

Quality Filtering:

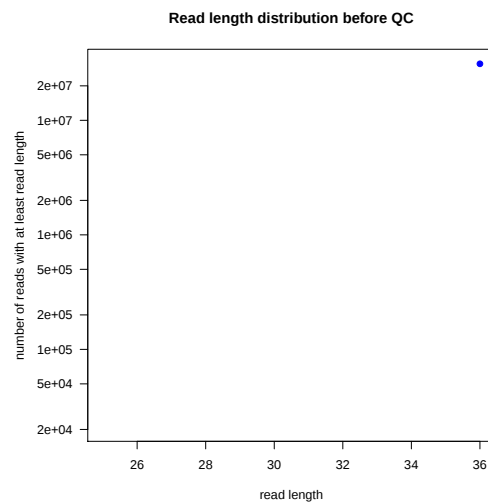


Figure 1: Read length reverse cumulative distribution before filtering

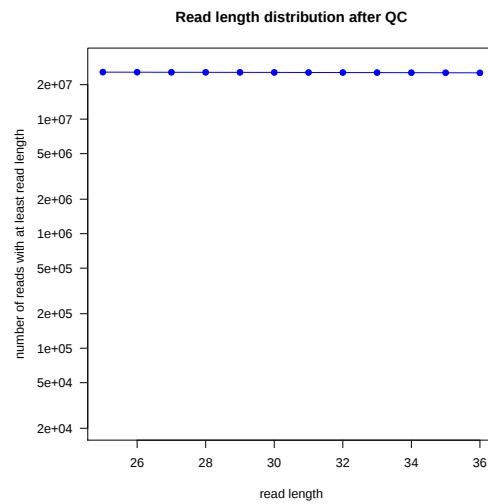


Figure 2: Read length reverse cumulative distribution after filtering

Adaptor Finding:

0.1

Chosen adaptor: GATCGGAAGAGCGGTTCAGCAGGAATGCCGAG

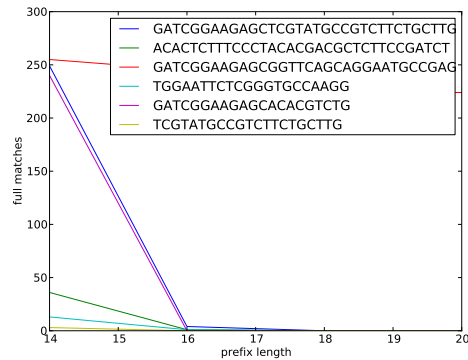


Figure 3: Plot for Adaptor Finding: Length of Adaptor versus Full Matches

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 19:33:00 2014: READING SEQUENCES
# Thu Dec 4 19:37:49 2014: done reading sequences (n=23050869)
# Thu Dec 4 19:37:49 2014: REMOVING ADAPTORS
# Thu Dec 4 20:02:32 2014: done removing adaptors (n=23048650)
# Thu Dec 4 20:02:32 2014: FILTERING AND OUTPUT
# Thu Dec 4 20:08:40 2014: done FILTERING AND OUTPUT
# SUMMARY:
# Number of sequences:
#
#           input sequences : 25734688 (100.0%)
#           unique sequences : 23050869 ( 89.6%)
#
# 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 : 24826 ( 0.1%)
#   inexact prefix matches to 3'-adaptor : 31710 ( 0.1%)
#           exact prefix matches to 3'-adaptor : 7691238 ( 29.9%)
#
#           total matches to 5'-adaptor : 0 ( 0.0%)
#           total matches to 3'-adaptor : 7747774 ( 30.1%)
#
#           unique sequences (w/o adaptors) : 23048650 ( 89.6%)
#
# Sequence filtering:
#           5'-adaptor alignment too short : 0 ( 0.0%)
```

```

#           3'-adaptor alignment too short :           0 (   0.0%)
#                                           too short :       25389 (   0.1%)
#           too many non-base characters :       41404 (   0.2%)
#                                           low entropy :       71315 (   0.3%)
#
# Final sequences:
#                                           passed :   25596580 ( 99.5%)

```

Top 10 tags before adaptor removal:

```

>seq0    22227
GATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCG
>seq1    2787
GGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAG
>seq2    2382
CTAACCCTAACCCTAACCCTAACCCTAACCCTAAC
>seq3    2183
AGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTT
>seq4    2061
GTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGG
>seq5    1993
CGGTTCAGCAGGAATGCCGAGATCGGAAGAGCGGT
>seq6    1737
GGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTA
>seq7    1339
CCTAACCCTAACCCTAACCCTAACCCTAACCCTAAC
>seq8    1047
TAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGT
>seq9     971
GATCGGAAGAGCGGTTCAGCAGGAATGCCGAGATCG

```

Mapping:

Number of mapped reads

```

Total reads mapped:           24269988
Total unmapped reads:        413947
Total overmapped reads (>100 hits):    912645

```

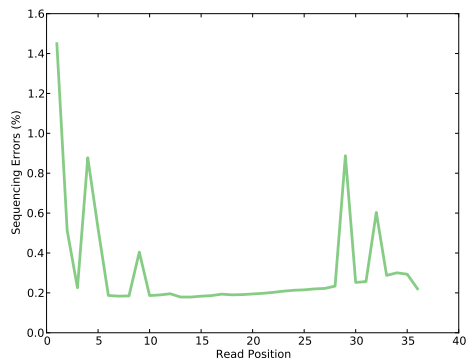


Figure 4: Fraction of all reads that had an error at read position

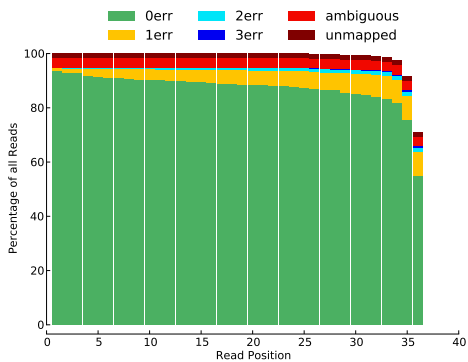


Figure 5: Fraction of all reads that have a certain mapping error up to read position

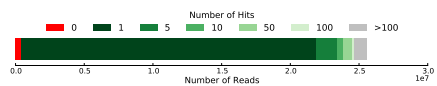


Figure 6: Number of reads that have a certain number of hits

Fragment Size Estimation:

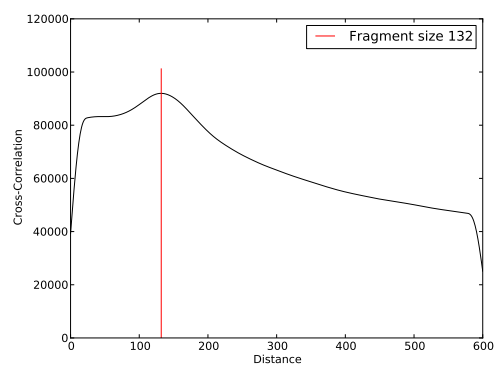


Figure 7: Fragment size correlation plot