Package ‘dinoR’

May 2, 2024

Title  Differential NOMe-seq analysis

Version  1.0.0

Description  dinoR tests for significant differences in NOMe-seq footprints between two conditions, using genomic regions of interest (ROI) centered around a landmark, for example a transcription factor (TF) motif. This package takes NOMe-seq data (GCH methylation/protection) in the form of a Ranged Summarized Experiment as input. dinoR can be used to group sequencing fragments into 3 or 5 categories representing characteristic footprints (TF bound, nucleosome bound, open chromatin), plot the percentage of fragments in each category in a heatmap, or averaged across different ROI groups, for example, containing a common TF motif. It is designed to compare footprints between two sample groups, using edgeR's quasi-likelihood methods on the total fragment counts per ROI, sample, and footprint category.

License  MIT + file LICENSE

Encoding  UTF-8

Roxygen  list(markdown = TRUE)

RoxygenNote  7.3.1

Suggests  knitr, rmarkdown, testthat (>= 3.0.0)

Config/testthat/edition  3

VignetteBuilder  knitr

Imports  BiocGenerics, circlize, ComplexHeatmap, cowplot, dplyr, edgeR, GenomicRanges, ggplot2, Matrix, methods, rlang, stats, stringr, tibble, tidy, tidyselect

Depends  R (>= 4.3.0), SummarizedExperiment

LazyData  false

biocViews  NucleosomePositioning, Epigenetics, MethylSeq, DifferentialMethylation, Coverage, Transcription, Sequencing, Software

URL  https://github.com/xxxmichi/xxx/dinoR
**BugReports** https://github.com/xxxmichixxx/dinoR/issues

**git_url** https://git.bioconductor.org/packages/dinoR

**git_branch** RELEASE_3_19

**git_last_commit** cb8acee

**git_last_commit_date** 2024-04-30

**Repository** Bioconductor 3.19

**Date/Publication** 2024-05-02

**Author** Michaela Schwaiger [aut, cre] (<https://orcid.org/0000-0002-4522-7810>)

**Maintainer** Michaela Schwaiger <michaela.schwaiger@fmi.ch>

**Contents**

<table>
<thead>
<tr>
<th>Function</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>dinoR-package</td>
<td>2</td>
</tr>
<tr>
<td>compareFootprints</td>
<td>3</td>
</tr>
<tr>
<td>createExampleData</td>
<td>4</td>
</tr>
<tr>
<td>diNOMeTest</td>
<td>5</td>
</tr>
<tr>
<td>footprintCalc</td>
<td>6</td>
</tr>
<tr>
<td>footprintPerc</td>
<td>7</td>
</tr>
<tr>
<td>footprintQuant</td>
<td>9</td>
</tr>
<tr>
<td>fpPercHeatmap</td>
<td>10</td>
</tr>
<tr>
<td>metaPlots</td>
<td>11</td>
</tr>
<tr>
<td>NomeData</td>
<td>12</td>
</tr>
</tbody>
</table>

**Index** 13

---

**dinoR-package**

**dinoR**

**Description**

*dinoR* tests for significant differences in NOMe-seq footprints between two conditions, using genomic regions of interest (ROI) centered around a landmark, for example a transcription factor (TF) motif. This package takes NOMe-seq data (GCH methylation/protection) in the form of a Ranged Summarized Experiment as input. *dinoR* can be used to group sequencing fragments into 3 or 5 categories representing characteristic footprints (TF bound, nucleosome bound, open chromatin), plot the percentage of fragments in each category in a heatmap, or averaged across different ROI groups, for example, containing a common TF motif. It is designed to compare footprints between two sample groups, using edgeR’s quasi-likelihood methods on the total fragment counts per ROI, sample, and footprint category.

**Author(s)**

Michaela Schwaiger
**Descripción**

Compare each footprint pattern in WT and KO samples (percentages and diNOMeTest results).

**Uso**

```r
compareFootprints(
  footprint_percentages,
  res,
  WTsamples = c("WT_1", "WT_2"),
  KOsamples = c("KO_1", "KO_2"),
  plotcols,
  facetROIgroup = FALSE,
  plot = TRUE
)
```

**Argumentos**

- `footprint_percentages` A tibble where each column corresponds to a sample-footprint percentage and each row to a ROI, with the rows clustered by similarity.
- `res` A tibble with the results of differential fragment count testing for each ROI-footprint combination.
- `WTsamples` The control sample names.
- `KOsamples` The treatment sample names.
- `plotcols` A character vector of colors to be used for distinguishing the ROI groups (has to be the same length as there are ROI groups).
- `facetROIgroup` If TRUE, split the plots for each pattern by ROI group.
- `plot` If TRUE, will output a plot.

**Detalles**

Plots the percentages of reads in each ROI in WT versus KO samples (mean of two replicates) in each footprint pattern. The color indicates the ROI group and the shape the results of the diNOMe-Test.
createExampleData

Value

A scatter plot for each footprint pattern comparing WT and KO percentages and significance test results.

Examples

```r
NomeData <- createExampleData()
NomeData <- footprintCalc(NomeData)
NomeData <- footprintQuant(NomeData)
res <- diNOMeTest(NomeData,
    WTsamples = c("WT_1", "WT_2"),
    KOsamples = c("KO_1", "KO_2")
)
footprint_percentages <- footprintPerc(NomeData)
compareFootprints(footprint_percentages, res,
    plotcols = "black", plot = TRUE)
```

description

Creates an RSE object with mock NOMe-seq data.

Usage

```r
createExampleData(
    samples = c("WT_1", "WT_2", "KO_1", "KO_2"),
    group = c("WT", "WT", "KO", "KO"),
    nROI = 4,
    randomMeth = TRUE
)
```

Arguments

- `samples` The sample names.
- `group` The sample group names.
- `nROI` The number of ROIs that should be constructed.
- `randomMeth` Logical indicating whether the methylation/protection values should be randomly generated.

Details

Creates an RSE object with mock NOMe-seq data.
**Value**

RSE object with mock data.

**Examples**

```r
createExampleData()
```

**Description**

Tests for differences in fragment counts for each NOMe footprint pattern compared to total fragment counts in two conditions.

**Usage**

```r
diNOMeTest(
  footprint_counts,
  WTsamples = c("WT_1", "WT_2"),
  KOsamples = c("KO_1", "KO_2"),
  minreads = 1,
  meanreads = 1,
  prior.count = 3,
  FDR = 0.05,
  FC = 2,
  combineNucCounts = FALSE
)
```

**Arguments**

- **footprint_counts**: A Summarized Experiment containing the sample names (colData), ROI names (rowData), and number of fragments in each NOMe footprint pattern category (assays). For example the output of the `footprintQuant` function.
- **WTsamples**: The control sample names as they appear in `colData(footprint_counts)$samples`.
- **KOsamples**: The treatment sample names as they appear in `colData(footprint_counts)$samples`.
- **minreads**: The minimum number of fragments to which a footprint could be assigned a ROI must have in all samples. All other ROIs are filtered out before the differential NOMe analysis.
- **meanreads**: The minimum number of fragments to which a footprint could be assigned a ROI must on average across all samples. All other ROIs are filtered out before the differential NOMe analysis.
- **prior.count**: The pseudocount used for `edgeR::glmQLFit`.
- **FDR**: False Discovery Rate.
- **FC**: Fold Change.
- **combineNucCounts**: Boolean indicating whether to combine nucleotide counts.
### footprintCalc

**FDR**

The FDR cutoff for a ROI - footprint combination to be called regulated in the output.

**FC**

The fold change cutoff for a ROI - footprint combination to be called regulated in the output.

**combineNucCounts**

If TRUE, the upNuc, downNuc, and Nuc fragment counts will be combined into the Nuc category.

#### Details

Uses edgeR’s quasi-likelihood methods to conveniently test for differential proportions of each one of 5 (or 3, if nucleosome footprints are combined) distinct footprints between at least two control and at least two treatment samples.

#### Value

A tibble with the results of differential fragment count testing for each ROI-footprint combination.

#### Examples

```r
NomeData <- createExampleData()
NomeData <- footprintCalc(NomeData)
footprint_counts <- footprintQuant(NomeData)
diNOMeTest(footprint_counts,
    WTsamples = c("WT_1", "WT_2"),
    KOsamples = c("KO_1", "KO_2")
)
```

---

**Description**

Assign a footprint type to each fragment based on GCH protection values in pre-defined windows.

**Usage**

```r
footprintCalc(
    NomeData,
    window_1 = c(-50, -25),
    window_2 = c(-8, 8),
    window_3 = c(25, 50)
)
```
Arguments

NomeData  A Ranged Summarized Experiment (RSE) with an entry for each ROI. The (rowData) should contain information about each ROI, including a ROIgroup. The (assays) should contain at least (nFragsAnalyzed) and (reads). (nFragsAnalyzed) describes the number of fragments that were analyzed for each sample/ROI combination. (reads) contains a Gpos object for each sample/ROI combination, with a position for each base in the ROI and two metadata columns (protection and methylation). protection is a sparse logical matrix where TRUE stands for Cs protected from methylation, and methylation is a sparse logical matrix where TRUE stands for methylated Cs.

window_1  Integer vector with two elements representing start and end positions of the first window relative to the ROI center.

window_2  Integer vector with two elements representing start and end positions of the second window relative to the ROI center.

window_3  Integer vector with two elements representing start and end positions of the third window relative to the ROI center.

Details

Selects 3 windows (default is -50:-25, -8:8, 25:50) around the center of the provided region of interest (ROI) and calculates the average GCH methylation protection for a given fragment across all GCHs in each window. If it is above 0.5 the window is deemed protected, below 0.5, unprotected. Depending on the protection pattern in all windows, a read is put into one of 5 footprint categories: tf bound (0 - 1 - 0), open chromatin (0 - 0 - 0), downstream positioned nucleosome (1 - 1 - 0), other nucleosome (1 - 1 - 1, 1 - 0 - 0, 0 - 0 - 1, 1 - 0 - 1), and upstream positioned nucleosome (0 - 1 - 1).

Value

The Ranged Summarized Experiment with an assay "footprints" added, which contains a footprint type assigned to each fragment.

Examples

NomeData <- createExampleData()
footprintCalc(NomeData)

Description

Calculates the percentage of all fragments in a ROI-sample combination corresponding to each footprint pattern.
Usage

```r
footprintPerc(
  footprint_counts,
  minreads = 1,
  meanreads = 1,
  ROIgroup = "motif",
  combineNucCounts = FALSE
)
```

Arguments

- `footprint_counts`: A Summarized Experiment containing the sample names (colData), ROI names (rowData), and number of fragments in each NOMe footprint pattern category as assays. For example the output of the (`footprintQuant`) function.
- `minreads`: The minimum number of fragments to which a footprint could be assigned a ROI must have in all samples. All other ROIs are removed.
- `meanreads`: The minimum number of fragments to which a footprint could be assigned a ROI must have on average across all samples. All other ROIs are removed.
- `ROIgroup`: Column name of a metadata column in the (rowData) of the RSE, describing a group each ROI belongs to, for example, different transcription factor motifs at the center of the ROI.
- `combineNucCounts`: If TRUE, the upNuc, downNuc, and Nuc fragment counts will be combined into the Nuc category.

Details

Calculates the percentage of all fragments in a ROI-sample combination in each footprint pattern. Then turns the table into wide format, where each column corresponds to a sample-footprint percentage and each row to a ROI and clusters the rows by similarity.

Value

A tibble where each column corresponds to a sample-footprint percentage and each row to a ROI.

Examples

```r
NomeData <- createExampleData()
NomeData <- footprintCalc(NomeData)
footprint_counts <- footprintQuant(NomeData)
footprintPerc(footprint_counts)
```
footprintQuant

Description

Quantify the footprint types.

Usage

footprintQuant(NomeData)

Arguments

NomeData  A Ranged Summarized Experiment (RSE) with an entry for each ROI. The (rowData) should contain information about each ROI, including a ROIgroup. The (assays) should contain at least (nFragsAnalyzed) and (reads). (nFragsAnalyzed) describes the number of fragments that were analyzed for each sample/ROI combination. (reads) contains a Gpos object for each sample/ROI combination, with a position for each base in the ROI and two metadata columns (protection and methylation). protection is a sparse logical matrix where TRUE stands for Cs protected from methylation, and methylation is a sparse logical matrix where TRUE stands for methylated Cs. In addition, there must be an assay called "footprints", which contains the assigned footprint ("tf","open","upNuc","Nuc", "downNuc") for each fragment (generated using the footprintCalc function).

Details

Count the number of fragments corresponding to a footprint type for each sample-ROI combination.

Value

The Ranged Summarized Experiment with an assay added for each footprint type, containing the number of fragments that contain that footprint. An assay with the total number of pattern-able fragments ("all") is also added. tf = transcription factor footprint open = open chromatin footprint upNuc = upstream nucleosome footprint downNuc = downstream nucleosome footprint Nuc = other nucleosome footprints

Examples

NomeData <- createExampleData()
NomeData <- footprintCalc(NomeData)
footprintQuant(NomeData)
fpPercHeatmap

Description

Draws heatmaps of the percentages of all fragments in a ROI-sample combination in each footprint pattern.

Usage

```r
fpPercHeatmap(
  footprint_percentages,
  breaks = rep(list(c(0, 50, 100)), 5),
  plotcols = c("#236467", "#AA9B39", "#822B56", "#822B26", "#822B99")
)
```

Arguments

- `footprint_percentages` A tibble where each column corresponds to a sample-footprint percentage and each row to a ROI, with the rows clustered by similarity.
- `breaks` A list of vectors indicating numeric breaks used in `ColorRamp2` to define the heatmap color gradient, with one element per pattern (usually 5, or 3 if the nucleosome patterns have been combined).
- `plotcols` A character vector of 5 colors to be used for the heatmaps of the 5 footprint patterns ("tf", "open", "upNuc", "Nuc", "downNuc"), or 3 colors if the nucleosome patterns have been combined.

Details

Draws heatmaps of the percentages of all fragments in a ROI-sample combination in each footprint pattern supplied (for example: "tf", "open", "upNuc", "Nuc", "downNuc"). The rows of the heatmaps are split by ROI group.

Value

Heatmaps of the percentages of all fragments in a ROI-sample combination in each footprint pattern.

Examples

```r
NomeData <- createExampleData()
NomeData <- footprintCalc(NomeData)
NomeData <- footprintQuant(NomeData)
footprint_percentages <- footprintPerc(NomeData)
fpPercHeatmap(footprint_percentages)
```
Description

Plot the summarized GCH methylation protections across selected ROIs.

Usage

metaPlots(NomeData, nr = 2, nROI = 2, ROIgroup = "motif", span = 0.05)

Arguments

NomeData  A Ranged Summarized Experiment (RSE) with an entry for each ROI. The (rowData) should contain information about each ROI, including a ROIgroup. The (assays) should contain at least (nFragsAnalyzed) and (reads). (nFragsAnalyzed) describes the number of fragments that were analyzed for each sample/ROI combination. (reads) contains a Gpos object for each sample/ROI combination, with a position for each base in the ROI and two metadata columns (protection and methylation). protection is a sparse logical matrix where TRUE stands for Cs protected from methylation, and methylation is a sparse logical matrix where TRUE stands for methylated Cs.

nr  Integer used as a cutoff to filter sample ROI combinations that have less than (nr) fragments analyzed (nFragsAnalyzed column).

nROI  The number of ROIs that need to have a GpC methylation measurement at a given position for this position to be included in the plot.

ROIgroup  Column name of a metadata column in the rowData of the RSE, describing a group each ROI belongs to, for example, different transcription factor motifs at the center of the ROI.

span  The (span) option to be used for the (loess) function (to draw a line through the datapoints).

Details

Summarizes the GCH methylation protections across selected ROIs.

Value

A tibble with the methylation protection profiles summarized across all ROIs in a certain group.

Examples

NomeData <- createExampleData()
metaPlots(NomeData)
Description

WT and AdnpKO mouse ES cells were subjected to guided NOMe-seq. 1500 regions were targeted using guideRNAs and Cas9, and bisulfite sequenced in 300bp paired-end mode. Reads were mapped to the genome using biscuit. Then we used UMI-tools to remove duplicated UMIs. The GCH protection was determined using the fetch-NOMe package, using the 1500 ROIs as input regions. The ROIs were all 600bp long, and centered around a transcription factor motif. The resulting tibble was converted into a RangedSummarizedExperiment using the NOMeConverteR package. To reduce file size, data were filtered to only those ROIs containing 20-180 fragments.

Usage

data(NomeData)

Format

NomeData: A RangedSummarizedExperiment with 219 ROIs and 4 samples:
  colData sample names
  rowData ROI names with the format: TFmotif_chromosome_start_end and motif type
  assays nFfragsAnalyzed: number of fragments that were analyzed for GCH methylation protection, reads: a Gpos element for every sample-ROI combination
  GPos_GCH_DataMatrix GPos elements with a "protection" metadata column containing a sparse logical matrix indicating if a GCH was protected from methylation and a "methylation" column containing a sparse logical matrix indicating if a GCH was methylated. The GPos has data for every position from 300bp upstream to 300bp downstream around a CTCF motif center.

Value

A RangedSummarizedExperiment with 219 ROIs and 4 samples

Source

generated by Lucas Kaaij
Index

* datasets
  NomeData, 12
* internal
  dinoR-package, 2

compareFootprints, 3
createExampleData, 4
diNOMeTest, 5
dinoR (dinoR-package), 2
dinoR-package, 2

footprintCalc, 6
footprintPerc, 7
footprintQuant, 9
fpPercHeatmap, 10

metaPlots, 11
NomeData, 12