genomationData: R package with high throughput genomic data

Altuna Akalin, Vedran Franke

October 25, 2013

Contents

1   Introduction

genomationData is a Bioconductor-compliant R package that contains raw and processed data from high throughput genomics experiments. The data is originaly intended for demo of the Bioconductor package *genomation*. *genomationData* holds data from CHiP Seq and Bisulphite sequencing experiments produced by Encode and Epigenetics Roadmap. All datasets come from human H1 embryonic stem cell line and are mapped to the hg19/GRCh37 version of the genome.

2   Getting started

To load the *genomationData* package into your R envirnoment type:

```
library(genomationData)
```

2.1   Sample description

To list the available files type:

```
list.files(system.file("extdata", package = "genomationData"))
```

The package currently holds 5 CHiP Sequencing and one Bisulfite sequencing samples. Files with the .bam extension contain raw CHiP seq reads, while *.broadPeak and *.narrowPeak represent processed peaks. Due to the large sizes of the samples, the *.bam files have been restricted to human chromosome 21 (chr21).

H1.Bisulfite-Seq.combined.chr21.bedGraph.gz' contains methylation calls for each CpG dinucleotide on human chromosome 21. The original file was converted from wig into bedGraph format using the standard UCSC applications.

To see complete information about the files, take a look into SamplesInfo.txt

```
samp.file = system.file('extdata/SamplesInfo.txt',
                        package='genomationData')
samp.info = read.table(samp.file, header=TRUE, sep='\t')
head(samp.info)
```

3 Session Info

```
sessionInfo()
```

```
## R version 4.3.0 RC (2023-04-18 r84287)
## Platform: x86_64-pc-linux-gnu (64-bit)
## Running under: Ubuntu 22.04.2 LTS
##
## Matrix products: default
## BLAS:   /home/biocbuild/bbs-3.18-bioc/R/lib/libRblas.so
## LAPACK: /usr/lib/x86_64-linux-gnu/lapack/liblapack.so.3.10.0
##
## locale:
##  [1] LC_CTYPE=en_US.UTF-8       LC_NUMERIC=C
##  [3] LC_TIME=en_GB              LC_COLLATE=C
##  [5] LC_MONETARY=en_US.UTF-8    LC_MESSAGES=en_US.UTF-8
##  [7] LC_PAPER=en_US.UTF-8       LC_NAME=C
##  [9] LC_ADDRESS=C               LC_TELEPHONE=C
## [11] LC_MEASUREMENT=en_US.UTF-8 LC_IDENTIFICATION=C
##
## time zone: America/New_York
## tzcode source: system (glibc)
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## loaded via a namespace (and not attached):
## [1] compiler_4.3.0 formatR_1.14   tools_4.3.0    highr_0.10     knitr_1.42
## [6] xfun_0.39      evaluate_0.20
```