check library size RAPseq

Qun Li

3 Sep 2024 (12:30:09)

Contents

```
rm(list = ls())
setwd("/Users/liqun/Desktop/Projects/RAPseq/0_RAPSeq/")
```

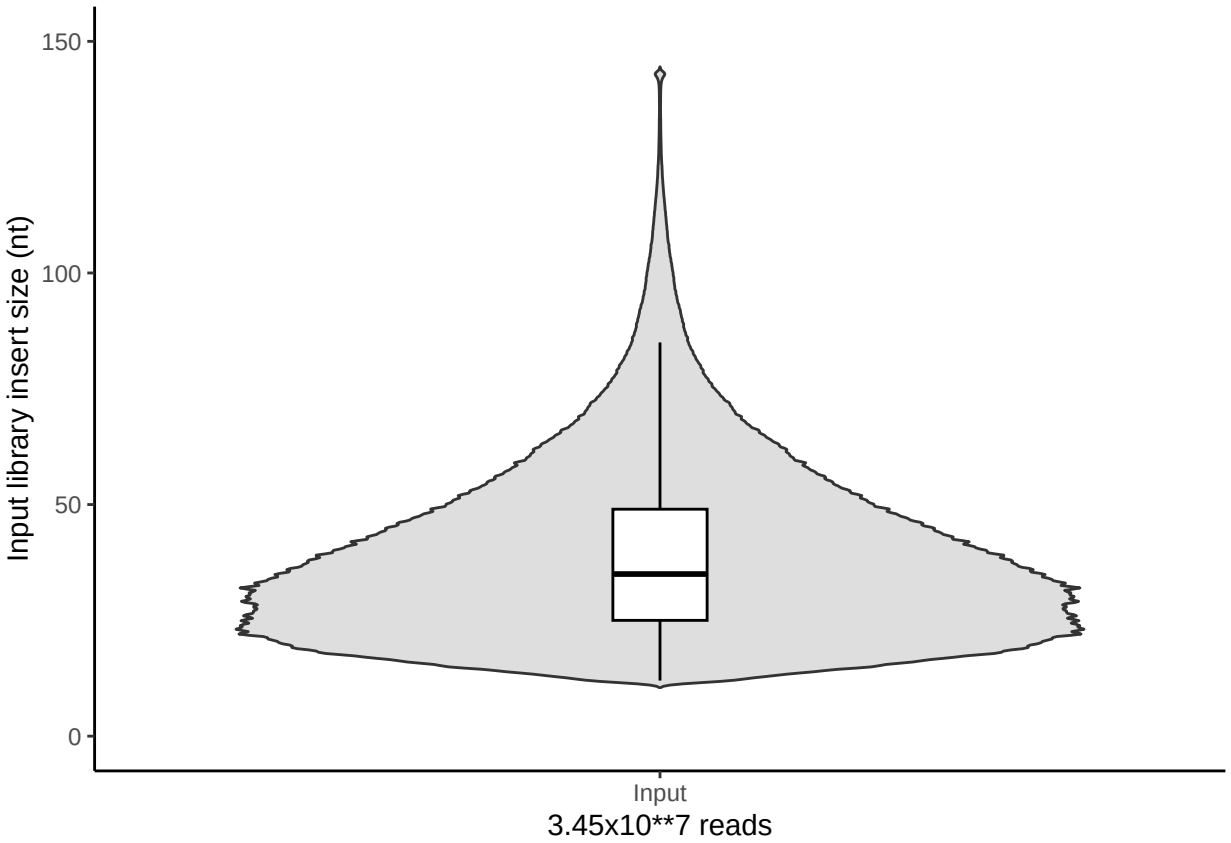
```
library(ggplot2)
library(data.table)
## Warning: package 'data.table' was built under R version 4.3.3
```

```
# 34514162 reads
read_length <- fread("../0_RAPSeq/Data/read_lengths.txt")

read_length$Name <- "Input"
colnames(read_length) <- c("length", "Name")
```

```
p <- ggplot(read_length, aes(x=Name, y=length)) +
  geom_violin(trim=FALSE, fill="grey", alpha=0.5) +
  geom_boxplot(width=0.1, fill="white", color="black", outlier.shape=NA) +
  coord_cartesian(ylim=c(0,150)) +
  theme_classic() +
  ylab("Input library insert size (nt)") +
  xlab("3.45x10**7 reads")

p
```




```
ggsave("../0_RAPSeq/inputLibrarySize.pdf", p, width = 3, height = 5)
```

```
summary(read_length$length)
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   12.00   25.00   35.00   39.43   49.00  143.00
```

```
# if your R cant load all read length files, so we use shell scripts to get read length summary informa

sort -n read_lengths.txt | awk '{
  a[NR]=$1; sum+=$1
}
END {
  mean = sum/NR
  if (NR%2==1) {
    median = a[(NR+1)/2]
  } else {
    median = (a[NR/2] + a[NR/2+1])/2
  }
  q1 = a[int(NR*0.25)]
  q3 = a[int(NR*0.75)]
  iqr = q3 - q1
  print "Mean:", mean
  print "Median:", median
  print "Q1:", q1
  print "Q3:", q3
```

```
  print "IQR:", iqr
}'
```

```
#Mean: 39.4285
#Median: 35
#Q1: 25
#Q3: 49
#IQR: 24
```