Package 'SeqCountGenome'

July 13, 2021

Title Counting sequences occurrences in arbitrary subsets of the genome

Depends R(>= 4.0.0), BSgenome, methods

Imports Biostrings, GenomeInfoDb, GenomicRanges

Suggests BSgenome.Hsapiens.UCSC.hg38, TxDb.Hsapiens.UCSC.hg38.knownGene, GenomicFeatures, knitr, rmarkdown, BiocStyle, testthat

Version 0.99.0

biocViews SequenceMatching, BiomedicalInformatics, Genetics

Description This package provides a function to count the occurrences of a set of search sequences in specific regions of the genome. The occurences are found through exact matching and reported in a count matrix.

License GPL-2

Encoding UTF-8

URL https://github.com/dacolombo/SeqCountGenome

BugReports https://github.com/dacolombo/SeqCountGenome/issues

Roxygen list(markdown = TRUE)

RoxygenNote 7.1.1

VignetteBuilder knitr

NeedsCompilation no

Author Daniele Colombo [aut, cre]

Maintainer Daniele Colombo <daniele20.colombo@mail.polimi.it>

R topics documented:

SeqCountGenome-package *SeqCountGenome - Counting sequences occurrences in subsets of the genome*

Description

The SeqCountGenome package implements a function for counting the occurrences of a set of search sequences in specific regions of the genome. The occurrences are found through exact matching and reported in a count matrix.

Details

Package:	SeqCountGenome
Type:	Package
Version:	0.99.0
Date:	2021-07-13
License:	GPL (>=2)

Author(s)

Daniele Colombo
Politecnico di Milano
Maintainer: Daniele Colombo
E-Mail: daniele20.colombo@mail.polimi.it

References

https://github.com/dacolombo/SeqCountGenome

SeqCountGenome *Count sequences occurrences in subsets of the genome*

Description

This function find exact matches of search sequences in subsets of the genome. The reference genome, the regions of interest and a character vector containing the search sequences must be passed. Optionally, regions for which no match is found can be dropped and the total count of matches can be computed for each search sequence. A count matrix is returned, with search sequences as columns and genomic regions as rows.

Usage

```
SeqCountGenome(genome,
regions,
patterns,
overall.count = FALSE,
remove.0.matches = FALSE)
```

Arguments

genome	A BSgenome object corresponding to the reference genome
regions	A GRanges object containing the genomic regions of interest
patterns	A character vector containing the search sequences
overall.count	A logical, if TRUE the sum of matches for each search sequence will be computed
remove.0.matches	A logical, if TRUE the regions with no matches will be removed from the output matrix

Value

A matrix containing the exact matches counts, with search sequences as columns and genomic regions as rows

Author(s)

Daniele Colombo
Politecnico di Milano
Maintainer: Daniele Colombo
E-Mail: daniele20.colombo@mail.polimi.it

Examples

```
# Reference genome: hg38
Hsapiens <- BSgenome.Hsapiens.UCSC.hg38::BSgenome.Hsapiens.UCSC.hg38
# Genomic regions: promoters
txdb <- TxDb.Hsapiens.UCSC.hg38.knownGene::TxDb.Hsapiens.UCSC.hg38.knownGene
promoters <- promoters(txdb, upstream=2000, downstream=200)
promoters <- promoters[grep("_",seqnames(promoters),invert=TRUE)]
# Search patterns: transcription factor consensus
transcription_consensus <- c("GTCAAGGTCA","CACGTG","GGGCGG")

SeqCountGenome(Hsapiens,promoters,transcription_consensus)
```

Index