

Quality Filtering:

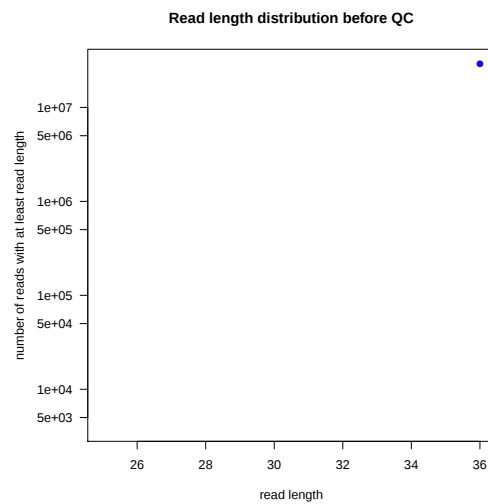


Figure 1: Read length reverse cumulative distribution before filtering

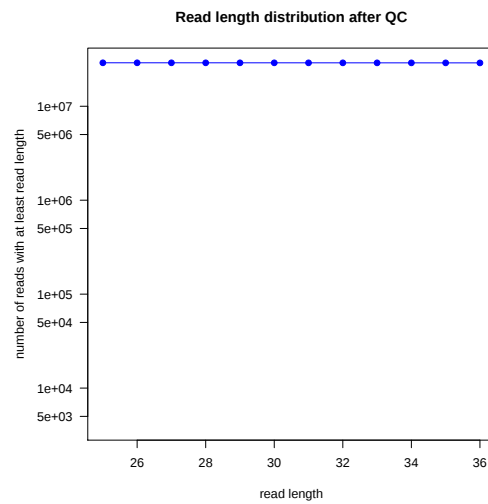


Figure 2: Read length reverse cumulative distribution after filtering

Adaptor Finding:

0.1

Chosen adaptor: TCGTATGCCGTCCTCTGCTTG

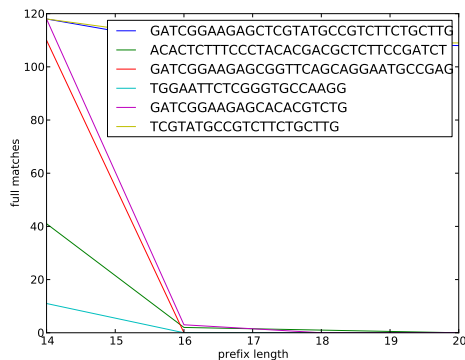


Figure 3: Plot for Adaptor Finding: Length of Adaptor versus Full Matches

Adaptor Removal and Quality Filtering:

0.2

```
# 5'/3' adaptors      : / TCGTATGCCGTCTTCTGCTTG
# 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 22:30:52 2015: READING SEQUENCES
# Sun Feb 1 22:36:42 2015: done reading sequences (n=26296487)
# Sun Feb 1 22:36:42 2015: REMOVING ADAPTORS
# Sun Feb 1 23:11:11 2015: done removing adaptors (n=26295964)
# Sun Feb 1 23:11:11 2015: FILTERING AND OUTPUT
# Sun Feb 1 23:20:30 2015: done FILTERING AND OUTPUT
# SUMMARY:
# Number of sequences:
#
#           input sequences : 29019758 (100.0%)
#           unique sequences : 26296487 ( 90.6%)
#
# 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 : 13012 ( 0.0%)
#   inexact prefix matches to 3'-adaptor : 25444 ( 0.1%)
#           exact prefix matches to 3'-adaptor : 10112298 ( 34.8%)
#
#           total matches to 5'-adaptor : 0 ( 0.0%)
#           total matches to 3'-adaptor : 10150754 ( 35.0%)
#
#           unique sequences (w/o adaptors) : 26295964 ( 90.6%)
#
# Sequence filtering:
#           5'-adaptor alignment too short : 0 ( 0.0%)
```

```
#          3'-adaptor alignment too short :          0 (  0.0%)
#                                     too short :       12909 (  0.0%)
#          too many non-base characters :          0 (  0.0%)
#                                     low entropy :    138286 (  0.5%)
#
# Final sequences:
#                                     passed :   28868563 ( 99.5%)
```

```
Top 10 tags before adaptor removal:
>seq0    12110
GATCGGAAGAGCTCGTATGCGGTCTTCTGCTTGAAA
>seq1    2474
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
>seq2    740
CTTCCTTGTTGTGTGTATTCAACTCACAGAGTTG
>seq3    706
GGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAG
>seq4    680
CATTCCATTCCATTCCATTCCATTCCATTCCATTCC
>seq5    675
CTAACCCTAACCCTAACCCTAACCCTAACCCTAACC
>seq6    617
CATTGACCTCAAAGCGGCTGAAATCTCCACTTGCAA
>seq7    614
CTGCTCTATGAAAAGAAAGGTTAAACTCTGTGAGTT
>seq8    606
CACACACACACACACACACACACACACACACACACA
>seq9    573
CCTAACCCTAACCCTAACCCTAACCCTAACCCTAAC
```

Mapping:

Number of mapped reads

```
Total reads mapped:      26375165
Total unmapped reads:    1371275
Total overmapped reads (>100 hits): 1122123
```

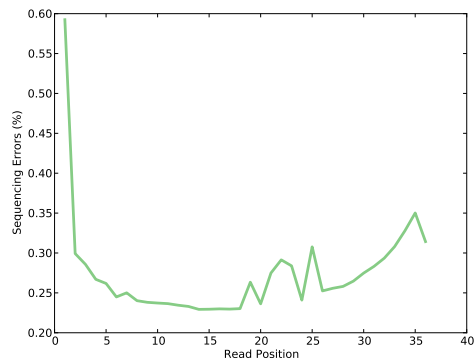


Figure 4: Fraction of all reads that had an error at read position

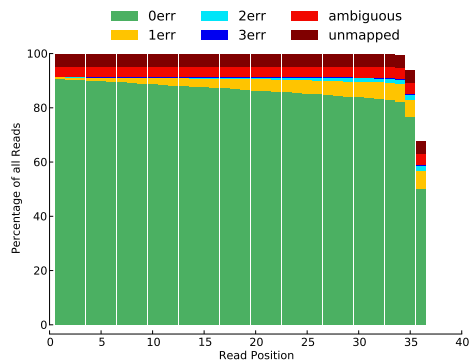


Figure 5: Fraction of all reads that have a certain mapping error up to read position

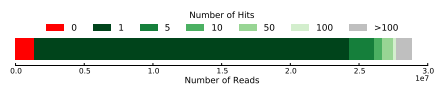


Figure 6: Number of reads that have a certain number of hits

Fragment Size Estimation:

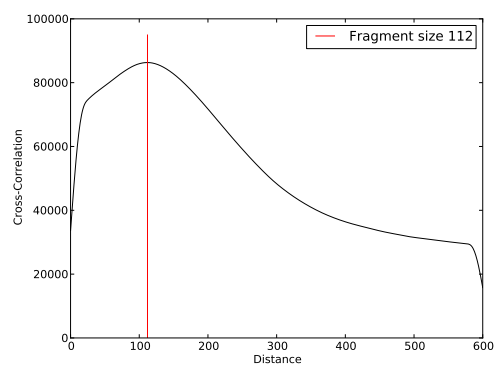


Figure 7: Fragment size correlation plot