

Quality Filtering:

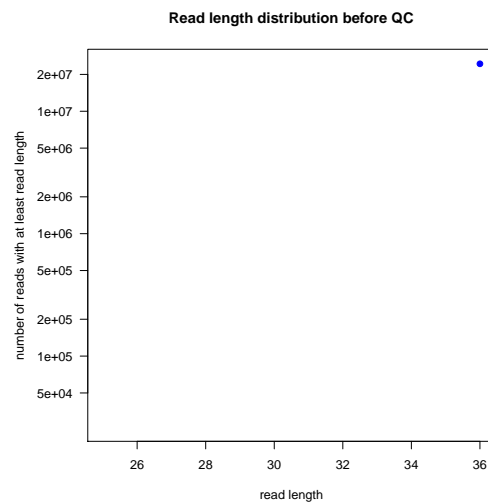


Figure 1: Read length reverse cumulative distribution before filtering

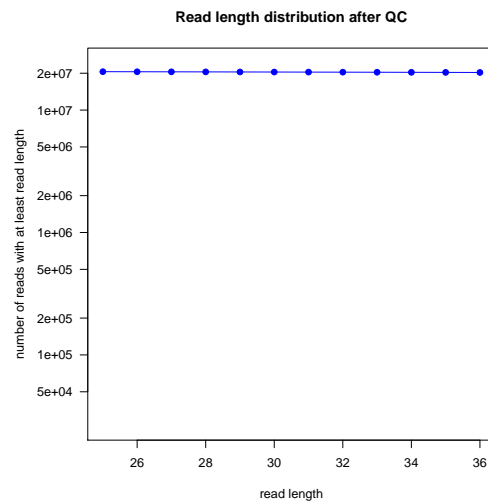


Figure 2: Read length reverse cumulative distribution after filtering

Adaptor Finding:

0.1

Chosen adaptor: GATCGGAAGAGCTCGTATGCGCTCTCTGCTTG

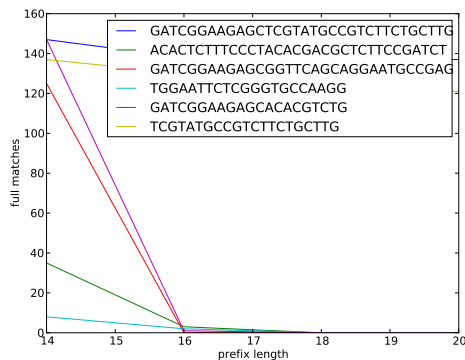


Figure 3: Plot for Adaptor Finding: Length of Adaptor versus Full Matches

Adaptor Removal and Quality Filtering:

0.2

```
# 5'/3' adaptors      : / GATCGGAAGAGCTCGTATGCCGCTTCTGCTTG
# 5'/3' min aln. len. : 0 / 0
# min len (1 mm)      : 7
# min len (2 mm)      : 10
# max mm (full)       : 2
# sequence filter     : on
# Sun Feb  1 20:55:55 2015: READING SEQUENCES
# Sun Feb  1 21:00:02 2015: done reading sequences (n=17654851)
# Sun Feb  1 21:00:02 2015: REMOVING ADAPTORS
# Sun Feb  1 21:19:46 2015: done removing adaptors (n=17652612)
# Sun Feb  1 21:19:46 2015: FILTERING AND OUTPUT
# Sun Feb  1 21:26:09 2015: done FILTERING AND OUTPUT
# SUMMARY:
# Number of sequences:
#
#           input sequences : 20593022 (100.0%)
#           unique sequences : 17654851 ( 85.7%)
#
# 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 : 9851 ( 0.0%)
#   inexact prefix matches to 3'-adaptor : 23371 ( 0.1%)
#           exact prefix matches to 3'-adaptor : 5891300 ( 28.6%)
#
#           total matches to 5'-adaptor : 0 ( 0.0%)
#           total matches to 3'-adaptor : 5924522 ( 28.8%)
#
#           unique sequences (w/o adaptors) : 17652612 ( 85.7%)
#
# Sequence filtering:
#           5'-adaptor alignment too short : 0 ( 0.0%)
```

```
#          3'-adaptor alignment too short :          0 (  0.0%)
#                                     too short :      10640 (  0.1%)
#          too many non-base characters :          0 (  0.0%)
#                                     low entropy :     53726 (  0.3%)
#
# Final sequences:
#                                     passed :    20528656 ( 99.7%)
```

```
Top 10 tags before adaptor removal:
>seq0    7434
GATCGGAAGAGCTCGTATGCGCTCTTCTGCTTGAAA
>seq1     618
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
>seq2     387
TGGAATGGAATGGAATGGAATGGAATGGAATGGAAT
>seq3     363
GAATGGAATGGAATGGAATGGAATGGAATGGAATGG
>seq4     359
GGAATGGAATGGAATGGAATGGAATGGAATGGAATG
>seq5     337
GTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGG
>seq6     316
GGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAG
>seq7     303
CATTCATTCCATTCCATTCCATTCCATTCCATTCC
>seq8     296
ATTCCATTCCATTCCATTCCATTCCATTCCATTCCA
>seq9     281
ACACACACACACACACACACACACACACACACAC
```

Mapping:

Number of mapped reads

```
Total reads mapped:      19148937
Total unmapped reads:    677998
Total overmapped reads (>100 hits): 701721
```

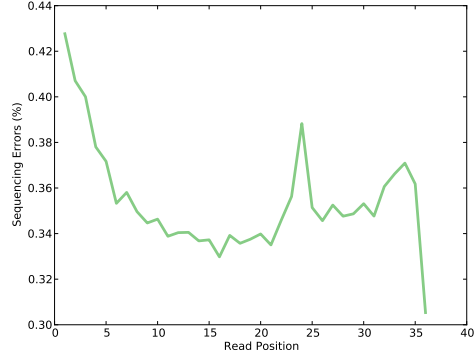


Figure 4: Fraction of all reads that had an error at read position

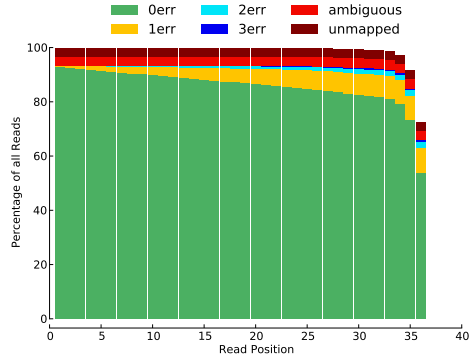


Figure 5: Fraction of all reads that have a certain mapping error up to read position

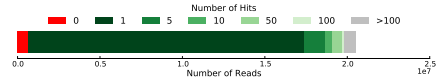


Figure 6: Number of reads that have a certain number of hits

Fragment Size Estimation:

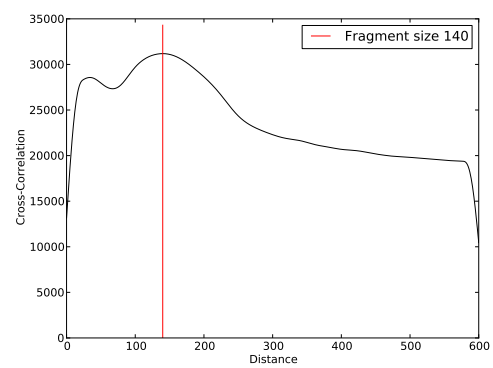


Figure 7: Fragment size correlation plot