Package 'GenomicScores'

October 17, 2017

Type Package

Title Infrastructure to work with genomewide position-specific scores

Description Provide infrastructure to store and access genomewide position-specific scores within R and Bioconductor.

Version 1.0.2

License Artistic-2.0

Depends R (>= 3.4), S4Vectors (>= 0.7.21), GenomicRanges, methods, BiocGenerics (>= 0.13.8)

Imports utils, XML, IRanges (>= 2.3.23), BSgenome, GenomeInfoDb, AnnotationHub

Suggests BiocStyle, knitr, rmarkdown, BSgenome.Hsapiens.UCSC.hg19, phastCons100way.UCSC.hg19, VariantAnnotation, TxDb.Hsapiens.UCSC.hg19.knownGene

VignetteBuilder knitr

URL `https://github.com/rcastelo/GenomicScores`

BugReports `https://github.com/rcastelo/GenomicScores/issues`

Encoding UTF-8

biocViews Infrastructure, Genetics, Annotation, Sequencing, Coverage

NeedsCompilation no

Author Robert Castelo [aut, cre],
Pau Puigdevall [ctb]

Maintainer Robert Castelo `<robert.castelo@upf.edu>`

R topics documented:

GScores-class 2
scores 3

Index **6**

`GScores-class` *The GenomicScores package and the GScores class*

Description

Provide support to store and retrieve genomic scores associated to nucleotide positions along a genome.

Usage

```
GScores(provider, provider_version, download_url,
        download_date, reference_genome,
        data_pkgname, data_dirpath, data_serialized_objnames)
```

Arguments

`provider`	Data provider.
`provider_version`	Version of the data, as given by the data provider.
`download_url`	URL from where data were downloaded.
`download_date`	Date when data were downloaded.
`reference_genome`	GenomeDescription object containing information about the reference genome to where data belongs to.
`data_pkgname`	Name of the package storing the data.
`data_dirpath`	Path to the directory where the actual data is stored.
`data_serialized_objnames`	Serialized object names.

Details

The `GenomicScores` package defines the `GScores` class and associated methods to provide support to annotation packages and resources that store genomic scores. In the context of the `GenomicScores` package, genomic scores are numeric values associated to physical nucleotide positions along a genome, such as UCSC phastCons conservation scores for human. Two such annotation packages are:

`phastCons100way.UCSC.hg19` Nucleotide-level phastCons conservation scores from the UCSC Genome Browser calculated from multiple genome alignments from the human genome version hg19 to 99 vertebrate species.

`phastCons100way.UCSC.hg38` Nucleotide-level phastCons conservation scores from the UCSC Genome Browser calculated from multiple genome alignments from the human genome version hg38 to 99 vertebrate species.

Value

The `GScores()` constructor function returns a `GScores-class` object. In principle, the end-user needs not to call this function.

Author(s)

R. Castelo

See Also

`phastCons100way.UCSC.hg19` `phastCons100way.UCSC.hg38`

Examples

```
## supporting annotation packages with genomic scores
if (require(phastCons100way.UCSC.hg19)) {
  library(GenomicRanges)

  gsco <- phastCons100way.UCSC.hg19
  gsco
  scores(gsco, GRanges(seqnames="chr7", IRanges(start=117232380, width=5)))
}

## supporting AnnotationHub resources
## Not run:
availableGScores()
gsco <- getGScores("phastCons100way.UCSC.hg19")
scores(gsco, GRanges(seqnames="chr7", IRanges(start=117232380, width=5)))

## End(Not run)

## meta information about a GScores object
provider(gsco)
providerVersion(gsco)
organism(gsco)
seqlevelsStyle(gsco)
seqinfo(gsco)
head(seqnames(gsco))
head(seqlengths(gsco))
```

`scores` *Accessing genomic scores*

Description

Functions to access genomic scores through `GScores` objects.

Usage

```
availableGSscores()
getGScores(x)
## S4 method for signature 'GScores,GenomicRanges'
scores(object, ranges, scores.only, ...)
```

Arguments

`x`	A `character` vector of length 1 specifiying the genomic scores resource to fetch. The function `availableGScores()` shows the available genomic scores resources.
`object`	A `GScores` object.
`ranges`	A `GenomicRanges` object with positions from where to retrieve genomic scores.
`scores.only`	Flag set to `FALSE` (default) when scores are return in a metadata column called `scores` from the input `GenomicRanges` object. When set to `TRUE`, the only the numeric vector of scores is returned.
`...`	In the call to the `scores`() method one can additionally set the following arguments: • summaryFunFunction to summarize genomic scores when more than one position is retrieved. By default, this is set to the arithmetic mean, i.e., the `mean()` function. • cachingFlag setting whether genomic scores per chromosome should be kept cached in memory (`TRUE`, default) or not (`FALSE`). The latter option minimizes the memory footprint but slows down the performance when the `scores()` method is called multiple times.

Details

The method `scores()` takes as first argument a `GScores-class` object that can be loaded from an annotation package or from an `AnnotationHub` resource. These two possibilities are illustrated in the examples below.

Value

The function `availableGScores()` returns a character vector with the names of the AnnotationHub resources corresponding to different available sets of genomic scores. The function `getGScores()` return a `GScores` object. The method `scores()` returns a numeric vector.

Author(s)

R. Castelo

See Also

`phastCons100way.UCSC.hg19` `phastCons100way.UCSC.hg38`

Examples

```
## accessing genomic scores from an annotation package
if (require(phastCons100way.UCSC.hg19)) {
  library(GenomicRanges)

  gsco <- phastCons100way.UCSC.hg19
  gsco
  scores(gsco, GRanges(seqnames="chr7", IRanges(start=117232380, width=5)))
}

## accessing genomic scores from AnnotationHub resources
## Not run:
```

```
availableGScores()
gsco <- getGScores("phastCons100way.UCSC.hg19")
scores(gsco, GRanges(seqnames="chr7", IRanges(start=117232380, width=5)))

## End(Not run)
```

Index

∗Topic **datasets**
    GScores-class, 2
    scores, 3

availableGScores *(*scores*)*, 3

GenomicScores *(*GScores-class*)*, 2
getGScores *(*scores*)*, 3
GScores *(*GScores-class*)*, 2
GScores-class, 2

organism,GScores-method *(*GScores-class*)*, 2

phastCons100way.UCSC.hg19, *3*, *4*
phastCons100way.UCSC.hg38, *3*, *4*
provider,GScores-method *(*GScores-class*)*, 2
providerVersion,GScores-method *(*GScores-class*)*, 2

referenceGenome,GScores-method *(*GScores-class*)*, 2

scores, 3
scores,GScores,GenomicRanges-method *(*scores*)*, 3
seqinfo,GScores-method *(*GScores-class*)*, 2
seqlengths,GScores-method *(*GScores-class*)*, 2
seqlevelsStyle,GScores-method *(*GScores-class*)*, 2
seqnames,GScores-method *(*GScores-class*)*, 2
show,GScores-method *(*GScores-class*)*, 2