leeBamViews

October 5, 2010

`bs1` *BamViews instance construction related to yeast RNA-seq*

Description

BamViews instance construction related to yeast RNA-seq

Format

The format is: Formal class 'BamViews' [package "Rsamtools"] with 5 slots
..@ bamPaths : chr [1:8] "/Users/stvjc/ExternalSoft/R-devel/library/leeBamViews/bam/isowt5_13e.bam"
"/Users/stvjc/ExternalSoft/R-devel/library/leeBamViews/bam/isowt6_13e.bam" "/Users/stvjc/ExternalSoft/R-
devel/library/leeBamViews/bam/rlp5_13e.bam" "/Users/stvjc/ExternalSoft/R-devel/library/leeBamViews/bam/rlp6_13e.
...
..@ bamIndicies : chr [1:8] "/Users/stvjc/ExternalSoft/R-devel/library/leeBamViews/bam/isowt5_13e.bam"
"/Users/stvjc/ExternalSoft/R-devel/library/leeBamViews/bam/isowt6_13e.bam" "/Users/stvjc/ExternalSoft/R-
devel/library/leeBamViews/bam/rlp5_13e.bam" "/Users/stvjc/ExternalSoft/R-devel/library/leeBamViews/bam/rlp6_13e.
...
..@ bamSamples :Formal class 'DataFrame' [package "IRanges"] with 6 slots
.. .. ..@ rownames : chr [1:8] "isowt.5" "isowt.6" "rlp.5" "rlp.6" ...
.. .. ..@ nrows : int 8
.. .. ..@ elementMetadata: NULL
.. .. ..@ elementType : chr "ANY
" .. .. ..@ metadata : list()
.. .. ..@ listData :List of 2
.. ..$ geno: chr [1:8] "isowt" "isowt" "rlp" "rlp" ...
.. ..$ lane: chr [1:8] "5" "6" "5" "6" ...
..@ bamRanges :Formal class 'GRanges' [package "GenomicRanges"] with 7 slots
.. .. ..@ seqnames :Formal class 'Rle' [package "IRanges"] with 5 slots
.. ..@ values : Factor w/ 1 level "Scchr13": 1
.. ..@ lengths : int 27
.. ..@ elementMetadata: NULL
.. ..@ elementType : chr "ANY"
.. ..@ metadata : list()
.. .. ..@ ranges :Formal class 'IRanges' [package "IRanges"] with 6 slots
.. ..@ start : int [1:27] 798517 801771 804455 808999 810465 811088 818826 820255
822762 832338 ...
.. ..@ width : int [1:27] 2862 933 636 234 114 108 1122 2199 1869 915 ...

.. ..@ NAMES : NULL
.. ..@ elementMetadata: NULL
.. ..@ elementType : chr "integer"
.. ..@ metadata : list()
.. .. ..@ strand :Formal class 'Rle' [package "IRanges"] with 5 slots
.. ..@ values : Factor w/ 3 levels "+","-","*": 1
.. ..@ lengths : int 27
.. ..@ elementMetadata: NULL
.. ..@ elementType : chr "ANY"
.. ..@ metadata : list()
.. .. ..@ seqlengths : Named int NA
.. ..- attr(*, "names")= chr "Scchr13"
.. .. ..@ elementMetadata:Formal class 'DataFrame' [package "IRanges"] with 6 slots
.. ..@ rownames : NULL
.. ..@ nrows : int 27
.. ..@ elementMetadata: NULL
.. ..@ elementType : chr "ANY"
.. ..@ metadata : list()
.. ..@ listData :List of 1
.. ..$ name: chr [1:27] "YMR266W" "YMR267W" "YMR269W" "YMRWdelta20" ...
.. .. ..@ elementType : chr "ANY"
.. .. ..@ metadata : list()
..@ bamExperiment:List of 1
.. ..$ annotation: chr "org.Sc.sgd.db"

Details

Illumina short reads from a very small segment of yeast chr XIII have been collected

Source

FASTQ data available at ftp://ftp.ncbi.nlm.nih.gov/sra/Studies/SRP000/SRP000632/

References

Albert Lee and Kasper Daniel Hansen and James Bullard and Sandrine Dudoit and Gavin Sherlock, Novel Low Abundance and Transient RNAs in Yeast Revealed by Tiling Microarrays and Ultra High–Throughput Sequencing Are Not Conserved Across Closely Related Yeast Species, PLoS Genet, v4, e1000299, Dec 2008

Examples

```
library(leeBamViews)  # bam files stored in package
bpaths = dir(system.file("bam", package="leeBamViews"), full=TRUE, patt="bam$")
#
# extract genotype and lane information from filenames
#
gt = gsub(".*/", "", bpaths)
gt = gsub("_.*", "", gt)
lane = gsub(".*(.)$", "\\1", gt)
geno = gsub(".$", "", gt)
#
# format the sample-level information appropriately
```

```
#
pd = DataFrame(geno=geno, lane=lane, row.names=paste(geno,lane,sep="."))
prd = new("DataFrame")  # protocol data could go here
#
# create the views object, adding some arbitrary experiment-level information
#
bs1 = BamViews(bamPaths=bpaths, bamSamples=pd,
        bamExperiment=list(annotation="org.Sc.sgd.db"))
bs1
# add ranges and tabulate reads

START=c(861250, 863000)
END=c(862750, 864000)
exc = GRanges(IRanges(start=START, end=END), seqnames="Scchr13", strand="+")
values(exc)$name = c("intv1", "intv2")  # necessary
bamRanges(bs1) = exc
bs1
tabulateReads(bs1, "+")
```

`tabulateReads` *tabulate counts of alignments occurring in specified genomic regions*

Description

tabulate counts of alignments occurring in specified genomic regions

Usage

```
tabulateReads(bv, strandmarker=NULL, as.GRanges=FALSE)
```

Arguments

`bv`	`BamViews-class` instance
`strandmarker`	character atom: '+' or '-'; if missing, ignore strand
`as.GRanges`	logical directive to return a GRanges instance instead of a matrix

Details

`readBamGappedAlignments` is the basic engine for this task

Value

annotated matrix with start, end, and samples as rows, regions as columns, and read counts as cell entries

Author(s)

VJ Carey <stvjc@channing.harvard.edu>

Examples

```
example(bs1)
#
# counts in a partition
#
myrn = GRanges(IRanges(start=seq(861250, 862750, 100), width=100),
    seqnames="Scchr13", strand="+")

values(myrn)$name = paste("til", 1:length(myrn), sep=".")
bamRanges(bs1) = myrn
tabulateReads(bs1, "+")
#
# a related computation based on countBam
lapply(bamPaths(bs1)[1:2], function(x)
    countBam(x,  param=ScanBamParam(which=bamRanges(bs1))))
```

`totalReadCounts` *scan BAM files for total read counts*

Description

scan BAM files for total read counts

Usage

```
totalReadCounts(x)
```

Arguments

`x`	`BamViews-class` instance

Details

slow procedure – does lightweight scan of entire file

Value

named integer vector of read counts per sample

Author(s)

VJ Carey <stvjc@channing.harvard.edu>

Examples

```
example(bs1)
totalReadCounts(bs1)
```

Index