

Quality Filtering:

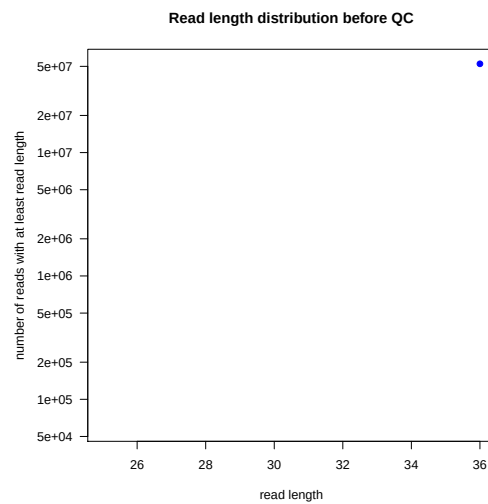


Figure 1: Read length reverse cumulative distribution before filtering

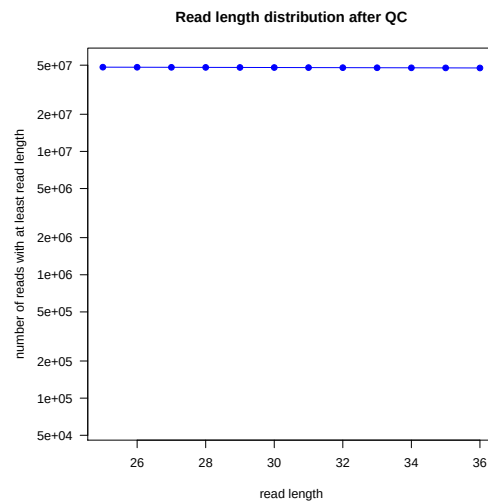


Figure 2: Read length reverse cumulative distribution after filtering

Adaptor Finding:

0.1

Chosen adaptor: GATCGGAAGAGCTCGTATGCGTCTTCTGCTTG

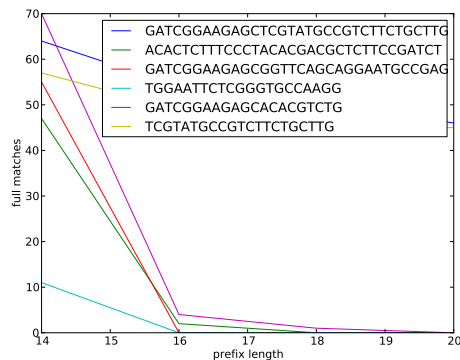


Figure 3: Plot for Adaptor Finding: Length of Adaptor versus Full Matches

Adaptor Removal and Quality Filtering:

0.2

```
# 5'/3' adaptors      : / GATCGGAAGAGCTCGTATGCCGCTCTTCTGCTTG
# 5'/3' min aln. len. : 0 / 0
# min len (1 mm)      : 7
# min len (2 mm)      : 10
# max mm (full)       : 2
# sequence filter      : on
# Wed Dec 3 12:33:50 2014: READING SEQUENCES
# Wed Dec 3 12:42:58 2014: done reading sequences (n=45849191)
# Wed Dec 3 12:42:58 2014: REMOVING ADAPTORS
# Wed Dec 3 13:30:35 2014: done removing adaptors (n=45847041)
# Wed Dec 3 13:30:35 2014: FILTERING AND OUTPUT
# Wed Dec 3 13:42:22 2014: done FILTERING AND OUTPUT
# SUMMARY:
# Number of sequences:
#
#           input sequences : 48183574 (100.0%)
#           unique sequences : 45849191 ( 95.2%)
#
# Adaptor removal:
#           full matches to 5'-adaptor : 0 ( 0.0%)
#   inexact suffix matches to 5'-adaptor : 0 ( 0.0%)
#           exact suffix matches to 5'-adaptor : 0 ( 0.0%)
#           full matches to 3'-adaptor : 5459 ( 0.0%)
#   inexact prefix matches to 3'-adaptor : 55068 ( 0.1%)
#           exact prefix matches to 3'-adaptor : 12870899 ( 26.7%)
#
#           total matches to 5'-adaptor : 0 ( 0.0%)
#           total matches to 3'-adaptor : 12931426 ( 26.8%)
#
#           unique sequences (w/o adaptors) : 45847041 ( 95.2%)
#
# Sequence filtering:
#           5'-adaptor alignment too short : 0 ( 0.0%)
```

```

#           3'-adaptor alignment too short :           0 (  0.0%)
#                                           too short :          6495 (  0.0%)
#           too many non-base characters :       595452 (  1.2%)
#                                           low entropy :       169488 (  0.4%)
#
# Final sequences:
#                                           passed :   47412139 ( 98.4%)

```

```

Top 10 tags before adaptor removal:
>seq0    52289
.....
>seq1    35006
A.....
>seq2    23227
C.....
>seq3    19656
G.....
>seq4    15971
T.....
>seq5    14215
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
>seq6    5232
CT.....
>seq7    4836
CA.....
>seq8    4744
AA.....
>seq9    4526
AT.....

```

Mapping:

Number of mapped reads

```

Total reads mapped:           36430732
Total unmapped reads:        10011848
Total overmapped reads (>100 hits):    969558

```

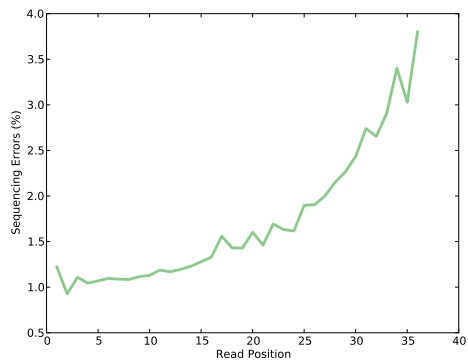


Figure 4: Fraction of all reads that had an error at read position

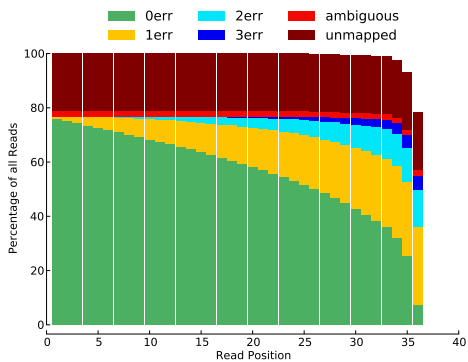


Figure 5: Fraction of all reads that have a certain mapping error up to read position

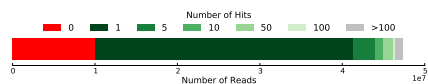


Figure 6: Number of reads that have a certain number of hits

Fragment Size Estimation:

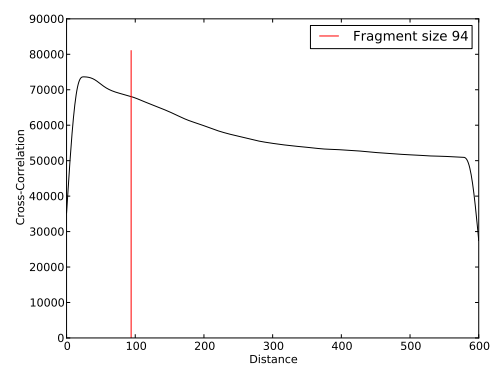


Figure 7: Fragment size correlation plot