

Alignment Stats

Lucas Michel Todó, Cristina Bancells
Alfred Cortes and Juan R. Gonzalez

Barcelona Global Health Institute (ISGlobal), Campus PRBB

February 21, 2017

Contents

1	Importing Data	2
2	Plots	2
2.1	MAPQ	2
2.2	Fragment Length	4

1 Importing Data

First we import the data for every alignment:

```
mapq <- read.csv2(file = "~/ISGlobal/TestSet/align_tests/params_1/A7K9_14456_TTA
lens <- read.csv2(file = "~/ISGlobal/TestSet/align_tests/params_1/A7K9_14456_TTA

abs_lens <- abs(as.numeric(lens))
max(abs.numeric(lens))

## [1] 1057118

which(lens == 1057118)

## [1] 86703

purged_lens <- abs_lens
purged_lens[86703:86704] <- c(0,0)
```

2 Plots

2.1 MAPQ

```
x <- as.data.frame(as.numeric(mapq))
colnames(x) <- "MAPQ"
ggplot(x, aes(x = MAPQ)) +
  geom_histogram(binwidth = 1) +
  labs(x = "MAPQ", y = "Count") +
  scale_x_continuous(breaks = seq(0, 45, by = 2), limits = c(0,48)) +
  scale_y_continuous(breaks = seq(0,35000, by = 2500))

table(x > 5)

##
## FALSE TRUE
## 32237 77763
```

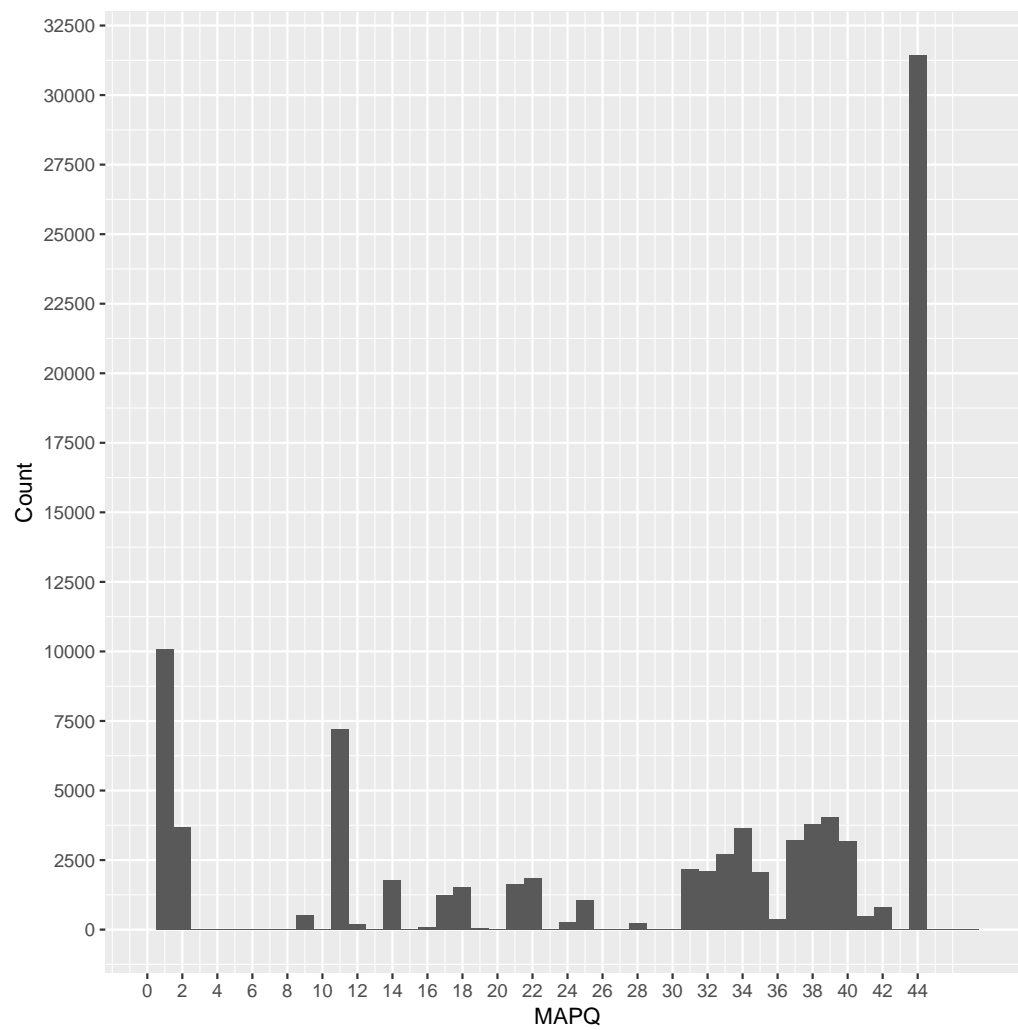


Figure 1: Histogram of read's MAPQ

2.2 Fragment Length

```
x <- as.data.frame(purged_lens[purged_lens != 0])
colnames(x) <- "len"
ggplot(x, aes(x = len)) +
  geom_histogram(bins = 100) +
  labs(x = "Lengths", y = "Count") +
  scale_x_continuous(breaks = seq(75, 550, by = 25), limits = c(75, 550))

## Warning: Removed 38 rows containing non-finite values (stat_bin).
## Warning: Removed 1 rows containing missing values (geom_bar).
```

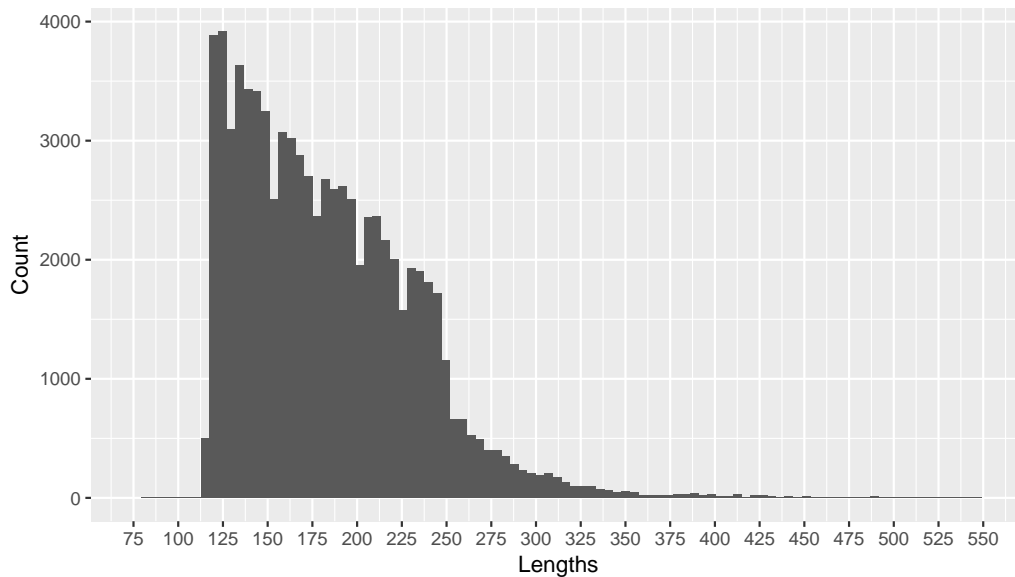


Figure 2: Histogram of fragment lengths

And some text after.