Package 'exomePeak2'

April 10, 2023

Type Package

Title Peak Calling and differential analysis for MeRIP-Seq

Version 1.10.0

Description exomePeak2 provides peak detection and differential methylation for Methylated RNA Immunoprecipitation Sequencing (MeRIP-Seq) data. MeRIP-Seq is a commonly applied sequencing assay that measures the location and abundance of RNA modification sites under specific cellular conditions. In practice, the technique is sensitive to PCR amplification biases commonly found in NGS data. In addition, the efficiency of immunoprecipitation often varies between different IP samples. exomePeak2 can perform peak calling and differential analysis independent of GC content bias and IP efficiency changes.

BugReports https://github.com/ZW-xjtlu/exomePeak2/issues

License Artistic-2.0

Imports
Rsamtools,GenomicAlignments,GenomicRanges,GenomicFeatures,DESeq2,ggplot2,mclust,BSgenome,Biostrings,Genom

Depends R (>= 3.5.0), SummarizedExperiment

biocViews Sequencing, MethylSeq, RNASeq, Coverage,
DifferentialMethylation, DifferentialPeakCalling, PeakDetection

Encoding UTF-8

RoxygenNote 7.1.2

Suggests knitr, rmarkdown

VignetteBuilder knitr

git_url https://git.bioconductor.org/packages/exomePeak2

git_branch RELEASE_3_16

git_last_commit ca31d66

git_last_commit_date 2022-11-01

Date/Publication 2023-04-10

Author Zhen Wei [aut, cre]

Maintainer Zhen Wei <zhen.wei10@icloud.com>

R **topics documented:**

---

exomePeak2   *Peak Calling and Differential Analysis of MeRIP-seq.*

Description

exomePeak2 conducts peak calling and differential methylation analysis using BAM files of aligned MeRIP-seq reads.

Usage

```
exomePeak2(
  bam_ip = NULL,
  bam_input = NULL,
  bam_ip_treated = NULL,
  bam_input_treated = NULL,
  txdb = NULL,
  genome = NULL,
  gff = NULL,
  strandness = c("unstrand", "1st_strand", "2nd_strand"),
  fragment_length = 100,
  bin_size = 25,
  step_size = 25,
  test_method = c("Poisson", "DESeq2"),
  p_cutoff = 1e-10,
  parallel = 1,
  plot_gc = TRUE,
  save_output = TRUE,
  save_dir = getwd(),
  experiment_name = "exomePeak2_output",
  mode = c("exon", "full_transcript", "whole_genome"),
  motif_based = FALSE,
  motif_sequence = "DRACH"
)
```

Arguments

bam_ip	a character vector for the BAM file directories of IP samples.
bam_input	a character vector for the BAM file directories of input samples.
bam_ip_treated	a character vector for the BAM file directories of treated IP samples.

bam_input_treated

          a `character` vector for the BAM file directories of treated input samples.

          The arguments of `BAM_ip` and `BAM_input` are only required for differential methylation analysis.

txdb   a [TxDb](#) object for the transcript annotation.

genome   a `character` or a [BSgenome](#) for the reference genome.

          The character should be the UCSC genome name which is acceptable by [getBSgenome](#) or/and [makeTxDbFromUCSC](#); example: `"hg19"`.

gff   optional, a `character` which specifies the directory toward a gene annotation GFF/GTF file, it is applied when the `TxDb` object is not available; default = `NULL`.

strandness   a `character` specifying the strand protocol type of the RNA-seq library, can be one of `c("unstrand", "1st_strand", "2nd_strand")`; default = `"unstrand"`.

          unstrand The randomly primed RNA-seq library type, i.e. both the strands generated during the first and the second strand sythesis are sequenced; example: Standard Illumina.

          1st_strand The first strand-specific RNA-seq library, only the strand generated during the first strand sythesis is sequenced; examples: dUTP, NSR, NNSR.

          2nd_strand The second strand-specific RNA-seq library, only the strand generated during the second strand sythesis is sequenced; examples: Ligation, Standard SOLiD.

fragment_length

          a positive integer number for the expected fragment length (in bp); default = `100`.

bin_size   a positive integer number for the width of the sliding window; default = `25`.

step_size   a positive integer number for the step size of the sliding window; default = `25`.

test_method   a `character` for the statistical testing method used in peak calling and differential analysis, can be one of `c("Poisson", "DESeq2")`; Default = `"Poisson"`

          `Poisson` Wald test of Poisson GLM.

          `DESeq2` Wald test of negative bionomial GLM with regularized estimation of over-dispersion parameters (implemented by DESeq2),

          Note that when using the test method of DESeq2, a larger p-value cut-off (e.g. 0.001) is often required. The cutoff can be set via the argument `p_cutoff`.

p_cutoff   a `numeric` value for the p value cutoff in peak calling; default = `1e-10`.

parallel   a `numeric` value specifying the number of cores for parallel computing; default = `1`.

plot_gc   a `logical` for saving the plots of bins' GC content v.s. bins' fitted coverage curves, which can be used as a diagnosis for GC content bias; default = `TRUE`.

save_output   a `logical` for saving the outcomes on disk; default = `TRUE`.

save_dir   a `character` for the output directory; default = `getwd`.

experiment_name

          a `character` for the folder name generated in the output directory that contains all the results; default: =`"exomePeak2_output"`

mode   a `character` specifies the scope of peak calling on genome, can be one of
                        `c("exon", "full_transcript", "whole_genome")`; Default = `"exon"`.

      `exon` generate sliding windows over exonic regions.

      `full_transcript` generate sliding windows over the full transcripts (include
         both introns and exons).

      `whole_genome` generate sliding windows over the whole genome (include in-
         trons, exons, and the intergenic regions).

      P.S. The full transcript mode and the whole genome mode demand big memory
      size (> 4GB) for large genomes.

motif_based   a `logical` for detecting (differential) modification over sites of motif; default =
                        `FALSE`. If = `TRUE`, sliding windows will be replaced into the single based sites of
                        the modification motif.

motif_sequence   a `character` for the motif sequence used for the reference sites, it is only applied
                        when `motif_based = TRUE`; default = `"DRACH"`.

Details

[exomePeak2](#) call (differential) RNA modification peaks and calculate peak statistics from **BAM**
files of a MeRIP-seq experiment.

The transcript annotation (from either the [TxDb](#) object or the **GFF** file) should be provided to per-
form analysis on exons.

The genome name or [BSgenome](#) object is required to perform the GC content bias correction. If the
genome argument is not provided (= NULL), the analysis will proceed without GC correction.

If the **BAM** files in treated samples are provided at the arguments `bam_ip_treated` and `bam_input_treated`,
the statistics of differential modification detection on peaks/sites will be reported.

Under the default setting, [exomePeak2](#) will save the results of (differential) modification analysis
under a folder named `'exomePeak2_output'`. The results generated include a **BED** file, a **RDS**
file, and a **CSV** table that stores the locations and statistics of the (differential) modified peaks/sites.

Value

a [GRangesList](#) object, the statistics and other annotations are saved in its metadata columns, which
can be accessed through `mcol()`. If `save_output = TRUE`, exomePeak2 will output results both as
BED, CSV, and RDS files on disk.

Examples

```
## Specify File Directories
GENE_ANNO_GTF = system.file("extdata", "example.gtf", package="exomePeak2")

f1 = system.file("extdata", "IP1.bam", package="exomePeak2")
f2 = system.file("extdata", "IP2.bam", package="exomePeak2")
f3 = system.file("extdata", "IP3.bam", package="exomePeak2")
f4 = system.file("extdata", "IP4.bam", package="exomePeak2")
IP_BAM = c(f1,f2,f3,f4)
f1 = system.file("extdata", "Input1.bam", package="exomePeak2")
f2 = system.file("extdata", "Input2.bam", package="exomePeak2")
```

```
f3 = system.file("extdata", "Input3.bam", package="exomePeak2")
INPUT_BAM = c(f1,f2,f3)

## Peak Calling
res <- exomePeak2(bam_ip = IP_BAM,
                  bam_input = INPUT_BAM,
                  gff = GENE_ANNO_GTF,
                  genome = "hg19")
res        ## Peak ranges
mcols(res) ## Peak statistics

## Differential Peak Detection (Comparison of Two Conditions)

f1 = system.file("extdata", "treated_IP1.bam", package="exomePeak2")
TREATED_IP_BAM = c(f1)
f1 = system.file("extdata", "treated_Input1.bam", package="exomePeak2")
TREATED_INPUT_BAM = c(f1)

res <- exomePeak2(bam_ip = IP_BAM,
                  bam_input = INPUT_BAM,
                  bam_ip_treated = TREATED_IP_BAM,
                  bam_input_treated = TREATED_INPUT_BAM,
                  gff = GENE_ANNO_GTF,
                  genome = "hg19")
res        ## Peak ranges
mcols(res) ## Peak statistics
```

---

quiet   *Silencing unwanted function output.*

Description

Silencing unwanted function output.

Usage

```
quiet(x)
```

Arguments

x   any R expression.

Value

none.

Index