

Quality Filtering:

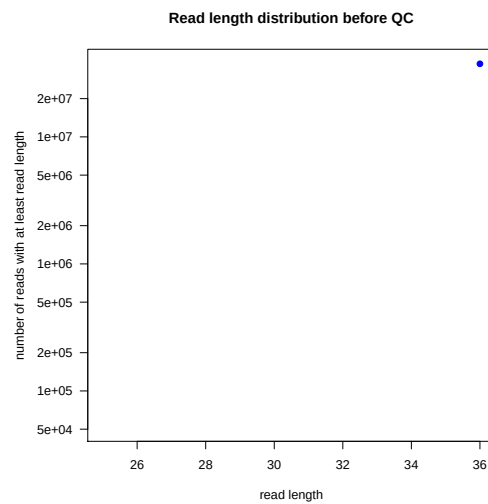


Figure 1: Read length reverse cumulative distribution before filtering

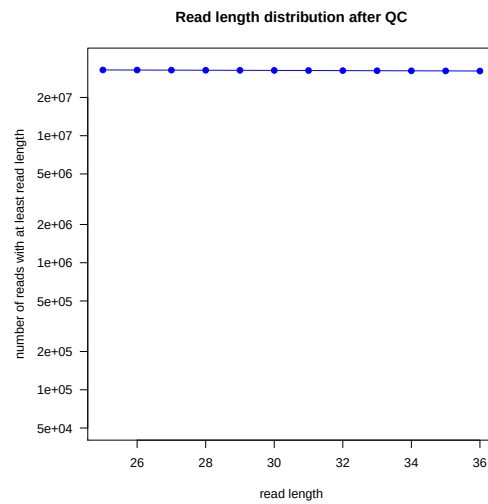


Figure 2: Read length reverse cumulative distribution after filtering

Adaptor Finding:

0.1

Chosen adaptor: GATCGGAAGAGCGGTTCAGCAGGAATGCCGAG

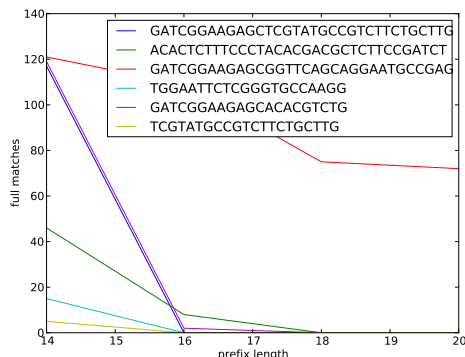


Figure 3: Plot for Adaptor Finding: Length of Adaptor versus Full Matches

Adaptor Removal and Quality Filtering:

0.2

```
# 5'/3' adaptors      : / GATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
# 5'/3' min aln. len. : 0 / 0
# min len (1 mm)      : 7
# min len (2 mm)      : 10
# max mm (full)       : 2
# sequence filter     : on
# Thu Dec 4 19:45:25 2014: READING SEQUENCES
# Thu Dec 4 19:51:40 2014: done reading sequences (n=30534880)
# Thu Dec 4 19:51:40 2014: REMOVING ADAPTORS
# Thu Dec 4 20:24:36 2014: done removing adaptors (n=30532283)
# Thu Dec 4 20:24:36 2014: FILTERING AND OUTPUT
# Thu Dec 4 20:32:22 2014: done FILTERING AND OUTPUT
# SUMMARY:
# Number of sequences:
#
#           input sequences : 32932491 (100.0%)
#           unique sequences : 30534880 ( 92.7%)
#
# 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 : 9311 ( 0.0%)
#   inexact prefix matches to 3'-adaptor : 45306 ( 0.1%)
#     exact prefix matches to 3'-adaptor : 10039266 ( 30.5%)
#
#           total matches to 5'-adaptor : 0 ( 0.0%)
#           total matches to 3'-adaptor : 10093883 ( 30.7%)
#
#           unique sequences (w/o adaptors) : 30532283 ( 92.7%)
#
# Sequence filtering:
#           5'-adaptor alignment too short : 0 ( 0.0%)
```

```
#          3'-adaptor alignment too short :          0 (  0.0%)
#                                     too short :       10118 (  0.0%)
#          too many non-base characters :    183049 (  0.6%)
#                                     low entropy :     82367 (  0.3%)
#
# Final sequences:
#                                     passed :   32656957 ( 99.2%)
```

Top 10 tags before adaptor removal:

```
>seq0    7093
GATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCG
>seq1    4354
CGGTTCAGCAGGAATGCCGAGATCGGAAGAGCGGT
>seq2    4053
CTAACCCTAACCCTAACCCTAACCCTAACCCTAACC
>seq3    4048
GGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAG
>seq4    2940
GGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTA
>seq5    2821
GTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGG
>seq6    2796
AGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTT
>seq7    2572
TAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGT
>seq8    2404
CGGAAGAGCGGTTCAGCAGGAATGCCGAGATCGGAA
>seq9    2315
CCTAACCCTAACCCTAACCCTAACCCTAACCCTAAC
```

Mapping:

Number of mapped reads

```
Total reads mapped:      29914587
Total unmapped reads:    1513120
Total overmapped reads (>100 hits): 1229250
```

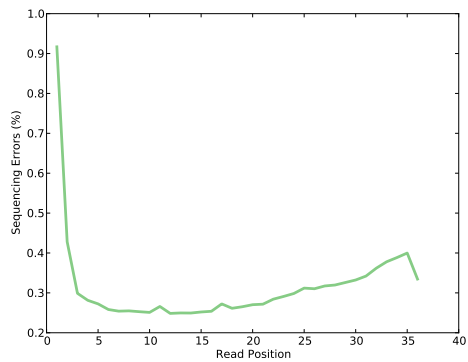


Figure 4: Fraction of all reads that had an error at read position

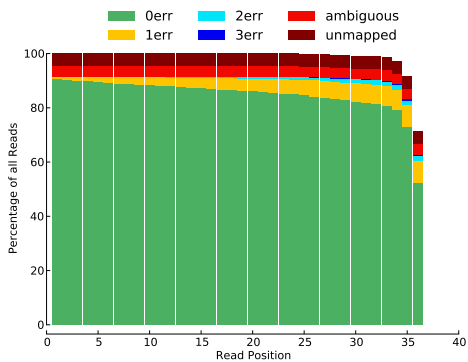


Figure 5: Fraction of all reads that have a certain mapping error up to read position

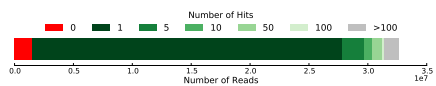


Figure 6: Number of reads that have a certain number of hits

Fragment Size Estimation:

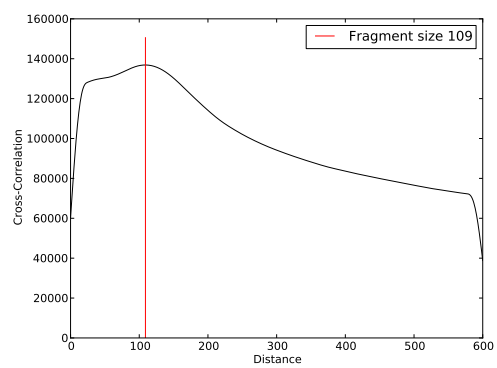


Figure 7: Fragment size correlation plot