Data for the VariantTools Tutorial

Michael Lawrence*

Genentech

*michafla@gene.com

November 23, 2021

Package

VariantToolsData 1.19.1

Contents

1 Overview

In support of the tutorial vignette for the *VariantTools* package, *VariantToolsData* provides a dataset derived from sequencing a 50/50 mixture of the HapMap samples NA12878 and NA19240, where the mixing was performed in triplicate ([1]). The data are subset over the region of TP53 (+/- 1 Mb).

There are several summarized objects accessible via `data()` and documented in the package manual. In addition, there are data files stored here:

```
> dir(system.file("extdata", package="VariantToolsData"))

 [1] "SAM7991860-p53-first.bam"      "SAM7991860-p53-first.bam.bai"
 [3] "SAM7991860-p53-first.fastq.gz" "SAM7991860-p53-last.fastq.gz"
 [5] "SAM7991861-p53-first.bam"      "SAM7991861-p53-first.bam.bai"
 [7] "SAM7991861-p53-first.fastq.gz" "SAM7991861-p53-last.fastq.gz"
 [9] "SAM7991862-p53-first.bam"      "SAM7991862-p53-first.bam.bai"
[11] "SAM7991862-p53-first.fastq.gz" "SAM7991862-p53-last.fastq.gz"
[13] "dbsnp-p53.vcf.gz"              "dbsnp-p53.vcf.gz.tbi"
```

These include FASTQ files with the raw reads and BAM files with the alignments, as well as a VCF file derived from dbSNP.

Please see the *VariantTools* vignette to learn how to work with these data with *Bioconductor*.

References

[1] Michael Lawrence, Melanie A Huntley, Eric Stawiski, Art Owen, Thomas D Wu, Leonard D Goldstein, Yi Cao, Jeremiah Degenhardt, Jason Young, Joseph Guillory, et al. Genomic variant calling: Flexible tools and a diagnostic data set. *bioRxiv*, page 027227, 2015.