",
    p_adjust_method = "BH",
    region_methylation_summary_function = colMeans,
    BPPARAM = BiocParallel::SerialParam(),
    ...
)

```

## Arguments

**meth\_rse** A RangedSummarizedExperiment with methylation values for CpG sites which will be used to calculate methylation values for genomic\_regions. There must be at least 3 samples in common between meth\_rse and transcript\_expression\_table.

**assay** The assay from meth\_rse to extract values from. Should be either an index or the name of an assay. Default is the first assay.

**transcript\_expression\_table** A table with the expression values for different transcripts in different samples. Row names should give be the transcript name and column names should be the name of samples.

**samples\_subset** Optional sample names used to subset meth\_rse and transcript\_expression\_table. Provided samples must be found in both meth\_rse and transcript\_expression\_table. Default is to use all samples in meth\_rse and transcript\_expression\_table.

**genomic\_regions** A GRanges object.

**genomic\_region\_names** A character vector of unique names to assign genomic\_regions in the output table. Defaults to names(genomic\_regions) if present or otherwise converts regions to character strings (e.g. "chr:1000-2000") to use as names.

**genomic\_region\_transcripts** Names of transcripts associated with each region in genomic\_regions. If not provided, attempts to use genomic\_regions\$transcript\_id. All transcripts must be present in transcript\_expression\_table.

**genomic\_region\_methylation** Optional preprovided table with methylation values for genomic\_regions such as created using summarizeRegionMethylation(). Table will be created if it is not provided which will increase running time. Row names should match genomic\_region\_names and column names should match those of transcript\_expression\_table

**cor\_method** A character string indicating which correlation coefficient is to be computed. One of either "pearson" or "spearman" or their abbreviations.

**p\_adjust\_method** Method used to adjust p-values. Same as the methods from p.adjust.methods. Default is Benjamini-Hochberg.

**region\_methylation\_summary\_function** A function that summarizes column values. Default is colMeans.

**BPPARAM** A BiocParallelParam object for parallel processing. Defaults to BiocParallel::SerialParam().

**...** Additional arguments to be passed to summary\_function.

## Value

A data.frame with the correlation values between the methylation of genomic regions and expression of transcripts associated with them

**Examples**

```
# Load TUBB6 TMRs, RangedSummarizedExperiment with methylation values for CpGs around TUBB6 TSS and TUBB6 transcripts
data(tubb6_tmrs, package = "methodical")
data(tubb6_meth_rse, package = "methodical")
tubb6_meth_rse <- eval(tubb6_meth_rse)
data(tubb6_transcript_counts, package = "methodical")

# Calculate correlation values between TMRs identified for TUBB6 and transcript expression
tubb6_tmrs_transcript_cors <- methodical::calculateRegionMethylationTranscriptCors(
  meth_rse = tubb6_meth_rse, transcript_expression_table = tubb6_transcript_counts,
  genomic_regions = tubb6_tmrs, genomic_region_names = tubb6_tmrs$tmr_name)
tubb6_tmrs_transcript_cors
```

---

```
calculateSmoothedMethodicalScores
```

*Calculate methodical score and smooth it using a exponential weighted moving average*

---

**Description**

Calculate methodical score and smooth it using a exponential weighted moving average

**Usage**

```
calculateSmoothedMethodicalScores(
  correlation_df,
  offset_length = 10,
  smoothing_factor = 0.75
)
```

**Arguments**

**correlation\_df** A data.frame with correlation values between methylation sites and a transcript as returned by calculateMethSiteTranscriptCors.

**offset\_length** Number of methylation sites added upstream and downstream of a central methylation site to form a window, resulting in a window size of  $2 \times \text{offset\_length} + 1$ . Default value is 10.

**smoothing\_factor** Smoothing factor for exponential moving average. Should be a value between 0 and 1 with higher values resulting in a greater degree of smoothing. Default is 0.75.

**Value**

A GRanges object

**Examples**

```
# Load data.frame with CpG methylation-transcription correlation results for TUBB6
data("tubb6_cpg_meth_transcript_cors", package = "methodical")

# Calculate smoothed Methodical scores from correlation values
smoothed_methodical_scores <- methodical::calculateSmoothedMethodicalScores(tubb6_cpg_meth_transcript_cors)
```

---

```
chr11_subset_hg38_cpgs      chr11_subset_hg38_cpgs
```

---

**Description**

All the CpG sites on both strands within the region chr11:67578812-67588812

**Usage**

```
chr11_subset_hg38_cpgs
```

**Format**

A GRanges object.

---

```
chr1_subset_hg38      chr1_subset_hg38
```

---

**Description**

A DNASTringSet for first 1,000,000 bp from chr1 from BSgenome.Hsapiens.UCSC.hg38

**Usage**

```
chr1_subset_hg38
```

**Format**

A DNASTringSet object.

---

```
chr4_subset_a_thal    chr4_subset_a_thal
```

---

**Description**

A DNASTringSet for first 1,000,000 bp from chr4 from BSgenome.Athaliana.TAIR.TAIR9

**Usage**

```
chr4_subset_a_thal
```

**Format**

A DNASTringSet object.

---

```
convert_rse_to_bsseq    Convert a methylation RSE to a BSseq object
```

---

**Description**

Convert a methylation RSE to a BSseq object

**Usage**

```
convert_rse_to_bsseq(meth_rse, proportion_assay = 1, coverage_assay = 2)
```

**Arguments**

meth_rse	A RangedSummarizedExperiment with methylation values.
proportion_assay	The assay of meth_rse which corresponds to the proportion of methylated reads. Can be either a numeric index or the name of the assay. Default is the first assay.
coverage_assay	The assay of meth_rse which corresponds to the coverage (the total number of reads). Can be either a numeric index or the name of the assay. Default is the second assay.

**Value**

A RangedSummarized experiment identical to meth\_rse with two additional assays added: one for methylated reads and another for unmethylated reads.

---

correct\_correlation\_pvalues

*Correct p-values for a list of methylation-transcription correlations results*

---

### Description

Correct p-values for a list of methylation-transcription correlations results

### Usage

```
correct_correlation_pvalues(correlation_list, p_adjust_method = "fdr")
```

### Arguments

correlation\_list

A list of data.frames with correlation values between methylation sites and a transcript as returned by calculateMethSiteTranscriptCors.

p\_adjust\_method

The method to use for p-value adjustment. Should be one of the methods in p.adjust.methods. Default is "fdr".

### Value

A list identical to correlation\_list except with p-values corrected using the indicated method.

---

expand\_granges

*Expand GRanges upstream and downstream*

---

### Description

Expand ranges in a GRanges object upstream and downstream by specified numbers of bases, taking account of strand. Unstranded ranges are treated like they on the "+" strand. If any of the resulting ranges are out-of-bounds given the seqinfo of genomic\_regions, they will be trimmed using trim().

### Usage

```
expand_granges(genomic_regions, upstream = 0, downstream = 0)
```

### Arguments

genomic\_regions

A GRanges object

upstream

Number of bases to add upstream of each region in genomic\_regions. Must be numeric vector of length 1 or else equal to the length of genomic\_regions. Default value is 0. Negative values result in upstream end of regions being shortened, however the width of the resulting regions cannot be less than zero.

downstream

Number of bases to add downstream of each region in genomic\_regions. Negative values result in downstream end of regions being shortened. Must be numeric vector of length 1 or else equal to the length of genomic\_regions. Default value is 0. Negative values result in upstream end of regions being shortened, however the width of the resulting regions cannot be less than zero.

**Value**

A GRanges object

**Examples**

```
data(tubb6_tss, package = "methodical")
tubb6_tss
methodical::expand_granges(tubb6_tss, upstream = 5000, downstream = 5000)
```

---

```
export_bedGraph_from_rse
```

*Export values for a sample in a RangedSummarizedExperiment as a bedGraph*

---

**Description**

Export values for a sample in a RangedSummarizedExperiment as a bedGraph

**Usage**

```
export_bedGraph_from_rse(meth_rse, assay_number = 1, sample_name, file_name)
```

**Arguments**

meth_rse	A RangedSummarizedExperiment for methylation data.
assay_number	The assay from meth_rse to extract values from. Default is the first assay.
sample_name	The name of a single sample in meth_rse.
file_name	The output filename.

**Value**

A data.frame with the methylation site values for all sites in meth\_rse which overlap genomic\_ranges. Row names are the coordinates of the sites as a character vector.

---

```
extractGRangesMethSiteValues
```

*Extract values for methylation sites overlapping genomic regions from a methylation RSE.*

---

**Description**

Extract values for methylation sites overlapping genomic regions from a methylation RSE.

**Usage**

```
extractGRangesMethSiteValues(
  meth_rse,
  genomic_regions = NULL,
  samples_subset = NULL,
  assay_number = 1
)
```

**Arguments**

meth\_rse            A RangedSummarizedExperiment for methylation data.  
 genomic\_regions    A GRanges object. If set to NULL, returns all methylation sites in meth\_rse  
 samples\_subset    Optional sample names used to subset meth\_rse.  
 assay\_number      The assay from meth\_rse to extract values from. Default is the first assay.

**Value**

A data.frame with the methylation site values for all sites in meth\_rse which overlap genomic\_ranges.  
 Row names are the coordinates of the sites as a character vector.

**Examples**

```
# Load sample RangedSummarizedExperiment with CpG methylation data
data(tubb6_meth_rse, package = "methodical")
tubb6_meth_rse <- eval(tubb6_meth_rse)

# Create a sample GRanges object to use
test_region <- GRanges("chr18:12305000-12310000")

# Get methylation values for CpG sites overlapping HDAC1 gene
test_region_methylation <- methodical::extractGRangesMethSiteValues(meth_rse = tubb6_meth_rse, genomic_regio
```

---

extractMethSitesFromGenome

*Create a GRanges with methylation sites of interest from a BSgenome or DNASTringSet*

---

**Description**

Create a GRanges with methylation sites of interest from a BSgenome or DNASTringSet

**Usage**

```
extractMethSitesFromGenome(  
  genome,  
  pattern = "CG",  
  stranded = TRUE,  
  standard_seqs_only = FALSE  
)
```

**Arguments**

genome            A BSgenome object or a DNASTringSet with names indicating the sequences.  
 pattern           A pattern to match in genome. Default is "CG".  
 stranded          TRUE or FALSE indicating whether to return matches on both strands or else  
                   just the "+" strand. Strand will be set to "\*" if FALSE. Default is TRUE.  
 standard\_seqs\_only   TRUE or FALSE indicating whether to only return ranges on standard sequences  
                   (those without "\_" in their names). Default is FALSE.

**Value**

A GRanges object with genomic regions matching the pattern.

**Examples**

```
# Get human CpG sites for a portion of chr1 from hg38 genome build
data(chr1_subset_hg38, package = "methodical")
chr1_subset_hg38_cpgs <- methodical::extractMethSitesFromGenome(chr1_subset_hg38)
head(chr1_subset_hg38_cpgs)

# Find CHG sites in Arabidopsis thaliana
data(chr4_subset_a_thal, package = "methodical")
chr4_subset_a_thal_chg_sites <- methodical::extractMethSitesFromGenome(
  genome = chr4_subset_a_thal, pattern = "CHG")
head(chr4_subset_a_thal_chg_sites)
```

---

 findTMRs

*Find TSS-Proximal Methylation-Controlled Regulatory Sites (TMRs)*


---

**Description**

Find TSS-Proximal Methylation-Controlled Regulatory Sites (TMRs)

**Usage**

```
findTMRs(
  correlation_list,
  offset_length = 10,
  p_adjust_method = "fdr",
  p_value_threshold = 0.05,
  smoothing_factor = 0.75,
  min_gapwidth = 150,
  min_meth_sites = 5,
  BPPARAM = BiocParallel::SerialParam()
)
```

**Arguments**

**correlation\_list** A list of data.frames with correlation values between methylation sites and a transcript as returned by calculateMethSiteTranscriptCors.

**offset\_length** Number of methylation sites added upstream and downstream of a central methylation site to form a window, resulting in a window size of  $2 * \text{offset\_length} + 1$ . Default value is 10.

**p\_adjust\_method** The method to use for p-value adjustment. Should be one of the methods in p.adjust.methods. Default is "fdr".

**p\_value\_threshold** The p\_value cutoff to use (after correcting p-values with p\_adjust\_method). Default value is 0.05.

smoothing_factor	Smoothing factor for exponential moving average. Should be a value between 0 and 1 with higher values resulting in a greater degree of smoothing. Default is 0.75.
min_gapwidth	Merge TMRs with the same direction separated by less than this number of base pairs. Default value is 150.
min_meth_sites	Minimum number of methylation sites that TMRs can contain. Default value is 5.
BPPARAM	A BiocParallelParam object for parallel processing. Defaults to <code>BiocParallel::SerialParam()</code> .

**Value**

A GRanges object with the location of TMRs.

---

hg19_chr18_cpgs	<i>hg19_chr18_cpgs</i>
-----------------	------------------------

---

**Description**

A GRanges object with CpG sites on chr18 for hg19

**Usage**

```
hg19_chr18_cpgs
```

**Format**

A GRanges object.

---

hg38_cpg_islands	<i>hg38_cpg_islands</i>
------------------	-------------------------

---

**Description**

A GRanges with CpG islands, shelves and shores for hg38

**Usage**

```
hg38_cpg_islands
```

**Format**

A GRanges object.

---

```
infinium_450k_probe_granges_hg19
      infinium_450k_probe_granges_hg19
```

---

**Description**

The hg19 genomic coordinates for methylation sites analysed by the Infinium HumanMethylation450K array.

**Usage**

```
infinium_450k_probe_granges_hg19
```

**Format**

GRanges object with 482,421 ranges and one metadata column name giving the name of the associated probe.

**Source**

Derived from the manifest file downloaded from [https://webdata.illumina.com/downloads/productfiles/humanmethylation2.csv?\\_gl<-1ocsx4f\\_gaMTk1Nzc4MDkwMy4xNjg3ODcxNjg0\\_ga\\_VVVPY8BDYL\\*MTY4Nzg3MTY4My4xLjEuMTY](https://webdata.illumina.com/downloads/productfiles/humanmethylation2.csv?_gl<-1ocsx4f_gaMTk1Nzc4MDkwMy4xNjg3ODcxNjg0_ga_VVVPY8BDYL*MTY4Nzg3MTY4My4xLjEuMTY)

---

```
liftoverMethRSE      Liftover rowRanges of a RangedSummarizedExperiment for methylation data from one genome build to another
```

---

**Description**

Removes methylation sites which cannot be mapped to the target genome build and those which result in many-to-one mappings. Also removes one-to-many mappings by default and can remove sites which do not map to allowed regions in the target genome e.g. CpG sites.

**Usage**

```
liftoverMethRSE(
  meth_rse,
  chain,
  remove_one_to_many_mapping = TRUE,
  permitted_target_regions = NULL,
  seqlevels = NULL
)
```

**Arguments**

meth_rse	A RangedSummarizedExperiment for methylation data
chain	A "Chain" object to be used with rtracklayer::liftOver
remove_one_to_many_mapping	TRUE or FALSE indicating whether to remove regions in the source genome which map to multiple regions in the target genome. Default is TRUE.
permitted_target_regions	An optional GRanges object used to filter the rowRanges by overlaps after liftover, for example CpG sites from the target genome. Any regions which do not overlap permitted_target_regions will be removed. GRangesList to GRanges if all remaining source regions can be uniquely mapped to the target genome.
seqlevels	An optional character vector giving the order to use for seqlevels of the rowRanges of the returned RangedSummarizedExperiment.

**Value**

A RangedSummarizedExperiment with rowRanges lifted over to the genome build indicated by chain.

**Examples**

```
# Load sample RangedSummarizedExperiment with CpG methylation data
data(tubb6_meth_rse, package = "methodical")
tubb6_meth_rse <- eval(tubb6_meth_rse)

# Get CpG sites for hg19
data(hg19_chr18_cpgs, package = "methodical")

# Get liftover chain for mapping hg38 to hg19
library(AnnotationHub)
ah <- AnnotationHub()
chain <- ah[["AH14108"]]

# Liftover tubb6_meth_rse from hg38 to hg19, keeping only sites that were mapped to CpG sites in hg19
tubb6_meth_rse_hg19 <- methodical::liftoverMethRSE(tubb6_meth_rse, chain = chain,
  permitted_target_regions = hg19_chr18_cpgs)
```

---

makeMethRSEFromInputFiles

*Create a HDF5-backed RangedSummarizedExperiment for methylation values in meth\_files*

---

**Description**

Create a HDF5-backed RangedSummarizedExperiment for methylation values in meth\_files

**Usage**

```

makeMethRSEFromInputFiles(
  meth_files,
  seqnames_col,
  start_col,
  total_reads_col = NULL,
  meth_reads_col = NULL,
  unmeth_reads_col = NULL,
  meth_fraction_col = NULL,
  zero_based,
  meth_sites,
  sequence_context = "CG",
  collapse_strands = TRUE,
  decimal_places = NA,
  sample_metadata = NULL,
  hdf5_dir,
  overwrite = FALSE,
  chunkdim = NULL,
  temporary_dir = NULL,
  BPPARAM = BiocParallel::SerialParam(),
  ...
)

```

**Arguments**

<code>meth_files</code>	A vector of paths to input methylation files. All sites in each file are assumed to be for the same sequence context e.g. CG or CHG. Automatically detects if <code>meth_files</code> contain a header if every field in the first line is a character.
<code>seqnames_col</code>	The column number in <code>meth_files</code> which corresponds to the sequence names.
<code>start_col</code>	The column number in <code>meth_files</code> which corresponds to the genomic start coordinate.
<code>total_reads_col</code>	The column number in <code>meth_files</code> which corresponds to the total number of reads for the position.
<code>meth_reads_col</code>	The column number in <code>meth_files</code> which corresponds to the number of methylated reads for the position.
<code>unmeth_reads_col</code>	The column number in <code>meth_files</code> which corresponds to the number of unmethylated reads for the position.
<code>meth_fraction_col</code>	The column number in <code>meth_files</code> which corresponds to the fraction of reads that support methylation at the position. Will be converted to a proportion if it appears to be a fraction.
<code>zero_based</code>	TRUE or FALSE indicating if files are zero-based.
<code>meth_sites</code>	A GRanges object with non-overlapping locations of methylation sites of interest e.g. CpG sites. Any methylation sites in <code>meth_files</code> that are not in <code>meth_sites</code> are ignored.
<code>sequence_context</code>	A single character string or DNASTring with the sequence context of the methylation sites e.g. CG or CHG. If a character, must be coercible to a DNASTring. Default is "CG".

collapse_strands	TRUE or FALSE indicating whether or not to collapse data on + and - strands. Only makes sense for symmetrically methylated contexts e.g. CG or CHG and meth_sites should include ranges on both the + and - strands if TRUE.
decimal_places	Optional integer indicating the number of decimal places to round beta values to. Default is not to round.
sample_metadata	Sample metadata to be used as colData for the RangedSummarizedExperiment.
hdf5_dir	Directory to save HDF5 file. Is created if it doesn't exist. HDF5 file is called assays.h5.
overwrite	TRUE or FALSE indicating whether to allow overwriting if hdf5_dir already exists. Default is FALSE.
chunkdim	The dimensions of the chunks for the HDF5 file. Should be a vector of length 2 giving the number of rows and then the number of columns in each chunk. Uses HDF5Array::getHDF5DumpChunkDim(length(meth_sites), length(meth_files)) by default.
temporary_dir	Name to give temporary directory created to store intermediate files. A directory with this name cannot already exist. Default is to create a subdirectory named "temporary_meth_chunks_" inside the directory given by tempdir(). Will be deleted after completion.
BPPARAM	A BiocParallelParam object for parallel processing. Defaults to BiocParallel::SerialParam().
...	Additional arguments to be passed to HDF5Array::HDF5RealizationSink() for controlling the physical properties of the created HDF5 file, such as compression level. Uses the defaults for any properties that are not specified.

## Value

A RangedSummarizedExperiment with two assays for all methylation sites in meth\_sites, beta with the proportion of methylated reads and Cov with the total number of reads for each site. methylation sites will be in the same order as sort(meth\_sites).

## Examples

```
# Load CpGs from subset of chromosome 11 as a GRanges object
data("chr11_subset_hg38_cpgs", package = "methodical")

# Get paths to meth_files
meth_files <- list.files(path = system.file('extdata', package = 'methodical'),
  pattern = ".CX_report.txt.gz", full.names = TRUE)

# Create sample metadata
sample_metadata <- data.frame(
  sample_type = ifelse(grepl("N", basename(meth_files)), "Normal", "Tumour"),
  row.names = gsub("_.*", "", basename(meth_files))
)

# Create a HDF5-backed RangedSummarizedExperiment from meth_files
meth_rse <- makeMethRSEFromInputFiles(meth_files = meth_files,
  seqnames_col = 1, start_col = 2, meth_reads_col = 4, unmeth_reads_col = 5,
  zero_based = FALSE, meth_sites = chr11_subset_hg38_cpgs, sample_metadata = sample_metadata,
  hdf5_dir = paste0(tempdir(), "/test_hdf5_1"))

# Show beta values and coverage
```

```
assay(meth_rse, "beta")
assay(meth_rse, "Cov")
```

---

maskRangesInRSE	<i>Mask regions in a ranged summarized experiment</i>
-----------------	---

---

### Description

Mask regions in a ranged summarized experiment

### Usage

```
maskRangesInRSE(rse, mask_ranges, assay_number = 1)
```

### Arguments

rse	A RangedSummarizedExperiment.
mask_ranges	Either a GRanges with regions to be masked in all samples (e.g. repetitive sequences) or a GRangesList object with different regions to mask in each sample (e.g. mutations). If using a GRangesList object, names of the list elements should be the names of samples in rse.
assay_number	Assay to perform masking. Default is first assay

### Value

A RangedSummarizedExperiment with the regions present in mask\_ranges masked

### Examples

```
# Load sample RangedSummarizedExperiment with CpG methylation data
data(tubb6_meth_rse, package = "methodical")
tubb6_meth_rse <- eval(tubb6_meth_rse)

# Create a sample GRanges object to use to mask tubb6_meth_rse
mask_ranges <- GRanges("chr18:12305000-12310000")

# Mask regions in tubb6_meth_rse
tubb6_meth_rse_masked <- methodical::maskRangesInRSE(tubb6_meth_rse, mask_ranges)

# Count the number of NA values before and after masking
sum(is.na(assay(tubb6_meth_rse)))
sum(is.na(assay(tubb6_meth_rse_masked)))
```

---

methrixToRSE	<i>Convert a Methrix object into a RangedSummarizedExperiment</i>
--------------	---

---

**Description**

Convert a Methrix object into a RangedSummarizedExperiment

**Usage**

```
methrixToRSE(methrix, assays = c("beta", "cov"))
```

**Arguments**

methrix	A methrix object
assays	A vector indicating the names of assays in methrix used to create a RangedSummarizedExperiment. Can be one or both of "beta" and "cov". Default is both "beta" and "cov" assays.

**Value**

A RangedSummarizedExperiment

**Examples**

```
# Load a sample methrix object
data("methrix_data", package = "methrix")

# Convert methrix to a RangedSummarizedExperiment with one assay for the methylation beta values
meth_rse <- methodical::methrixToRSE(methrix_data, assays = "beta")
print(meth_rse)
```

---

plotMethodicalScores	<i>Create plot of Methodical score values for methylation sites around a TSS</i>
----------------------	--

---

**Description**

Create plot of Methodical score values for methylation sites around a TSS

**Usage**

```
plotMethodicalScores(
  genomic_region_values,
  reference_tss = NULL,
  p_value_threshold = 0.005,
  smooth_scores = TRUE,
  offset_length = 10,
  smoothing_factor = 0.75,
  smoothed_curve_colour = "black",
  linewidth = 1,
```

```

curve_alpha = 0.75,
title = NULL,
xlabel = "Genomic Position",
low_colour = "#7B5C90",
high_colour = "#BFAB25"
)

```

## Arguments

**genomic\_region\_values** A data.frame with correlation values for methylation sites. There should be one column called "cor". and another called "p\_val" which are used to calculate the Methodical score. row.names should be the names of methylation sites and all methylation sites must be located on the same sequence.

**reference\_tss** An optional GRanges object with a single range. If provided, the x-axis will show the distance of methylation sites to the start of this region with methylation sites upstream. relative to the reference\_tss shown first. If not, the x-axis will show the start site coordinate of the methylation site.

**p\_value\_threshold** The p-value threshold used to identify TMRs. Default value is 0.005. Set to NULL to turn off significance thresholds.

**smooth\_scores** TRUE or FALSE indicating whether to display a curve of smoothed Methodical scores on top of the plot. Default is TRUE.

**offset\_length** Offset length to be supplied to calculateSmoothedMethodicalScores. Default is 10.

**smoothing\_factor** Smoothing factor to be provided to calculateSmoothedMethodicalScores. Default is 0.75.

**smoothed\_curve\_colour** Colour of the smoothed curve. Default is "black".

**linewidth** Line width of the smoothed curve. Default value is 1.

**curve\_alpha** Alpha value for the curve. Default value is 0.75.

**title** Title of the plot. Default is no title.

**xlabel** Label for the X axis in the plot. Default is "Genomic Position".

**low\_colour** Colour to use for low values. Default value is "#7B5C90".

**high\_colour** Colour to use for high values. Default value is "#BFAB25".

## Value

A ggplot object

## Examples

```

# Load methylation-transcript correlation results for TUBB6 gene
data("tubb6_cpg_meth_transcript_cors", package = "methodical")

# Calculate and plot Methodical scores from correlation values
methodical::plotMethodicalScores(tubb6_cpg_meth_transcript_cors, reference_tss = attributes(tubb6_cpg_meth_t

```

---

plotMethSiteCorCoefs *Plot the correlation coefficients for methylation sites within a region and an associated feature of interest*

---

### Description

Plot the correlation coefficients for methylation sites within a region and an associated feature of interest

### Usage

```
plotMethSiteCorCoefs(
  meth_site_cor_values,
  reference_tss = FALSE,
  title = NULL,
  xlabel = NULL,
  ylabel = "Correlation Coefficient",
  value_colours = c("#7B5C90", "#bfab25"),
  reverse_x_axis = FALSE
)
```

### Arguments

**meth\_site\_cor\_values** A data.frame with correlation values associated with methylation sites, such as returned by `calculateMethSiteTranscriptCors`. There should be one column called `meth_site` giving the coordinates of methylation sites in character format and another column called `cor` giving the correlation between the methylation values of the methylation sites and a feature of interest. All methylation sites must be located on the same sequence.

**reference\_tss** TRUE or FALSE indicating whether to show distances on the X-axis relative to the TSS stored as an attribute `tss_range` of `meth_site_cor_values`. Alternatively, can provide a `GRanges` object with a single range for such a TSS site. In either case, will show the distance of methylation sites to the start of this region with methylation sites upstream relative to the `reference_tss` shown first. If FALSE (the default), the x-axis will instead show the start site coordinate of the methylation site.

**title** Title of the plot. Default is no title.

**xlabel** Label for the X axis in the plot. Defaults to "Distance to TSS" if `reference_tss` is used or "seqname position" where `seqname` is the name of the relevant sequence.

**ylabel** Label for the Y axis in the plot. Default is "Correlation Coefficient".

**value\_colours** A vector with two colours to use, the first for low values and the second for high values. Defaults are `c("#7B5C90", "#bfab25")`.

**reverse\_x\_axis** TRUE or FALSE indicating whether x-axis should be reversed, for example if plotting a region on the reverse strand so that left side of plot corresponds to upstream.

### Value

A ggplot object

**Examples**

```
# Load methylation-transcript correlation results for TUBB6 gene
data("tubb6_cpg_meth_transcript_cors", package = "methodical")

# Plot methylation-transcript correlation values around TUBB6 TSS
methodical::plotMethSiteCorCoefs(tubb6_cpg_meth_transcript_cors, ylabel = "Spearman Correlation")

# Create same plot but showing the distance to the TUBB6 TSS on the x-axis
methodical::plotMethSiteCorCoefs(tubb6_cpg_meth_transcript_cors,
  ylabel = "Spearman Correlation", reference_tss = attributes(tubb6_cpg_meth_transcript_cors)$tss_range)
```

---

plotRegionValues	<i>Create a scatter plot with smoothed curve for values along adjacent loci in a genomic region</i>
------------------	---

---

**Description**

Create a scatter plot with smoothed curve for values along adjacent loci in a genomic region

**Usage**

```
plotRegionValues(
  genomic_region_values,
  sample_name = NULL,
  reference_tss = FALSE,
  geom_point_params = list(),
  geom_smooth_params = list(),
  title = NULL,
  xlabel = NULL,
  ylabel = "Genomic Region Value",
  value_colours = c("#53868B", "#CD2626"),
  reverse_x_axis = FALSE
)
```

**Arguments**

genomic_region_values	A data.frame with values associated with genomic regions. Row names must be the coordinates of genomic regions in character format (e.g chr1:1000-2000) and all regions must be located on the same sequence. The position of the first base in each region is used as the x-axis coordinate for the plot.
sample_name	Name of column in genomic_region_values to plot. Defaults to first column if none provided.
reference_tss	TRUE or FALSE indicating whether to show distances on the X-axis relative to the TSS stored as an attribute tss_range of genomic_region_values. Alternatively, can provide a GRanges object with a single range for such a TSS site. In either case, will show the distance of genomic regions to the start of this region with genomic regions upstream relative to the reference_tss shown first. If FALSE (the default), the x-axis will instead show the start site coordinate of the genomic region.

<code>geom_point_params</code>	An optional list to explicitly set values of parameters to use with <code>geom_point()</code> . Use <code>list(alpha = 0)</code> to make points invisible.
<code>geom_smooth_params</code>	An optional list to explicitly set values of parameters to use with <code>geom_smooth()</code> . Use <code>list(alpha = 0)</code> to make line invisible.
<code>title</code>	Title of the plot. Default is no title.
<code>xlabel</code>	Label for the X axis in the plot. Defaults to "Distance to TSS" if <code>reference_tss</code> is used or "seqname position" where <code>seqname</code> is the name of the relevant sequence.
<code>ylabel</code>	Label for the Y axis in the plot. Default is "Genomic Region Value".
<code>value_colours</code>	A vector with two colours to use, the first for low values and the second for high values. Defaults are <code>c("#53868B", "#CD2626")</code> .
<code>reverse_x_axis</code>	TRUE or FALSE indicating whether x-axis should be reversed, for example if plotting a region on the reverse strand so that left side of plot corresponds to upstream.

**Value**

A ggplot object

**Examples**

```
# Load methylation-values around the TUBB6 TSS
data("tubb6_meth_rse", package = "methodical")
tubb6_meth_rse <- eval(tubb6_meth_rse)

# Extract methylation values from tubb6_meth_rse
tubb6_methylation_values = methodical::extractGRangesMethSiteValues(meth_rse = tubb6_meth_rse)

# Plot methylation values around TUBB6 TSS
methodical::plotRegionValues(tubb6_methylation_values, sample_name = "N1", ylabel = "Methylation Value")

# Create same plot but showing the distance to the TUBB6 TSS on the x-axis
data("tubb6_tss", package = "methodical")
methodical::plotRegionValues(tubb6_methylation_values, sample_name = "N1",
  reference_tss = tubb6_tss, ylabel = "Methylation Value")
```

---

plotTMRs

*Add TMRs to a methylation site value plot*

---

**Description**

Add TMRs to a methylation site value plot

**Usage**

```
plotTMRs(
  meth_site_plot,
  tmrs_gr,
  reference_tss = NULL,
```

```

transcript_id = NULL,
tmr_colours = c("#A28CB1", "#D2C465"),
linewidth = 5
)

```

### Arguments

**meth\_site\_plot** A plot of Value around a TSS.

**tmsr\_gr** A GRanges object giving the position of TMRs.

**reference\_tss** An optional GRanges object with a single range. If provided, the x-axis will show the distance of methylation sites to the start of this region with methylation sites upstream relative to the reference\_tss shown first. If not, the x-axis will show the start site coordinate of the methylation site.

**transcript\_id** An optional transcript ID. If provided, will attempt to filter tmsr\_gr and reference\_tss using a metadata column called transcript\_id with a value identical to the provided one.

**tmr\_colours** A vector with colours to use for negative and positive TMRs. Defaults to "#7B5C90" for negative and "#BFAB25" for positive TMRs.

**linewidth** A numeric value to be provided as linewidth for geom\_segment().

### Value

A ggplot object

### Examples

```

# Load methylation-transcript correlation results for TUBB6 gene
data("tubb6_cpg_meth_transcript_cors", package = "methodical")

# Plot methylation-transcript correlation values around TUBB6 TSS
tubb6_correlation_plot <- methodical::plotMethSiteCorCoefs(tubb6_cpg_meth_transcript_cors, ylabel = "Spearman")

# Find TMRs for TUBB6
tubb6_tmsr <- findTMRs(correlation_list = list(ENST00000591909 = tubb6_cpg_meth_transcript_cors))

# Plot TMRs on top of tubb6_correlation_plot
methodical::plotTMRs(tubb6_correlation_plot, tmsr_gr = tubb6_tmsr)

```

---

`rangesRelativeToTSS` *Find locations of genomic regions relative to transcription start sites.*

---

### Description

Find locations of genomic regions relative to transcription start sites.

### Usage

```
rangesRelativeToTSS(genomic_regions, tss_gr)
```

**Arguments**

`genomic_regions` A GRanges object.

`tss_gr` A GRanges object with transcription start sites. Each range should have width 1. Upstream and downstream are relative to strand of `tss_gr`.

**Value**

A GRanges object where all regions have "relative" as the sequence names and ranges are the location of TMRs relative to the TSS.

**Examples**

```
# Create query and subject GRanges
genomic_regions <- GenomicRanges::GRanges(c("chr1:100-1000:+", "chr1:2000-3000:-"))
tss_gr <- GenomicRanges::GRanges(c("chr1:1500:+", "chr1:4000:-"))

# Calculate distances between query and subject
methodical::rangesRelativeToTSS(genomic_regions, tss_gr)
```

---

<code>rapidCorTest</code>	<i>Rapidly calculate the correlation and the significance of pairs of columns from two data.frames</i>
---------------------------	--

---

**Description**

Rapidly calculate the correlation and the significance of pairs of columns from two data.frames

**Usage**

```
rapidCorTest(
  table1,
  table2,
  cor_method = "pearson",
  table1_name = "table1",
  table2_name = "table2",
  p_adjust_method = "BH",
  n_covariates = 0,
  min_number_complete_pairs = 30
)
```

**Arguments**

`table1` A data.frame

`table2` A data.frame

`cor_method` A character string indicating which correlation coefficient is to be computed. One of either "pearson" or "spearman" or their abbreviations.

`table1_name` Name to give the column giving the name of features in table1. Default is "table1".

table2\_name      Name to give the column giving the name of features in table2. Default is "table2".

p\_adjust\_method      Method used to adjust p-values. Same as the methods from p.adjust.methods. Default is Benjamini-Hochberg. Setting to "none" will result in no adjusted p-values being calculated.

n\_covariates      Number of covariates if calculating partial correlations. Defaults to 0.

min\_number\_complete\_pairs      The minimum number of complete pairs required to return a p-value for a correlation. Correlations with less than this number are given a p-value of NaN. Default value is 30.

### Value

A data.frame with the correlation and its significance for all pairs consisting of a variable from table1 and a variable from table2.

### Examples

```
# Divide mtcars into two tables
table1 <- mtcars[, 1:5]
table2 <- mtcars[, 6:11]

# Calculate correlation between table1 and table2
cor_results <- methodical::rapidCorTest(table1, table2, cor_method = "spearman",
  table1_name = "feature1", table2_name = "feature2")
head(cor_results)
```

---

sampleMethSites      *Randomly sample sites from a methylation RSE.*

---

### Description

Randomly sample sites from a methylation RSE.

### Usage

```
sampleMethSites(
  meth_rse,
  n_sites = 1000,
  seqnames_filter = NULL,
  genomic_ranges_filter = NULL,
  invert_granges_filter = FALSE,
  samples_subset = NULL
)
```

**Arguments**

meth_rse	A RangedSummarizedExperiment for methylation data.
n_sites	Number of sites to randomly sample. Default is 1000. Will give an error if there are less than this number of sites available to sample after applying any of the optional filters.
seqnames_filter	An optional character vector giving the names of sequences to filter meth_rse for.
genomic_ranges_filter	An optional GRanges object used to first subset meth_rse. Sites will then be chosen randomly from those overlapping these ranges.
invert_granges_filter	TRUE or FALSE indicating whether to invert the genomic_ranges_filter so as to exclude sites overlapping these regions. Default value is FALSE.
samples_subset	Optional sample names used to subset meth_rse.

**Value**

A RangedSummarizedExperiment with the specified number of randomly sampled sites after applying the different filters.

**Examples**

```
# Load sample RangedSummarizedExperiment with CpG methylation data
data(tubb6_meth_rse, package = "methodical")
tubb6_meth_rse <- eval(tubb6_meth_rse)

# Create a sample GRanges object to use to mask tubb6_meth_rse
mask_ranges <- GRanges("chr18:12305000-12310000")

# Get 20 random CpG sites outside mask_ranges
random_cpgs <- methodical::sampleMethSites(tubb6_meth_rse, n_sites = 20, genomic_ranges_filter = mask_ranges,
  invert_granges_filter = TRUE)

# Check that no CpGs overlap repeats
intersect(rowRanges(random_cpgs), mask_ranges)
```

---

strandedDistance	<i>Calculate distances of query GRanges upstream or downstream of subject GRanges</i>
------------------	---

---

**Description**

Upstream ranges are assigned negative distances and downstream regions positive distances and are relative to the strand of subject\_gr. Unstranded ranges are treated the same as regions on the "+" strand. If subject\_gr has a length of 1, then distances are calculated between each range in query\_gr and this range, otherwise distances are calculated in a pairwise manner between ranges in query\_gr and subject\_gr.

**Usage**

```
strandedDistance(query_gr, subject_gr)
```

**Arguments**

```
query_gr      A GRanges object.
subject_gr    A GRanges object.
```

**Value**

A numeric vector of distances.

**Examples**

```
# Create query and subject GRanges
query_gr <- GenomicRanges::GRanges(c("chr1:100-1000:+", "chr1:2000-3000:-"))
subject_gr <- GenomicRanges::GRanges(c("chr1:1500-1600:+", "chr1:4000-4500:-"))

# Calculate distances between query and subject
methodical::strandedDistance(query_gr, subject_gr)
```

---

```
summarizeRegionMethylation
```

*Summarize methylation of genomic regions within samples*

---

**Description**

Summarize methylation of genomic regions within samples

**Usage**

```
summarizeRegionMethylation(
  meth_rse,
  assay = 1,
  genomic_regions,
  genomic_region_names = NULL,
  col_summary_function = "colMeans2",
  keep_metadata_cols = FALSE,
  max_sites_per_chunk = floor(62500000/ncol(meth_rse)),
  na.rm = TRUE,
  BPPARAM = BiocParallel::SerialParam(),
  ...
)
```

**Arguments**

```
meth_rse      A RangedSummarizedExperiment with methylation values.
assay         The assay from meth_rse to extract values from. Should be either an index or
              the name of an assay. Default is the first assay.
genomic_regions GRanges object with regions to summarize methylation values for.
```

genomic_region_names	A character vector of unique names to assign genomic_regions in the output table. Defaults to names(genomic_regions) if present or otherwise converts regions to character strings (e.g. "chr:1000-2000") to use as names.
col_summary_function	A function that summarizes column values. Should be the name of one of the column summary functions from MatrixGenerics. Default is "colMeans2".
keep_metadata_cols	TRUE or FALSE indicating whether to add the metadata columns of genomic_regions to the output. Default is FALSE.
max_sites_per_chunk	The approximate maximum number of methylation sites to try to load into memory at once. The actual number loaded may vary depending on the number of methylation sites overlapping each region, but so long as the size of any individual regions is not enormous ( $\geq$ several MB), it should vary only very slightly. Some experimentation may be needed to choose an optimal value as low values will result in increased running time, while high values will result in a large memory footprint without much improvement in running time. Default is $\text{floor}(62500000/\text{ncol}(\text{meth\_rse}))$ , resulting in each chunk requiring approximately 500 MB of RAM.
na.rm	TRUE or FALSE indicating whether to remove NA values when calculating summaries. Default value is TRUE.
BPPARAM	A BiocParallelParam object. Defaults to BiocParallel::SerialParam().
...	Additional arguments to be passed to col_summary_function.

**Value**

A data.table with the summary of methylation of each region in genomic\_regions for each sample.

**Examples**

```
# Load sample RangedSummarizedExperiment with CpG methylation data
data(tubb6_meth_rse, package = "methodical")
tubb6_meth_rse <- eval(tubb6_meth_rse)

# Create a sample GRanges
test_gr <- GRanges(c("chr18:12303400-12303500", "chr18:12303600-12303750", "chr18:12304000-12306000"))
names(test_gr) <- paste("region", 1:3, sep = "_")

# Calculate mean methylation values for regions in test_gr
test_gr_methylation <- methodical::summarizeRegionMethylation(tubb6_meth_rse, genomic_regions = test_gr,
  genomic_region_names = names(test_gr))
```

---

tubb6\_correlation\_plot

*tubb6\_correlation\_plot*


---

**Description**

A plot of the correlation values between methylation-transcription correlations for CpG sites around the TUBB6 TSS.

**Usage**

```
tubb6_correlation_plot
```

**Format**

A ggplot object.

---

```
tubb6_cpg_meth_transcript_cors
      tubb6_cpg_meth_transcript_cors
```

---

**Description**

A data.frame with the methylation-transcription correlation results for CpGs around the TUBB6 TSS.

A data.frame with the correlation results for CpG sites within +/- 5 KB of the TUBB6 (ENST00000591909) TSS.

**Usage**

```
tubb6_cpg_meth_transcript_cors
```

```
tubb6_cpg_meth_transcript_cors
```

**Format**

A ggplot object.

A data.frame with 5 columns giving the name of the CpG site (meth\_site), name of the transcript associated with the TSS, Spearman correlation value between the methylation of the CpG site and expression of the transcript, p-value associated with the correlations and distance from the CpG site to the TSS.

---

```
tubb6_meth_rse      tubb6_meth_rse
```

---

**Description**

A RangedSummarizedExperiment with methylation data for TUBB6.

**Usage**

```
tubb6_meth_rse
```

**Format**

A call to create a RangedSummarizedExperiment with methylation data for 355 CpG sites within +/- 5,000 base pairs of the TUBB6 TSS in 126 normal prostate samples. Should be evaluated after loading using `tubb6_meth_rse <- tubb6_meth_rse <- eval(tubb6_meth_rse)` to restore the RangedSummarizedExperiment.

**Source**

WGBS data from 'Li, Jing, et al. "A genomic and epigenomic atlas of prostate cancer in Asian populations." Nature 580.7801 (2020): 93-99.'

---

`tubb6_tmrs``tubb6_tmrs`

---

**Description**

TMRs identified for TUBB6

**Usage**`tubb6_tmrs`**Format**

A GRanges object with two ranges.

---

`tubb6_transcript_counts``tubb6_transcript_counts`

---

**Description**

Transcript counts for TUBB6 in normal prostate samples.

**Usage**`tubb6_transcript_counts`**Format**

A data.frame with normalized transcript counts for TUBB6 in 126 normal prostate samples.

**Source**

RNA-seq data from 'Li, Jing, et al. "A genomic and epigenomic atlas of prostate cancer in Asian populations." Nature 580.7801 (2020): 93-99.'

---

tubb6\_tss

*tubb6\_tss*

---

**Description**

The location of the TSS for TUBB6.

**Usage**

tubb6\_tss

**Format**

GRanges object with 1 ranges for the TUBB6 TSS.

**Source**

The TSS of the ENST00000591909 transcript.

# Index

## \* datasets

- chr11\_subset\_hg38\_cpgs, [19](#)
- chr1\_subset\_hg38, [19](#)
- chr4\_subset\_a\_thal, [20](#)
- hg19\_chr18\_cpgs, [25](#)
- hg38\_cpg\_islands, [25](#)
- infinium\_450k\_probe\_granges\_hg19, [26](#)
- tubb6\_correlation\_plot, [41](#)
- tubb6\_cpg\_meth\_transcript\_cors, [42](#)
- tubb6\_meth\_rse, [42](#)
- tubb6\_tmrs, [43](#)
- tubb6\_transcript\_counts, [43](#)
- tubb6\_tss, [44](#)
- .check\_input\_files, [4](#)
- .chunk\_regions, [4](#)
- .collapse\_strands, [5](#)
- .create\_meth\_rse\_from\_hdf5, [5](#)
- .find\_tmrs\_single, [6](#)
- .make\_meth\_rse\_setup, [7](#)
- .set\_meth\_df\_columns, [8](#)
- .split\_meth\_file, [8](#)
- .split\_meth\_files\_into\_chunks, [9](#)
- .summarize\_chunk\_methylation, [10](#)
- .test\_tmrs, [10](#)
- .tss\_correlations, [11](#)
- .tss\_iterator, [11](#)
- .write\_chunks\_to\_hdf5, [12](#)
- annotatePlot, [13](#)
- calculateMethSiteTranscriptCors, [14](#)
- calculateRegionMethylationTranscriptCors, [16](#)
- calculateSmoothedMethodicalScores, [18](#)
- chr11\_subset\_hg38\_cpgs, [19](#)
- chr1\_subset\_hg38, [19](#)
- chr4\_subset\_a\_thal, [20](#)
- convert\_rse\_to\_bsseq, [20](#)
- correct\_correlation\_pvalues, [21](#)
- expand\_granges, [21](#)
- export\_bedGraph\_from\_rse, [22](#)
- extractGRangesMethSiteValues, [22](#)
- extractMethSitesFromGenome, [23](#)
- findTMRs, [24](#)
- hg19\_chr18\_cpgs, [25](#)
- hg38\_cpg\_islands, [25](#)
- infinium\_450k\_probe\_granges\_hg19, [26](#)
- liftoverMethRSE, [26](#)
- makeMethRSEFromInputFiles, [27](#)
- maskRangesInRSE, [30](#)
- methodical (methodical-package), [3](#)
- methodical-package, [3](#)
- methrixToRSE, [31](#)
- plotMethodicalScores, [31](#)
- plotMethSiteCorCoefs, [33](#)
- plotRegionValues, [34](#)
- plotTMRs, [35](#)
- rangesRelativeToTSS, [36](#)
- rapidCorTest, [37](#)
- sampleMethSites, [38](#)
- strandedDistance, [39](#)
- summarizeRegionMethylation, [40](#)
- tubb6\_correlation\_plot, [41](#)
- tubb6\_cpg\_meth\_transcript\_cors, [42](#)
- tubb6\_meth\_rse, [42](#)
- tubb6\_tmrs, [43](#)
- tubb6\_transcript\_counts, [43](#)
- tubb6\_tss, [44](#)