Package 'RRBSdata'

November 10, 2026

Type Package

Title An RRBS data set with 12 samples and 10,000 simulated DMRs

Version 1.32.0

Date 2021-11-21

Author Katja Hebestreit, Hans-Ulrich Klein

Maintainer Katja Hebestreit <katja.hebestreit@gmail.com>

Depends R (>= 3.1.0), BiSeq (>= 1.9.2)

Description RRBS data set comprising 12 samples with simulated differentially methylated regions (DMRs).

License LGPL-3

biocViews ExperimentData, CancerData, SequencingData, CpGIslandData

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

git_branch RELEASE_3_23

git_last_commit a9be0e2

git_last_commit_date 2026-04-28

Repository Bioconductor 3.23

Date/Publication 2026-11-10

Contents

diffMethCpGs 2
islands 2
rrbs 4

Index **6**

diffMethCpGs *The differentially methylated CpG sites.*

Description

This object comprises all CpG sites that are differentially methylated between 'cancer' and 'control' samples.

Usage

```
data(diff.meth.cpgs)
```

Format

A `GRanges-class` object with information on the CpG island (`cluster.id`), the covered island region (`cluster.id.2`) and the simulated methylation difference (`meth.diff`).

References

Hebestreit K, Klein HU. Global test and BiSeq are the methods of choice for testing genomic regions for differential methylation in bisulfite sequencing data. In preparation.

Examples

```
data(diffMethCpGs)
head(diffMethCpGs)
```

islands *CpG islands with information on DMRs.*

Description

The 27,718 CpG islands were dowloaded from UCSC database. Only islands with not less than 10 covered CpG sites were considered to receive a DMR (24,698 islands). Within these islands we incorporated 10,000 DMRs.

Usage

```
data(islands)
```

Format

A GRanges-class object with metdata columns:

- cluster.id An ID of the CpG island.
- cpg.no Number of CpG sites within CpG island.
- cpg.perc Percentage of CpG sites on all nucleotides within island (cpg.no/width(islands)).
- ave.cov Average coverage of CpG sites within CpG island.
- KStest.clust Test statistic of test on complete spatial randomness of CpG sites within CpG island.
- KStest.pval.clust P value of test on complete spatial randomness of CpG sites within CpG island.
- dmr.start Start site position of the incorporated DMR. Is NA, if no DMR is incorporated for this CpG island.
- dmr.end End site position of the incorporated DMR. Is NA, if no DMR is incorporated for this CpG island.
- dmr.meth.diff Methylation difference of the incorporated DMR. Is NA, if no DMR is incorporated for this CpG island.
- cpg.no.dmr Number of CpG sites within incorporated DMR. Is NA, if no DMR is incorporated for this CpG island.
- cpg.perc.dmr Percentage of CpG sites within incorporated DMR on all CpG sites within this island. Is NA, if no DMR is incorporated for this CpG island.
- ave.cov.dmr The average coverage of all CpG sites within incorporated DMR. Is NA, if no DMR is incorporated for this CpG island.
- KStest.clust.dmr Test statistic of test on complete spatial randomness of CpG sites within incorporated DMR. Is NA, if no DMR is incorporated for this CpG island.
- KStest.pval.clust.dmr P value of test on complete spatial randomness of CpG sites within incorporated DMR. Is NA, if no DMR is incorporated for this CpG island.

References

Hebestreit K, Klein HU. Global test and BiSeq are the methods of choice for testing genomic regions for differential methylation in bisulfite sequencing data. In preparation.

Examples

```
data(islands)
head(islands)
```

rrbs *An RRBS data set with 12 samples and 10,000 simulated DMRs.*

Description

To obtain a dataset with known DMRs, we used 12 human control samples from a previously published RRBS data set (Schoofs et al., Blood, 2013). We simulated DMRs of different lengths and intensities by altering the number of methylated reads of some of the CpG sites in half of the samples, which could be considered as the cancer samples. For more information on the simulation, see Details.

Usage

```
data(rrbs)
```

Format

A `BSraw-class` object.

Details

We downloaded CpG island positions from UCSC database and filtered out all islands that were not covered in any of the samples (27,718 remaining islands). Only islands with not less than 10 covered CpG sites were considered to receive a DMR (24,698 islands). Within these 24,698 CpG islands we incorporated 10,000 DMRs with methylation differences of 10, 20, 30, or 40%. The DMRs spanned 10, 20, 30, or 40% of the CpG sites of the respective islands. We made sure that we gained the same amount of DMRs for each of the 16 combinations of methylation difference and percentage of modified CpG sites, that is, we gained 1,250 DMRs per combination.

We incorporated the DMRs as follows: 1.) We devided the 12 control samples into 6 "cancer" samples and 6 "control" samples. 2.) Within each CpG island we only considered regions to receive a DMR if each of its CpG sites was covered in at least half of all samples. Those regions are referred to as "covered island regions". 3.) For each CpG site within covered island regions we determined its minimum and maximum smoothed methylation level across all samples. 4.) For each CpG island we determined the maximum percentage of neighbored CpG sites that can be altered by determining the percentage of CpG sites within its biggest covered island region on all CpG sites within the island. 5.) Each of the 10,000 DMRs was sampled into a covered island region of a CpG island that was appropriate to harbor the DMR, in terms of minimum and maximum DNA methylation and CpG percentage of its largest covered island region on all CpG sites within the island. Not more than one DMR was incorporated per CpG island. 6.) Whenever it was possible to increase the DNA methylation by the amount of the difference of the respective DMR (that is, the resulting methylation level is below 1), it was increased in the cancer samples. Otherwise, the DNA methylation was decreased by the amount of difference. 7.) To increase the DNA methylation in a cancer sample within a region by a certain amount, the number of methylated reads of the respective CpG sites was increased. For instance, a CpG site of coverage 12 with 3 methylated reads (and a relative methylation level of 0.25) within a region that should get a methylation difference of 0.3, received 4 additional methylated reads (because: 0.3x12 = 4).

The scripts to simulate the data can be found under /R.

Source

Schoofs T, Rohde C, Hebestreit K, Klein HU, Goellner S, Schulze I, Lerdrup M, Dietrich N, Agrawal-Singh S, Witten A, Stoll M, Lengfelder E, Hofmann WK, Schlenke P, Buechner T, Hansen K, Berdel WE, Rosenbauer F, Dugas M, Mueller-Tidow C (2012). DNA methylation changes are a late event in acute promyelocytic leukemia and coincide with loss of transcription factor binding. Blood.

References

Hebestreit K, Klein HU. Global test and BiSeq are the methods of choice for testing genomic regions for differential methylation in bisulfite sequencing data. In preparation.

Examples

```
data(rrbs)
rrbs
```

Index

∗ **datasets**
diffMethCpGs, 2
islands, 2
rrbs, 4

diffMethCpGs, 2

islands, 2

rrbs, 4