

Background: Quality Filtering:

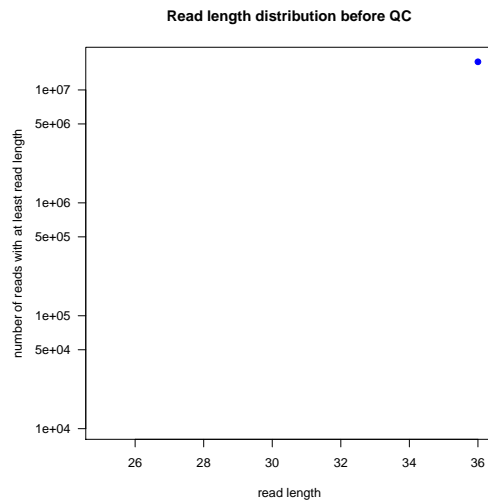


Figure 1: Read length reverse cumulative distribution before filtering

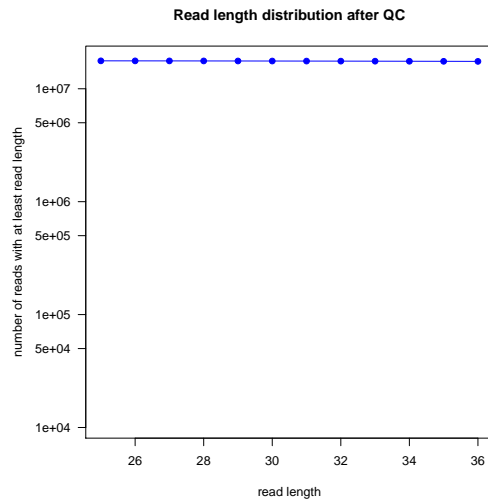


Figure 2: Read length reverse cumulative distribution after filtering

Adaptor Finding:

0.1

Chosen adaptor: `ACACTCTTTCCCTACACGACGCTCTTCCGATCT`

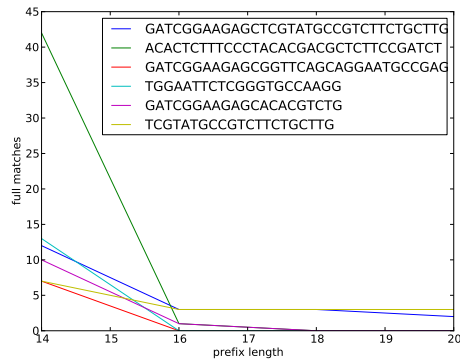


Figure 3: Plot for Adaptor Finding: Length of Adaptor versus Full Matches

Adaptor Removal and Quality Filtering:

0.2

```

# 5'/3' adaptors      : / ACACTCTTTCCTACACGACGCTCTCCGATCT
# 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 Nov 26 19:09:35 2014: READING SEQUENCES
# Wed Nov 26 19:12:42 2014: done reading sequences (n=16592758)
# Wed Nov 26 19:12:42 2014: REMOVING ADAPTORS
# Wed Nov 26 19:28:57 2014: done removing adaptors (n=16591750)
# Wed Nov 26 19:28:57 2014: FILTERING AND OUTPUT
# Wed Nov 26 19:33:14 2014: done FILTERING AND OUTPUT
# SUMMARY:
# Number of sequences:
#
#           input sequences : 17676727 (100.0%)
#           unique sequences : 16592758 ( 93.9%)
#
# 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 : 0 ( 0.0%)
#   inexact prefix matches to 3'-adaptor : 60012 ( 0.3%)
#     exact prefix matches to 3'-adaptor : 6171683 ( 34.9%)
#
#           total matches to 5'-adaptor : 0 ( 0.0%)
#           total matches to 3'-adaptor : 6231695 ( 35.3%)
#
#           unique sequences (w/o adaptors) : 16591750 ( 93.9%)
#
# Sequence filtering:
#           5'-adaptor alignment too short : 0 ( 0.0%)

```

```

#           3'-adaptor alignment too short :           0 (  0.0%)
#                                           too short :           1 (  0.0%)
#           too many non-base characters :           0 (  0.0%)
#                                           low entropy :       42044 (  0.2%)
#
# Final sequences:
#                                           passed : 17634682 ( 99.8%)

```

Top 10 tags before adaptor removal:

```

>seq0  510
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
>seq1  502
TGGAATGGAATGGAATGGAATGGAATGGAATGGAAT
>seq2  428
GAATGGAATGGAATGGAATGGAATGGAATGGAATGG
>seq3  420
GGAATGGAATGGAATGGAATGGAATGGAATGGAATG
>seq4  396
ATTCATTCCATTCCATTCCATTCCATTCCATTCCA
>seq5  359
GGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAG
>seq6  350
AATGGAATGGAATGGAATGGAATGGAATGGAATGGA
>seq7  341
CATTCCATTCCATTCCATTCCATTCCATTCCATTCC
>seq8  336
ATGGAATGGAATGGAATGGAATGGAATGGAATGGAA
>seq9  303
TAACCCTAACCCTAACCCTAACCCTAACCCTAACCC

```

Mapping:

Number of mapped reads

```

Total reads mapped:           16557795
Total unmapped reads:        470842
Total overmapped reads (>100 hits): 606045

```

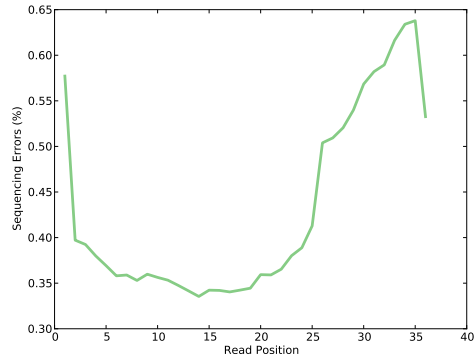


Figure 4: Fraction of all reads that had an error at read position

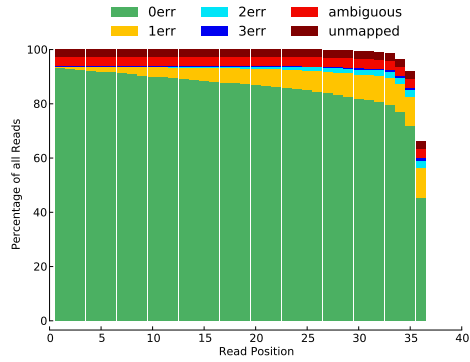


Figure 5: Fraction of all reads that have a certain mapping error up to read position

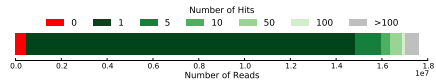


Figure 6: Number of reads that have a certain number of hits

Fragment Size Estimation:

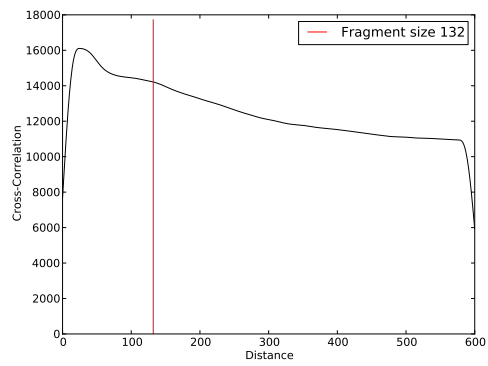


Figure 7: Fragment size correlation plot