

R Package for T4RR S16

Schuyler

May 6, 2016

binnit

Binnit is a function designed to split **BED format** data into ‘bins’, or smaller subsamples, either of a designated size or proportion of the total size of the primary. BED format is used for genome data but may be applied to other datasets as well. BED format lists the chromosome in column 1 and the size of the chromosome in column 2.

- BED format

Chromosome	Size
1	3000
2	5000
3	2000

How to run binnit

The first thing you need is a dataset in the correct format:

```
data <- data.frame("chromosome"=sequence(3), "size"=round(runif(3,5000000,5000000)))
data
```

```
##   chromosome      size
## 1           1 36882590
## 2           2 27025563
## 3           3 25035323
```

With that dataset and the binnit function loaded into the R workspace the function can then be run.

```
source("./R/Create_Genome_Intervals.R")
```

The function takes three arguments, the first being the BED format dataset. The second argument asks for intervals (TRUE|FALSE). When FALSE each chromosome will be split into bins of equal size, when TRUE then they will be split into bins of a given size. The third argument is the number of bins or nubmer of Megabases for each bin that’s being created.

```
create.genome.intervals(chromosome.sizes=data, intervals=TRUE, size=5)
```

```
##      [,1]      [,2]      [,3]
## [1,]    1         1 5268941
## [2,]    1 5268942 10537883
## [3,]    1 10537884 15806824
## [4,]    1 15806825 21075766
```

```
## [5,] 1 21075767 26344707
## [6,] 1 26344708 31613649
## [7,] 1 31613650 36882590
## [8,] 2 1 5405113
## [9,] 2 5405114 10810225
## [10,] 2 10810226 16215338
## [11,] 2 16215339 21620450
## [12,] 2 21620451 27025563
## [13,] 3 1 5007065
## [14,] 3 5007066 10014129
## [15,] 3 10014130 15021194
## [16,] 3 15021195 20028258
## [17,] 3 20028259 25035323
```

vs

```
create.genome.intervals(chromosome.sizes=data, intervals=FALSE, size=5)
```

```
##      [,1]      [,2]      [,3]
## [1,] 1      1 7376518
## [2,] 1 7376519 14753036
## [3,] 1 14753037 22129554
## [4,] 1 22129555 29506072
## [5,] 1 29506073 36882590
## [6,] 2 1 5405113
## [7,] 2 5405114 10810225
## [8,] 2 10810226 16215338
## [9,] 2 16215339 21620450
## [10,] 2 21620451 27025563
## [11,] 3 1 5007065
## [12,] 3 5007066 10014129
## [13,] 3 10014130 15021194
## [14,] 3 15021195 20028258
## [15,] 3 20028259 25035323
```