Package ‘VariantToolsData’

April 2, 2024

Type Package
Version 1.26.0
Maintainer Michael Lawrence <michafla@gene.com>
License Artistic-2.0
Title Data for the VariantTools tutorial
Author Michael Lawrence
Description Data from the sequencing of a 50/50 mixture of HapMap trio samples NA12878 (CEU) and NA19240 (YRI), subset to the TP53 region.
Suggests VariantTools (&ge; 1.3.4), gmapR (&ge; 1.3.3), BiocStyle
Imports BiocGenerics, GenomicRanges
Depends R (&ge; 2.10), VariantAnnotation (&ge; 1.7.35)
biocViews ExperimentData, SequencingData, HapMap, SNPData
git_url https://git.bioconductor.org/packages/VariantToolsData
git_branch RELEASE_3_18
git_last_commit 946f731
git_last_commit_date 2023-10-24
Repository Bioconductor 3.18
Date/Publication 2024-04-02

R topics documented:

<table>
<thead>
<tr>
<th>topic</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>geno</td>
<td>2</td>
</tr>
<tr>
<td>repeats</td>
<td>2</td>
</tr>
<tr>
<td>sumDepths</td>
<td>3</td>
</tr>
<tr>
<td>tallies</td>
<td>4</td>
</tr>
<tr>
<td>TP53Region</td>
<td>5</td>
</tr>
</tbody>
</table>

Index 6
Reference genotypes

Description
Reference genotypes for NA12878 and NA19240, as called by the HapMap project, and the Broad GATK project.

Usage
data(geno)

Format
A VRanges with the genotypes for NA12878 and NA19240 from the HapMap pilot project and the Broad/GATK calling of NA12878. The genotypes, stored in the "freq." columns, are represented by the alt frequency, so 0/0.5/1 for hom-ref/het/hom-alt. The "expected.freq" column indicates the alt frequency expected in the 50/50 mixture.

Source
The HapMap Pilot and Broad GATK projects.

Examples
data(geno)table(geno$expected.freq)

Simple repeats

Description
Repeat regions from the RepeatMasker track of the hg19 UCSC genome browser database, subset to low complexity and simple repeats.

Usage
data(repeats)

Format
A GRanges object with the repeat ranges, including variables classifying the repeats by name, class, and family, and information about the alignment of the repeat consensus to the genome.
**sumDepths**

**Source**

The "rmsk" table in the UCSC table browser (hg19). Click the "Describe Schema" button for details.

**Examples**

```r
data(repeats)
tab <- table(repeats$repFamily)
tab[tab > 0]
```

<table>
<thead>
<tr>
<th>sumDepths</th>
<th>Sum Replicate Depths</th>
</tr>
</thead>
</table>

**Description**

Finds the unique variants across every element of a list of VRanges, with depths computed by summing the depths over the elements. If a variant is not found in a sample, the depth is assumed to zero. That is a very dangerous assumption.

**Usage**

```r
sumDepths(x)
```

**Arguments**

- `x` A VRangesList, typically of replicates

**Value**

A VRanges

**Author(s)**

Michael Lawrence

**Examples**

```r
data(tallies)
sumDepths(tallies)
```
**tallies**

**Tally VRanges**

**Description**

Nucleotide tallies computed over the TP53 region (+/- 1Mb) for the 50/50 NA12878/NA19240 mixture, separately for each replicate. Each replicate corresponds to a separate biochemical mixing.

**Usage**

data(tallies)

**Format**

A VRangesList, each VRanges of which corresponds to one of the three biochemical replicates.

**Source**

Computed from the alignments of the FASTQ files found in the `inst/extdata` directory. Repeat regions (see `repeats`) were excluded. For example, for one replicate,

```r
library(gmapR)
extdata.dir <- system.file("extdata", package="VariantToolsData")
bams <- BamFileList(tools::list_files_with_exts(extdata.dir, "bam"))
data(repeats, package = "VariantToolsData")
param <- TallyVariantsParam(TP53Genome(), mask = repeats, read_pos=TRUE, read_length=75L)
tallies <- split(tallyVariants(bams, param), ~ sampleNames)
```

This assumes that the BAM files have been generated for the current version of the TP53 genome:

```r
param <- GsnapParam(TP53Genome(), unique_only = TRUE, molecule = "DNA")
first.fastq <- dir(extdata.dir, "first.fastq", full.names=TRUE)
last.fastq <- dir(extdata.dir, "last.fastq", full.names=TRUE)
output <- gsnap(first.fastq, last.fastq, param)
bams <- as(output, "BamFileList")
```

**References**

TP53Region

Examples

```r
data(tallies)
VariantTools::callVariants(tallies[[1L]])
```

---

**TP53Region**  
*Range around TP53*

**Description**

Returns a GRanges object consisting of the TP53 coordinates in hg19. All coordinates in these data are relative to that region.

**Usage**

```r
TP53Region()
```

**Value**

A GRanges of the extents of the TP53 gene in hg19.

**Author(s)**

Michael Lawrence

**Examples**

```r
TP53Region()
```
Index

* datasets
  geno, 2
  repeats, 2
  tallies, 4

geno, 2
repeats, 2, 4
sumDepths, 3
tallies, 4
TP53Region, 5