Package 'reptDir'

June 20, 2024

Title Annotate replication timing direction

Version 0.0.0.9000

Description determines the direction of replication timing domains from previously processed replication timing data found at popular resources such as the UCSC genome browser. Leading and lagging strand annotations are easily derived from the found directions across the genome.

License `use_mit_license()`, `use_gpl3_license()` or friends to pick a license

Encoding UTF-8

Roxygen list(markdown = TRUE)

RoxygenNote 7.2.0

RdMacros Rdpack

Imports data.table, Rdpack

Depends R (>= 2.10)

LazyData true

Suggests knitr, rmarkdown

VignetteBuilder knitr

R topics documented:

mcf7Rept *Replication timing data*

Description

Chromosome 10 Wavelet-smoothed signal data from MCF7 cells as provided by the UCSC genome browser (see source)

Usage

```
mcf7Rept
```

Format

`mcf7Rept`:

A `data.table` with 133,719 rows and 4 columns:

chr: Chromosome name

start: start positions of genomic ranges (open)

end: end positions of genomic ranges (closed)

signal: replication timing signal

Source

https://genome-euro.ucsc.edu/cgi-bin/hgTables?db=hg19&hgta_group=regulation&hgta_track=wgEncodeUwRepliSeq&hgta_table=wgEncodeUwRepliSeqMcf7WaveSignalRep1&hgta_doSchema=describe+table+schema

reptDir *reptDir: Replication Direction Finder*

Description

Determine the replication direction of genomic ranges based on replication timing signal.

Usage

```
reptDir(reptdt, minLen, minSlope)
```

Arguments

reptdt	A data.table with at least the following columns: start: start positions of genomic ranges (open) end: end positions of genomic ranges (closed) signal: replication time signal. The input data.table is assumed to have only genomic ranges of a single chromosome ordered by start position. The ranges are also assumed to be completely disjoint from one another. The replication signal is assumed to be processed already such that it meaningfully depicts the replication time profile of interest. Such data is typically found already in this format in the UCSC genome browser for example. See the reptDir vignette through utils::browseVignettes("reptDir") for a full usage example.
minLen	A scalar integer, minimum number of consecutive bases of direction domains.
minSlope	A scalar numeric, minimum slope at both sides of ranges to be included in direction domains.

Details

The method used for assigning replication direction to each range is inspired by the descriptions of Morganella et al. (2016). Briefly, replication time signal extrema are found by identifying the first genomic ranges switching their slope from positive to negative (peaks) or visceversa (valleys) after a streak of consecutive ranges with the same slope sign. Slopes are defined as the difference between the signal of any range and the previous neighbor range as defined by standard genomic coordinates in the reference strand. Ranges with 0 slope and whose next neighbor also has 0 slope (flat regions), as well as extrema themselves, are assigned NA replication direction. Ranges for which a putative direction assignment is possible are therefore those between extrema or between extrema and flat regions. These collections of consecutive ranges are deemed replication direction domains. Additional ranges may be assigned NA direction by minLen and then by minSlope criteria in that order. These parameters correspond to filters applied by Morganella et al. (2016) and Haradhvala et al. (2016) respectively. if a domain spans less than minLen nucleotides, its ranges are determined to have unreliable replication direction. If a range or its next neighbor has a slope less than minSlope, its replication direction is deemed unreliable. The remaining domains with only reliable ranges are assigned replication direction based on the peaks being replication origins and valleys termination zones. Therefore, a domain whose ranges have a positive slope indicates replication from peak to valley in the "q-to-p" direction relative to the chromosomal arms. A negative slope indicates peak to valley replication in the "p-to-q" direction. The leading strand is the reference strand in q-to-p replication while the lagging strand is the reference in the p-to-q case.

Value

A data.table with the following columns in addition to input columns:
dl: signal difference between current range and previous range to the left
dr: signal difference between next range to the right and current
extrema: whether or not the range is a replication peak or valley
sdl: sign of dl
sdr: sign of dr
direction: replication direction.

References

Haradhvala NJ, Polak P, Stojanov P, Covington KR, Shinbrot E, Hess JM, Rheinbay E, Kim J, Maruvka YE, Braunstein LZ, Kamburov A, Hanawalt PC, Wheeler DA, Koren A, Lawrence MS, Getz G (2016). "Mutational strand asymmetries in cancer genomes reveal mechanisms of DNA damage and repair." *Cell*, **164**(3), 538–549.

Morganella S, Alexandrov LB, Glodzik D, Zou X, Davies H, Staaf J, Sieuwerts AM, Brinkman AB, Martin S, Ramakrishna M, Butler A, Kim H, Borg A, Sotiriou C, Futreal PA, Campbell PJ, Span PN, Van Laere S, Lakhani SR, Eyfjord JE, Thompson AM, Stunnenberg HG, van de Vijver MJ, Martens JWM, Børresen-Dale A, Richardson AL, Kong G, Thomas G, Sale J, Rada C, Stratton MR, Birney E, Nik-Zainal S (2016). "The topography of mutational processes in breast cancer genomes." *Nat. Commun.*, **7**(1), 11383.

Index