Package 'MethylSeqData'

November 10, 2026

Title Collection of Public DNA Methylation Sequencing Datasets

Version 1.22.0

Date 2020-10-25

Description Base-level (i.e. cytosine-level) counts for a collection of public bisulfite-seq datasets (e.g., WGBS and RRBS), provided as SummarizedExperiment objects with sample- and base-level metadata.

License CC0

NeedsCompilation no

Depends SummarizedExperiment

Imports ExperimentHub, HDF5Array, rhdf5, GenomeInfoDb, S4Vectors, GenomicRanges, stats, IRanges, utils

Suggests BiocFileCache, BiocStyle, data.table, knitr, rmarkdown

VignetteBuilder knitr

Encoding UTF-8

biocViews ExperimentHub, ExperimentData, SequencingData, Homo_sapiens_Data, Mus_musculus_Data, MethylSeqData

BuildResaveData no

RoxygenNote 7.1.1

git_url https://git.bioconductor.org/packages/MethylSeqData

git_branch RELEASE_3_23

git_last_commit 4ac30fc

git_last_commit_date 2026-04-28

Repository Bioconductor 3.23

Date/Publication 2026-11-10

Author Peter Hickey [aut, cre]

Maintainer Peter Hickey <peter.hickey@gmail.com>

Contents

ChenMammaryData 2
listDatasets 3
RizzardiHickeyBrain 4

Index **6**

`ChenMammaryData` *Obtain the Chen mammary data*

Description

Obtain the mouse mammary reduced representation bisulfite sequencing (RRBS) data from Chen et al. (2017).

Usage

```
ChenMammaryData(HDF5Array = TRUE)
```

Arguments

`HDF5Array`	Logical scalar indicating whether the assays should be returned as HDF5Array objects or as ordinary matrix objects.

Details

The data consists of RRBS methylation profiles of epithelial populations from the mouse mammary gland, with $n = 2$ biological replicates for each of three cell populations. The dataset contains unstranded CpG loci and the counts have not been aggregated by strand.

Column metadata include:

- `Sample` The sample ID
- `Population` The epithelial population ID
- `Description` The description of the epithelial population

If `HDF5Array = TRUE` then the each assay is returned as a HDF5Array object. Each HDF5Array occupies approximately 2 Kib in memory. If `HDF5Array = FALSE` then the each assay is returned as an ordinary matrix objects. Each matrix occupies approximately 81 MiB in memory.

All data are downloaded from ExperimentHub and cached for local re-use. Specific resources can be retrieved by searching for `MethylSeqdata/chen_mammary`.

Value

A SummarizedExperiment object with assays containing a matrix of methylated read counts (`M`) and a matrix of unmethylated read counts (`U`).

Author(s)

Peter Hickey

References

Chen, et al. (2017). Differential methylation analysis of reduced representation bisulfite sequencing experiments using edgeR. *F1000Research* 6, 23-32. https://doi.org/10.12688/f1000research.13196.2

Examples

```
se <- ChenMammaryData()
```

`listDatasets` *List all available datasets*

Description

Summary information for all available datasets in the **MethylSeqData** package.

Usage

```
listDatasets()
```

Details

A study may contribute multiple datasets if they cannot be reasonably combined (e.g., different sets methylation loci or different species). The reported number of samples refers only to the dataset as it is stored in **MethylSeqData**; this may be different to the number of samples used by the authors in their analysis, e.g., due to filtering.

Value

A DataFrame where each row corresponds to a dataset, containing the fields:

- `Reference`, a Markdown-formatted citation to `scripts/ref.bib` in the **MethylSeqData** installation directory.
- `Taxonomy`, an identifier for the organism.
- `Part`, the part of the organism being studied.
- `Number`, the total number of samples in the dataset.
- `Call`, the relevant R call required to construct the dataset.

Author(s)

Peter Hickey

Examples

```
listDatasets()
```

`RizzardiHickeyBrain` *Obtain the Rizzardi/Hickey brain data*

Description

Obtain the human brain nuclei whole genome bisulfite sequencing (WGBS) data from Rizzardi, Hickey, et al. (2019).

Usage

```
RizzardiHickeyBrain(
  context = c("CpG", "CpA", "CpT"),
  strand = c("*", "+", "-"),
  HDF5Array = TRUE
)
```

Arguments

`context`	A string specifying the context of the methylation loci to be obtained.
`strand`	A string specifying the strand of the methylation loci to be obtained.
`HDF5Array`	Logical scalar indicating whether the assays should be returned as HDF5Array objects or as ordinary matrix objects.

Details

The data consists of WGBS methylation profiles of neuronal, non-neuronal, and unsorted nuclei populations from four different regions of the human brain, with $n = 4 - 9$ biological replicates for each condition (i.e. the combination of nuclei population and brain region).

Column metadata include:

- `donor` The donor ID
- `neun` The nuclei population as determined using fluorescence-activated nuclear sorting based on the neuronal marker NeuN
  - `"pos"` Neuronal (NeuN+) nuclei
  - `"neg"` Non-neuronal (NeuN-) nuclei
  - `"unsorted"` Unsorted nuclei (i.e. including both neuronal (Neun+) and non-neuronal (NeuN-) nuclei in unknown proportions).
- `tissue` The brain region
  - `"BA9"` Brodmann area 9 (part of the frontal cortex)
  - `"BA24"` Brodmann area 9 (part of the anterior cingulate)
  - `"HC"` Hippocampus

  - "NAcc" Nucleus accumbens

The majority of donors were matched between sorted and unsorted populations, but not all.

Separate datasets are provided for five sets of methylation loci:

- CpGs (context = "CpG") aggregated across strands (and thus unstranded, strand = "*")
- CpAs (context = "CpA") on the forward strand (strand = "+")
- CpAs (context = "CpA") on the reverse strand (strand = "-")
- CpTs (context = "CpT") on the forward strand (strand = "+")
- CpTs (context = "CpT") on the reverse strand (strand = "-")

Data from the unsorted nuclei are not available for the CpA or CpT datasets.

If HDF5Array = TRUE then the each assay is returned as a HDF5Array object. Each HDF5Array occupies approximately 2 KB in memory. If HDF5Array = FALSE then the each assay is returned as an ordinary matrix objects. Each matrix occupies approximately $6 - 33$ GiB in memory (smaller for CpG, larger for CpA and CpT).

All data are downloaded from ExperimentHub and cached for local re-use. Specific resources can be retrieved by searching for MethylSeqdata/rizzardi_hickey_brain.

Value

A SummarizedExperiment object with assays containing a matrix of methylated read counts (M) and a matrix of total read counts (Cov).

Author(s)

Peter Hickey

References

Rizzardi, Hickey, et al. (2019) Neuronal brain-region-specific DNA methylation and chromatin accessibility are associated with neuropsychiatric trait heritability. *Nature neuroscience* 22(2), 307-316. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6348048/

Examples

```
se <- RizzardiHickeyBrain()
```

Index

ChenMammaryData, 2

DataFrame, *3*

HDF5Array, *2*, *4*, *5*

listDatasets, 3

matrix, *2*, *4*, *5*

RizzardiHickeyBrain, 4

SummarizedExperiment, *2*, *5*