

```

source("https://bioconductor.org/biocLite.R")

## Bioconductor version 3.4 (BiocInstaller 1.24.0), ?biocLite for help
biocLite("GenomicRanges")

## BioC_mirror: https://bioconductor.org
## Using Bioconductor 3.4 (BiocInstaller 1.24.0), R 3.3.3 (2017-03-06).
## Installing package(s) 'GenomicRanges'
## installation path not writeable, unable to update packages: XML,
## cluster,
## mgcv, survival

biocLite("rtracklayer")

## BioC_mirror: https://bioconductor.org
## Using Bioconductor 3.4 (BiocInstaller 1.24.0), R 3.3.3 (2017-03-06).
## Installing package(s) 'rtracklayer'
## installation path not writeable, unable to update packages: XML,
## cluster,
## mgcv, survival

biocLite("GenomicAlignments")

## BioC_mirror: https://bioconductor.org
## Using Bioconductor 3.4 (BiocInstaller 1.24.0), R 3.3.3 (2017-03-06).
## Installing package(s) 'GenomicAlignments'
## installation path not writeable, unable to update packages: XML,
## cluster,
## mgcv, survival

library("GenomicRanges")

## Loading required package: stats4
## Loading required package: BiocGenerics
## Loading required package: parallel
##
## Attaching package: 'BiocGenerics'
## The following objects are masked from 'package:parallel':
##
##   clusterApply, clusterApplyLB, clusterCall, clusterEvalQ,
##   clusterExport, clusterMap, parApply, parCapply, parLapply,
##   parLapplyLB, parRapply, parSapply, parSapplyLB
## The following objects are masked from 'package:stats':
##
##   IQR, mad, xtabs
## The following objects are masked from 'package:base':
##

```

```
## anyDuplicated, append, as.data.frame, cbind, colnames,
## do.call, duplicated, eval, evalq, Filter, Find, get, grep,
## grepl, intersect, is.unsorted, lapply, lengths, Map, mapply,
## match, mget, order, paste, pmax, pmax.int, pmin, pmin.int,
## Position, rank, rbind, Reduce, rownames, sapply, setdiff,
## sort, table, tapply, union, unique, unsplit, which, which.max,
## which.min
## Loading required package: S4Vectors
##
## Attaching package: 'S4Vectors'
## The following objects are masked from 'package:base':
##
## colMeans, colSums, expand.grid, rowMeans, rowSums
## Loading required package: IRanges
## Loading required package: GenomeInfoDb

library("GenomicAlignments")

## Loading required package: SummarizedExperiment
## Loading required package: Biobase
## Welcome to Bioconductor
##
## Vignettes contain introductory material; view with
## 'browseVignettes()'. To cite Bioconductor, see
## 'citation("Biobase")', and for packages 'citation("pkgname)".
## Loading required package: Biostrings
## Loading required package: XVector
## Loading required package: Rsamtools

library("rtracklayer")
library(parallel)
```

```
filepaths <- "/home/lucas/ISGlobal/TestSet/align_tests/params_1/"
files <- c("/home/lucas/ISGlobal/TestSet/align_tests/params_1/A7K9_14456_TTAGGC_sort.bam", '
#samples <- files[substrR(files, 8) == "sort.bam"]

aligns <- lapply(files, readGAlignmentPairs)
```

0.1 Calculating Coverage

```
cov <- lapply(aligns, coverage)
```

1 Normalization

1.1 By Number of Mapped Reads

To get the number of mapped reads in the sample we use `samtools` as follows:

```
samtools view -F 0x4 foo.sorted.bam | cut -f 1 | sort | uniq | wc -l
```

-F : tells samtools to exclude given flag. Flag 0x4 means unaligned.

```
nreads <- c()
for (file in files) {
  cmd <- paste0("/home/lucas/Programs/samtools-1.3.1/samtools view -F 0x4 ", file, " | cut -f 1 | sort | uniq | wc -l")
  nreads <- c(nreads, system(cmd, intern = TRUE))
}
```

1.2 Normalize Coverage

```
norm_cov <- c()
for (i in range(1, length(nreads))) {
  norm_cov <- c(norm_cov, cov[[i]]/as.numeric(nreads[i])*1000000)
}
```

```
for (i in range(1, length(norm_cov))) {
  con <- gsub("_sort.bam", "_cov.bed", files[i])
  export(norm_cov[[i]], con = con, format = "BED")
}
```

2 Subtractions

```
subtraction_A7K9_E5K9 <- norm_cov[[1]] - norm_cov[[2]]
export(subtraction_A7K9_E5K9, con = "/home/lucas/ISGlobal/TestSet/align_tests/params_1/A7K9_E5K9.bed")
```

3 Session Info

```
sessionInfo()
```

```

## R version 3.3.3 (2017-03-06)
## Platform: x86_64-pc-linux-gnu (64-bit)
## Running under: Ubuntu 16.04.2 LTS
##
## locale:
##  [1] LC_CTYPE=en_US.UTF-8      LC_NUMERIC=C
##  [3] LC_TIME=es_ES.UTF-8      LC_COLLATE=en_US.UTF-8
##  [5] LC_MONETARY=es_ES.UTF-8  LC_MESSAGES=en_US.UTF-8
##  [7] LC_PAPER=es_ES.UTF-8     LC_NAME=C
##  [9] LC_ADDRESS=C             LC_TELEPHONE=C
## [11] LC_MEASUREMENT=es_ES.UTF-8 LC_IDENTIFICATION=C
##
## attached base packages:
## [1] parallel stats4 stats graphics grDevices utils datasets
## [8] methods base
##
## other attached packages:
##  [1] rtracklayer_1.34.2      GenomicAlignments_1.10.1
##  [3] Rsamtools_1.26.1       Biostrings_2.42.1
##  [5] XVector_0.14.1         SummarizedExperiment_1.4.0
##  [7] Biobase_2.34.0          GenomicRanges_1.26.4
##  [9] GenomeInfoDb_1.10.3    IRanges_2.8.2
## [11] S4Vectors_0.12.2       BiocGenerics_0.20.0
## [13] BiocInstaller_1.24.0    knitr_1.15.1
##
## loaded via a namespace (and not attached):
##  [1] magrittr_1.5           zlibbioc_1.20.0      BiocParallel_1.8.1
##  [4] lattice_0.20-34       highr_0.6            stringr_1.2.0
##  [7] tools_3.3.3           grid_3.3.3          Matrix_1.2-8
## [10] bitops_1.0-6          RCurl_1.95-4.8      evaluate_0.10
## [13] stringi_1.1.3         XML_3.98-1.3

```