

Quality Filtering:

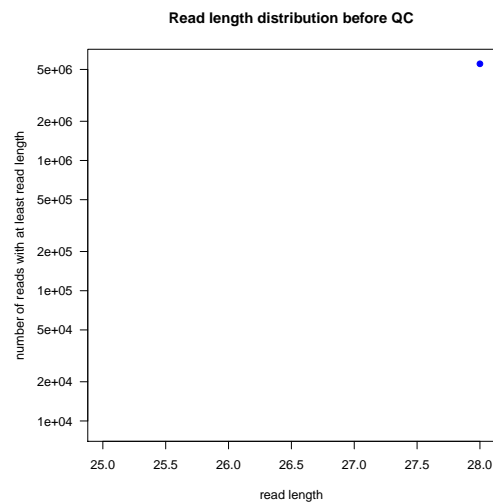


Figure 1: Read length reverse cumulative distribution before filtering

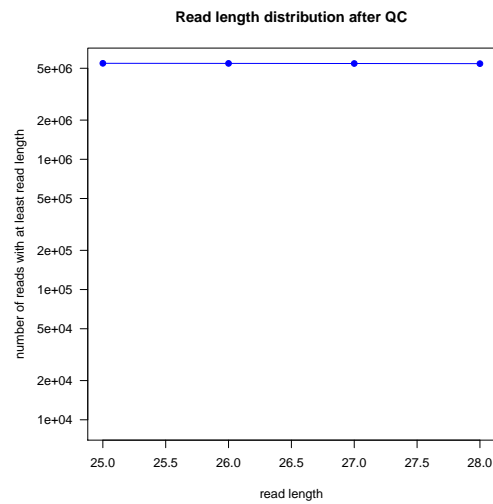


Figure 2: Read length reverse cumulative distribution after filtering

Adaptor Finding:

0.1

Chosen adaptor: GATCGGAAGAGCTCGTATGCGGTCTTCTGCTTG

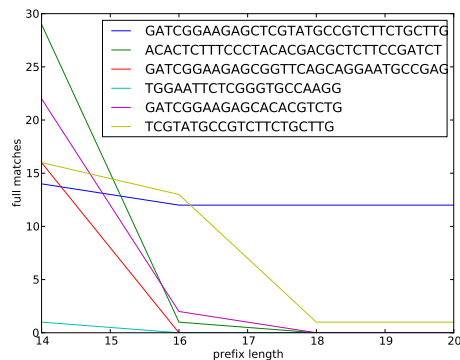


Figure 3: Plot for Adaptor Finding: Length of Adaptor versus Full Matches

Adaptor Removal and Quality Filtering:

0.2

```
# 5'/3' adaptors      : / GATCGGAAGAGCTCGTATGCCGTCTTCTGCTTG
# 5'/3' min aln. len. : 0 / 0
# min len (1 mm)      : 7
# min len (2 mm)      : 10
# max mm (full)       : 2
# sequence filter      : on
# Mon Feb  2 02:37:48 2015: READING SEQUENCES
# Mon Feb  2 02:38:53 2015: done reading sequences (n=5157292)
# Mon Feb  2 02:38:53 2015: REMOVING ADAPTORS
# Mon Feb  2 02:43:50 2015: done removing adaptors (n=5157143)
# Mon Feb  2 02:43:50 2015: FILTERING AND OUTPUT
# Mon Feb  2 02:45:24 2015: done FILTERING AND OUTPUT
# SUMMARY:
# Number of sequences:
#
#           input sequences : 5459026 (100.0%)
#           unique sequences : 5157292 ( 94.5%)
#
# 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 : 6191 ( 0.1%)
#     exact prefix matches to 3'-adaptor : 1540855 ( 28.2%)
#
#           total matches to 5'-adaptor : 0 ( 0.0%)
#           total matches to 3'-adaptor : 1547046 ( 28.3%)
#
#           unique sequences (w/o adaptors) : 5157143 ( 94.5%)
#
# Sequence filtering:
#           5'-adaptor alignment too short : 0 ( 0.0%)
```

```
#          3'-adaptor alignment too short :          0 (  0.0%)
#                                     too short :          316 (  0.0%)
#          too many non-base characters :          0 (  0.0%)
#                                     low entropy :       30852 (  0.6%)
#
# Final sequences:
#                                     passed :   5427858 ( 99.4%)
```

Top 10 tags before adaptor removal:

```
>seq0    2153
AAAAAAAAAAAAAAAAAAAAAAAAAAAAA
>seq1    255
ACACACACACACACACACACACACAC
>seq2    240
GATCGGAAGAGCTCGTATGCOGTCCTTCT
>seq3    187
CACACACACACACACACACACACACA
>seq4    173
TGGAATGGAATGGAATGGAATGGAATGG
>seq5    167
GAATGGAATGGAATGGAATGGAATGGAA
>seq6    158
AATGGAATGGAATGGAATGGAATGGAAT
>seq7    153
GGAATGGAATGGAATGGAATGGAATGGA
>seq8    116
ATGGAATGGAATGGAATGGAATGGAATG
>seq9    113
GGTTAGGGTTAGGGTTAGGGTTAGGGTT
```

Mapping:

Number of mapped reads

```
Total reads mapped:      4855998
Total unmapped reads:    312674
Total overmapped reads (>100 hits): 259186
```

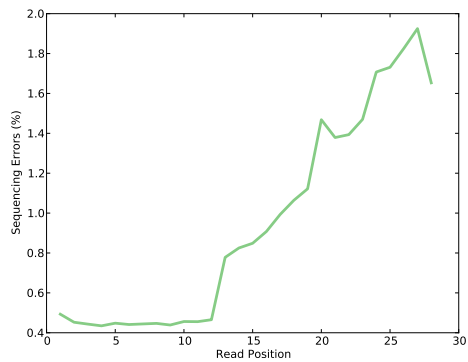


Figure 4: Fraction of all reads that had an error at read position

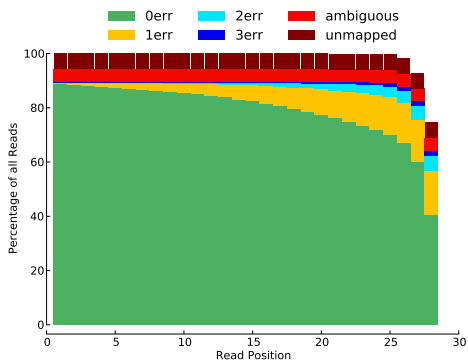


Figure 5: Fraction of all reads that have a certain mapping error up to read position

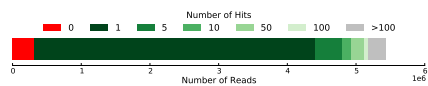


Figure 6: Number of reads that have a certain number of hits

Fragment Size Estimation:

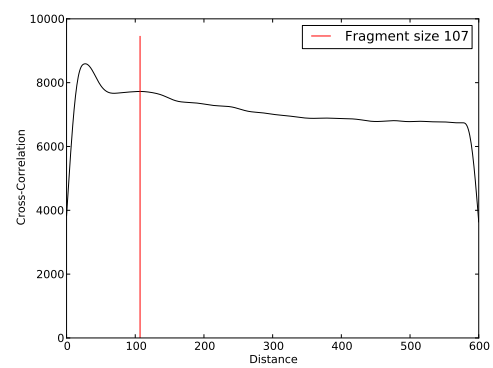


Figure 7: Fragment size correlation plot