GenomicFeatures

April 19, 2010

`extractTranscriptsFromGenome` *Extract transcripts from a genome*

Description

`extractTranscriptsFromGenome` extracts the transcript sequences from a BSgenome data package using transcript information (exon boundaries) stored in a "gene table".

Usage

```
extractTranscriptsFromGenome(genome, genes)
```

Arguments

`genome`	A `BSgenome` object. See the `available.genomes` function in the BSgenome package for how to install a genome.
`genes`	A data.frame like that returned by `geneHuman`.

Value

A `DNAStringSet` object.

Note

`extractTranscriptsFromGenome` is based on the `extractTranscripts` function defined in the Biostrings package. See `¿extractTranscripts‘` for more information and related functions like `transcriptLocs2refLocs` for converting transcript-based locations into chromosome-based (aka reference-based) locations.

Author(s)

H. Pages

See Also

`available.genomes`, `geneHuman`, `transcriptLocs2refLocs`

Examples

```
library(GenomicFeatures.Hsapiens.UCSC.hg18)  # load the gene table
genes <- geneHuman()
library(Biostrings)  # for transcriptWidths()
tw <- transcriptWidths(genes$exonStarts, genes$exonEnds)

if (interactive()) {
  library(BSgenome.Hsapiens.UCSC.hg18)  # load the genome
  ## Takes about 30 sec.:
  transcripts <- extractTranscriptsFromGenome(Hsapiens, genes)
  ## Sanity check:
  stopifnot(identical(width(transcripts), tw))
}

## Get the reference-based locations of the first 4 (5' end)
## and last 4 (3' end) nucleotides in each transcript:
tlocs <- lapply(tw, function(w) c(1:4, (w-3):w))
rlocs <- transcriptLocs2refLocs(tlocs, genes$exonStarts, genes$exonEnds,
             genes$strand, reorder.exons.on.minus.strand=TRUE)
```

`regions` *Functions that compute genomic regions of interest.*

Description

Functions that compute genomic regions of interest such as promotor, upstream regions etc, from the genomic locations provided in data like the data.frame returned by `geneHuman`.

Usage

```
transcripts(genes, proximal = 500, distal = 10000)
exons(genes)
introns(genes)
```

Arguments

`genes`	A data.frame like that returned by `geneHuman`.
`proximal`	The number of bases on either side of TSS and 3'-end for the promoter and end region, respectively.
`distal`	The number of bases on either side for upstream/downstream, i.e. enhancer/silencer regions.

Details

The assumption made for introns is that there must be more than one exon, and that the introns are between the end of one exon and before the start of the next exon.

Value

All of these functions return a `RangedData` object with a `gene` column with the UCSC ID of the gene. For `transcripts`, each element corresponds to a transcript, and there are columns for each type of region (promoter, threeprime, upstream, and downstream). For `exons`, each element corresponds to an exon. For `introns`, each element corresponds to an intron.

Author(s)

M. Lawrence.

Examples

```
library(GenomicFeatures.Hsapiens.UCSC.hg18)
## promoter 300bp up and down from TSS (threeprime from TES)
transcripts(geneHuman(), proximal = 300)
```

Index