

Package ‘MAAPER’

January 2, 2021

Title Model-based Analysis of Alternative Polyadenylation Using 3' End-linked Reads

Version 1.0.0

Description MAAPER is a computational method developed for model-based analysis of alternative polyadenylation using 3' end-linked reads. It accurately assigns near-Site reads to PASs through statistical modeling, and generates multiple statistics for APA analysis.

License GPL-3

Encoding UTF-8

LazyData true

Roxygen list(markdown = TRUE)

RoxygenNote 7.1.1

Imports parallel,
GenomicRanges,
GenomicAlignments,
GenomicFeatures,
GenomeInfoDb,
stats,
utils,
Rsamtools,
IRanges

R topics documented:

maaper 1

Index 4

maaper	<i>Model-based analysis of alternative polyadenylation using 3' end-linked reads</i>
--------	--------------------------------------------------------------------------------------

Description

Model-based analysis of alternative polyadenylation using 3' end-linked reads

Usage

```
maaper(
  gtf,
  pas_annotation,
  output_dir,
  bam_c1,
  bam_c2,
  read_len,
  ncores = 1,
  num_pas_thre = 25,
  frac_pas_thre = 0.05,
  dist_thre = 600,
  num_thre = 50,
  run = "all",
  subset = NULL,
  region = "all",
  gtf_rds = NULL,
  verbose = FALSE
)
```

Arguments

<code>gtf</code>	A character specifying the full path of the GTF file (reference genome);
<code>pas_annotation</code>	A list containing the pas annotation. MAAPER provides processed annotation information from PolyA_DB v3 on its Github page.
<code>output_dir</code>	A character specifying the full path of the output directory, which is used to store all intermediate and final outputs.
<code>bam_c1</code>	A character vector specifying the full paths to the bam files for condition 1 (control). The length of the vector equals the number of samples.
<code>bam_c2</code>	A character vector specifying the full paths to the bam files for condition 2 (experiment). The length of the vector equals the number of samples.
<code>read_len</code>	An integer specifying the read length.
<code>ncores</code>	An integer specifying the number of cores used in parallel computation.
<code>num_pas_thre</code>	An integer specifying the threshold on PAS's read number. Defaults to 25.
<code>frac_pas_thre</code>	A numeric specifying the threshold on PAS's fraction. Defaults to 0.05.
<code>dist_thre</code>	An integer specifying the threshold on fragment length. Defaults to 600.
<code>num_thre</code>	An integer specifying the threshold on gene's read number. Defaults to 50.
<code>run</code>	"all" (default) or "skip-train". For test and debug only.
<code>subset</code>	A character vector specifying genes' Ensembl IDs if only a subset of genes need to be analyzed. Check the <code>pas_annotation</code> files for ID formats.
<code>region</code>	"all" (default). For test and debug only.
<code>gtf_rds</code>	NULL (default). For test and debug only.
<code>verbose</code>	FALSE (default). For test and debug only.

Value

mapper saves two text files, `gene.txt` and `pas.txt`, to `out_dir`. `pas.txt` contains the gene names, predicted PASs, and their corresponding fractions in the two conditions. `gene.txt` contains the genes' PAS number, p values, RED, RLDu, and RLDi scores.

Author(s)

Wei Vivian Li, <vivian.li@rutgers.edu>

Index

maaper, [1](#)