Package ‘DMRsegaldata’

May 28, 2026

Type Package

Title Example DNAm Data for DMRsegal

Version 1.1.0

Description Data package providing example DNA methylation files used in the DMRsegal vignette and examples. Includes a sorted beta matrix as a tab-delimited, bgzip-compressed file and a matching phenotype table. The data contains 10 healthy and 10 cancer samples, and preprocessing has already been performed on the beta values.

License GPL (>= 2)

Encoding UTF-8

URL `https://github.com/CMG-UA/DMRsegal`

BugReports `https://github.com/CMG-UA/DMRsegal/issues`

biocViews ExperimentHub, ExperimentData, SNPData, SequencingData, MicroarrayData, Genome

Depends R (>= 4.5.0), ExperimentHub

LazyData yes

LazyDataCompression xz

Suggests knitr, rmarkdown, BiocStyle, ExperimentHubData

VignetteBuilder knitr

NeedsCompilation no

RoxygenNote 7.3.3

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

git_branch devel

git_last_commit ce339a5

git_last_commit_date 2026-04-28

Repository Bioconductor 3.24

Date/Publication 2026-05-28

Author Vasileios Lemonidis [aut, cre, cph] (ORCID: <`https://orcid.org/0000-0002-6446-3536`>), Center for Oncological Research, University of Antwerp [cph, fnd], Stichting Tegen Kanker [fnd]

Maintainer Vasileios Lemonidis <`vasileios.lemonidis@uantwerpen.be`>

Contents

`DMRsegaldata-package` *Example DNA Methylation Data for DMRsegal*

Description

The DMRsegaldata package provides example DNA methylation datasets for use with the DMRsegal package. It contains preprocessed Illumina HumanMethylation450K array data from 10 cancer and 10 normal/healthy samples, along with associated phenotype information, differentially methylated positions (DMPs), and array type metadata.

This data package is designed to support the DMRsegal vignette and examples, providing users with realistic methylation data to explore differential methylation analysis workflows.

Details

Package:	DMRsegaldata
Type:	Package
Version:	0.99.6
License:	GPL (>= 2)
LazyData:	yes

The package contains four main data objects:

`beta` A matrix of DNA methylation beta values (proportion of methylation) for CpG sites across 20 samples. Rows represent CpG sites and columns represent samples (10 cancer: T1-T10, 10 normal: H1-H10).

`pheno` A data frame containing phenotype information for the 20 samples, including sample group (cancer vs. normal), age, and gender.

`dmps` A data frame of differentially methylated positions identified by limma analysis comparing cancer vs. normal samples, filtered at FDR < 0.05. Includes p-values, q-values, and adjusted p-values.

`array_type` A character string indicating the Illumina array platform used ("450K").

All datasets are provided as .rda files with xz compression and are loaded automatically via LazyData.

Data Preparation

The methylation data has been preprocessed and quality controlled. The preprocessing included:

- Quality control of raw Illumina 450K array data
- Normalization of beta values
- Filtering of low-quality probes
- Calculation of differentially methylated positions using limma

The DMP analysis was performed using empirical Bayes moderated t-statistics on M-values (log2 ratio of methylated to unmethylated signal), with adjustment for age and gender covariates.

Accessing Data

Resources can be accessed through the standard ExperimentHub interface:

```
library(ExperimentHub)
eh <- ExperimentHub()

# Query for all DMRsegaldata resources
myfiles <- query(eh, "DMRsegaldata")
myfiles

# Load specific resources by name
beta <- myfiles[["EH10275"]]
pheno <- myfiles[["EH10276"]]

# Or use package helper functions
library(ExperimentHub)
eh <- ExperimentHub()

# List all resources in this package
ExperimentHub::listResources(eh, "DMRsegaldata")

# Load a specific resource
beta <- ExperimentHub::loadResource(eh, "DMRsegaldata", "beta.rda")
```

Integration with DMRsegal

This data package is designed to work seamlessly with the DMRsegal package for differential methylation region analysis. The data objects can be directly used as inputs to DMRsegal functions.

See the DMRsegal package documentation for analysis workflows using this data.

Author(s)

Vasileios Lemonidis

Maintainer: Vasileios Lemonidis <vasileios.lemonidis@uantwerpen.be>

References

- DMRsegal package: `https://github.com/CMG-UA/DMRsegal`
- Illumina HumanMethylation450K array documentation
- Bioconductor workflows for DNA methylation analysis

See Also

`beta`, `pheno`, `dmps`, `array_type`

Related packages:

- `DMRsegal` - Differential methylation region analysis
- `minfi` - Illumina methylation array analysis
- `limma` - Linear models for microarray and RNA-seq data

Examples

```
## Not run:
## Access via ExperimentHub (recommended for end users)
library(ExperimentHub)
eh <- ExperimentHub()


## List all resources in this package
myfiles <- query(eh, "DMRsegaldata")
myfiles

## Load resources using helper functions
beta <- eh[["EH10275"]]
pheno <- eh[["EH10276"]]
dmps <- eh[["EH10277"]]
array_type <- eh[["EH10278"]]

## Explore the data
dim(beta)
head(beta[, 1:5])

## Examine phenotype data
head(pheno)
table(pheno$Sample_Group)

## View DMPs
head(dmps)
cat("Number of significant DMPs:", nrow(dmps), "\n")

## Check array type
array_type

## Calculate mean methylation per sample group
cancer_samples <- rownames(pheno)[pheno$Sample_Group == "cancer"]
normal_samples <- rownames(pheno)[pheno$Sample_Group == "normal"]

mean_cancer <- mean(colMeans(beta[, cancer_samples], na.rm = TRUE))
mean_normal <- mean(colMeans(beta[, normal_samples], na.rm = TRUE))

cat("Mean methylation (cancer):", round(mean_cancer, 4), "\n")
cat("Mean methylation (normal):", round(mean_normal, 4), "\n")

## Examine a top DMP
if (nrow(dmps) > 0) {
  top_cpg <- rownames(dmps)[1]
  if (top_cpg %in% rownames(beta)) {
```

```
    boxplot(beta[top_cpg, ] ~ pheno$Sample_Group,
            main = paste("Top DMP:", top_cpg),
            xlab = "Sample Group", ylab = "Beta Value",
            col = c("lightcoral", "lightblue"))
  }
}

## End(Not run)
```

`array_type` *Methylation Array Type*

Description

A character string indicating the type of Illumina methylation array used for the data in this package.

Format

A character string with value "450K", indicating the data was generated using the Illumina HumanMethylation450 BeadChip array.

Details

This variable specifies that the methylation data in this package comes from the Illumina 450K array platform, which measures methylation at approximately 450,000 CpG sites across the human genome.

Accessing the Data

This resource can be accessed via ExperimentHub:

```
library(ExperimentHub)
eh <- ExperimentHub()

# Query and load
myfiles <- query(eh, "DMRsegaldata")
array_type <- myfiles[["EH10278"]]
```

Source

Example data for the DMRsegal package vignette and examples.

Examples

```
## Not run:
library(ExperimentHub)
eh <- ExperimentHub()
array_type <- eh[["EH10278"]]

## End(Not run)
```

`beta` *DNA Methylation Beta Values Matrix*

Description

A matrix of DNA methylation beta values for CpG sites across 20 samples (10 healthy and 10 cancer samples). The matrix contains preprocessed beta values from Illumina 450K methylation array data.

Format

A numeric matrix with CpG sites as rows and sample identifiers as columns. Beta values range from 0 to 1, representing the proportion of methylation at each CpG site.

Details

The beta values have been preprocessed and quality controlled. The data includes measurements from 10 healthy samples (H1-H10) and 10 cancer samples (T1-T10).

Data Processing: The preprocessing steps included:

- Quality control of raw Illumina 450K array data
- Normalization of beta values
- Filtering of low-quality probes
- Sample QC and filtering

Accessing the Data

This resource can be accessed via ExperimentHub:

```
library(ExperimentHub)
eh <- ExperimentHub()
beta <- eh[["EH10275"]]
```

Source

Example data for the DMRsegal package vignette and examples.

Examples

```
## Not run:
library(ExperimentHub)
eh <- ExperimentHub()
beta <- eh[["EH10275"]]
head(beta[, 1:5])
# Summary statistics
summary(as.vector(beta[1:100, ]))

## End(Not run)
```

`dmps` *Differentially Methylated Positions (DMPs)*

Description

A data frame containing differentially methylated positions (CpG sites) identified by comparing cancer and normal samples. The DMPs were identified using limma with adjusted p-value < 0.05.

Format

A data frame with differentially methylated CpG sites as rows and the following columns:

`intercept`	average expression (M-value) across all samples
`pval`	raw p-value from the statistical test
`f`	F-statistic or B-statistic value
`qval`	q-value (FDR-adjusted p-value using Storey's method)
`pval_adj`	adjusted p-value using Benjamini-Hochberg method

Details

The DMPs were identified using limma's empirical Bayes moderated t-statistics on M-values (log2 ratio of methylated to unmethylated signal). Only CpG sites with adjusted p-value (Benjamini-Hochberg) < 0.05 are included. The results are sorted by p-value.

Analysis Details: The differential methylation analysis included:

- Conversion of beta values to M-values for statistical testing
- Linear modeling with adjustment for age and gender covariates
- Empirical Bayes moderation of t-statistics (limma)
- Multiple testing correction using Benjamini-Hochberg method
- Filtering at FDR < 0.05

Accessing the Data

This resource can be accessed via ExperimentHub:

```
library(ExperimentHub)
eh <- ExperimentHub()

# Query and load
myfiles <- query(eh, "DMRsegaldata")
dmps <- myfiles[["EH10277"]]
```

Source

Derived from the beta matrix and pheno data using limma differential methylation analysis.

Examples

```
## Not run:
library(ExperimentHub)
eh <- ExperimentHub()
dmps <- eh[["EH10277"]]

head(dmps)
nrow(dmps)

# Summary of significance
summary(dmps$pval_adj)
sum(dmps$pval_adj < 0.01)  # Very significant DMPs
sum(dmps$pval_adj < 0.001) # Highly significant DMPs

## End(Not run)
```

---

pheno *Phenotype Data for DNA Methylation Samples*

Description

A data frame containing phenotype information for 20 samples, including sample group (cancer vs. normal), age, and gender.

Format

A data frame with 20 observations on the following 3 variables:

`Sample_Group` a character vector indicating sample group: "cancer" or "normal"

`Age` a numeric vector containing the age of each individual

`Gender` a character vector indicating gender: "m" (male) or "f" (female)

Details

The phenotype data contains information for 10 cancer samples (T1-T10) and 10 normal/healthy samples (H1-H10). Row names correspond to sample identifiers.

Accessing the Data

This resource can be accessed via ExperimentHub:

```
library(ExperimentHub)
eh <- ExperimentHub()

# Query and load
myfiles <- query(eh, "DMRsegaldata")
pheno <- myfiles[["EH10276"]]
```

Source

Example data for the DMRsegal package vignette and examples.

Examples

```
## Not run:
# Via ExperimentHub
library(ExperimentHub)
eh <- ExperimentHub()
pheno <- eh[["EH10276"]]

# Check sample distribution
table(pheno$Sample_Group, pheno$Gender)
summary(pheno$Age)

## End(Not run)
```

Index