

## Quality Filtering:

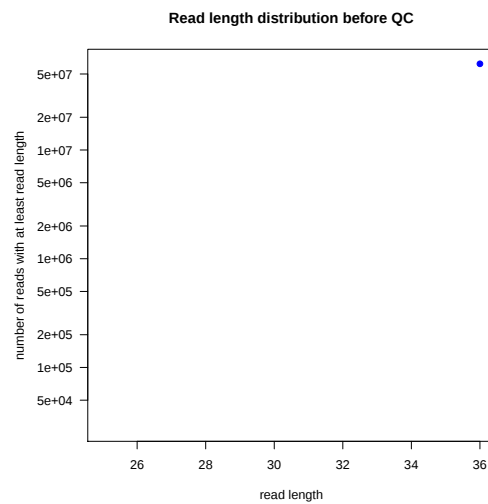


Figure 1: Read length reverse cumulative distribution before filtering

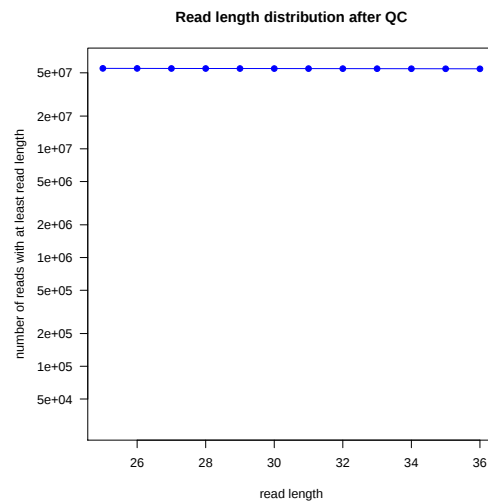


Figure 2: Read length reverse cumulative distribution after filtering

## Adaptor Finding:

0.1

Chosen adaptor: GATCGGAAGAGCGGTTCAGCAGGAATGCCGAG

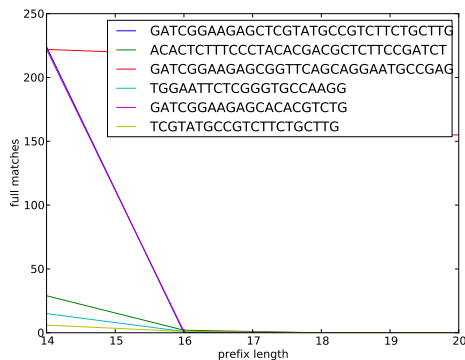


Figure 3: Plot for Adaptor Finding: Length of Adaptor versus Full Matches

## 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
# Wed Dec 3 22:14:23 2014: READING SEQUENCES
# Wed Dec 3 22:23:43 2014: done reading sequences (n=24593674)
# Wed Dec 3 22:23:43 2014: REMOVING ADAPTORS
# Wed Dec 3 22:49:39 2014: done removing adaptors (n=24591704)
# Wed Dec 3 22:49:39 2014: FILTERING AND OUTPUT
# Wed Dec 3 22:56:07 2014: done FILTERING AND OUTPUT
# SUMMARY:
# Number of sequences:
#
#           input sequences : 54936648 (100.0%)
#           unique sequences : 24593674 ( 44.8%)
#
# 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 : 31160 ( 0.1%)
#   inexact prefix matches to 3'-adaptor : 93582 ( 0.2%)
#     exact prefix matches to 3'-adaptor : 16671364 ( 30.3%)
#
#           total matches to 5'-adaptor : 0 ( 0.0%)
#           total matches to 3'-adaptor : 16796106 ( 30.6%)
#
#           unique sequences (w/o adaptors) : 24591704 ( 44.8%)
#
# Sequence filtering:
#           5'-adaptor alignment too short : 0 ( 0.0%)
```

```
#          3'-adaptor alignment too short :          0 (  0.0%)
#                                     too short :       33148 (  0.1%)
#          too many non-base characters :       41782 (  0.1%)
#                                     low entropy :    156340 (  0.3%)
#
# Final sequences:
#                                     passed :   54705378 ( 99.6%)
```

Top 10 tags before adaptor removal:

```
>seq0    26516
GATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCG
>seq1    16644
CGGTTCAGCAGGAATGCCGAGATCGGAAGAGCGGT
>seq2    8714
CGGAAGAGCGGTTCAGCAGGAATGCCGAGATCGGAA
>seq3    2786
GATCGGAAGAGCGGTTCAGCAGGAATGCCGAGATCG
>seq4    2306
CGGAAGAGCGGTTCAGCAGGAATGCCGAGACCGATC
>seq5    1144
TGTTGCGCTGTAATATTGAACGTAGGTGCGATAAAT
>seq6    1014
CACACACACACACACACACACACACACACACACACA
>seq7    962
TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG
>seq8    958
GGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAG
>seq9    926
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
```

## Mapping:

### Number of mapped reads

```
Total reads mapped:          50697602
Total unmapped reads:        1922380
Total overmapped reads (>100 hits): 2085396
```

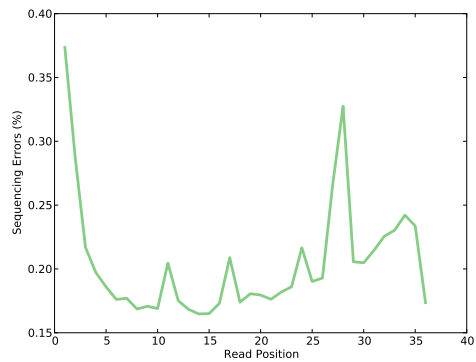


Figure 4: Fraction of all reads that had an error at read position

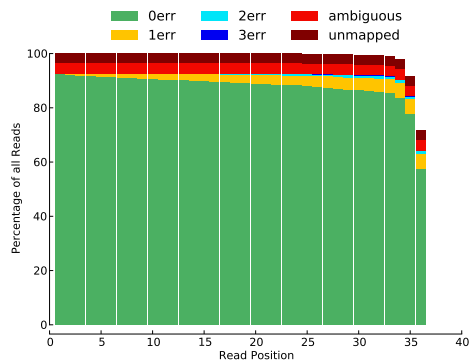


Figure 5: Fraction of all reads that have a certain mapping error up to read position

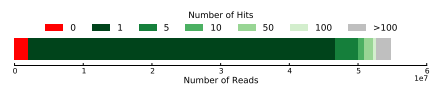


Figure 6: Number of reads that have a certain number of hits

## Fragment Size Estimation:

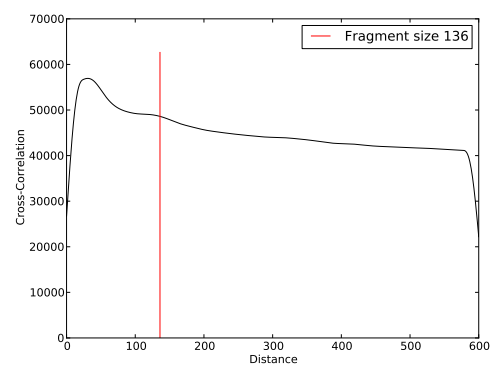


Figure 7: Fragment size correlation plot