

Background: Quality Filtering:

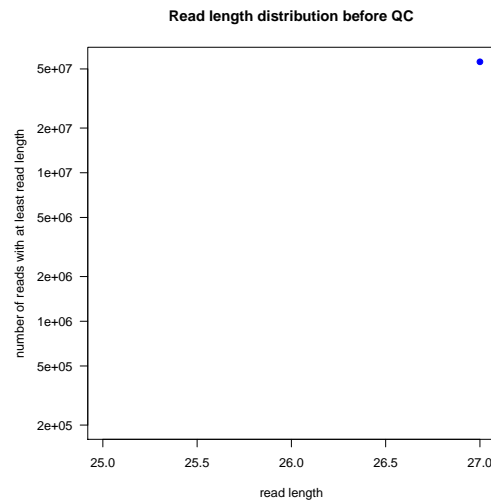


Figure 1: Read length reverse cumulative distribution before filtering

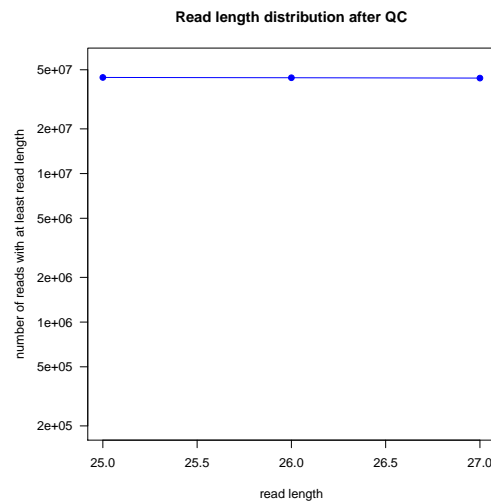


Figure 2: Read length reverse cumulative distribution after filtering

Adaptor Finding:

0.1

Chosen adaptor: GATCGGAAGAGCTCGTATGCGGTCTTCTGCTTG

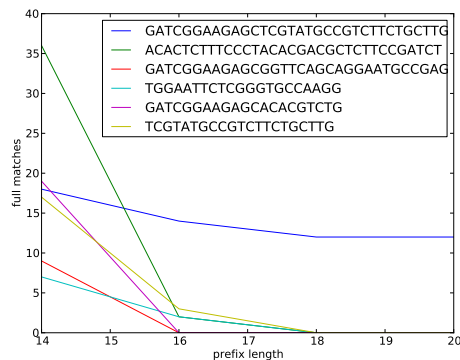


Figure 3: Plot for Adaptor Finding: Length of Adaptor versus Full Matches

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
# Sat Nov 29 17:48:04 2014: READING SEQUENCES
# Sat Nov 29 17:56:23 2014: done reading sequences (n=39326692)
# Sat Nov 29 17:56:23 2014: REMOVING ADAPTORS
# Sat Nov 29 18:30:06 2014: done removing adaptors (n=39321624)
# Sat Nov 29 18:30:06 2014: FILTERING AND OUTPUT
# Sat Nov 29 18:38:37 2014: done FILTERING AND OUTPUT
# SUMMARY:
# Number of sequences:
#
#           input sequences : 44385915 (100.0%)
#           unique sequences : 39326692 ( 88.6%)
#
# 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 : 0 ( 0.0%)
#   inexact prefix matches to 3'-adaptor : 48457 ( 0.1%)
#     exact prefix matches to 3'-adaptor : 12392902 ( 27.9%)
#
#           total matches to 5'-adaptor : 0 ( 0.0%)
#           total matches to 3'-adaptor : 12441359 ( 28.0%)
#
#           unique sequences (w/o adaptors) : 39321624 ( 88.6%)
#
# Sequence filtering:
#           5'-adaptor alignment too short : 0 ( 0.0%)
```

```

#           3'-adaptor alignment too short :           0 (  0.0%)
#                                           too short :       2363 (  0.0%)
#           too many non-base characters :           0 (  0.0%)
#                                           low entropy :    183109 (  0.4%)
#
# Final sequences:
#                                           passed :   44200443 ( 99.6%)

```

Top 10 tags before adaptor removal:

```

>seq0    2066
AAAAAAAAAAAAAAAAAAAAAAAAAAAA
>seq1    2054
GAATGGAATGGAATGGAATGGAATGGA
>seq2    2022
GGAATGGAATGGAATGGAATGGAATGG
>seq3    1984
ATTCCATTCCATTCCATTCCATTCCAT
>seq4    1671
TGGAATGGAATGGAATGGAATGGAATG
>seq5    1533
CATTCCATTCCATTCCATTCCATTCCA
>seq6    1414
TGIGIGIGIGIGIGIGIGIGIGIGIGT
>seq7    1231
AATGGAATGGAATGGAATGGAATGGAA
>seq8    1229
GATCGGAAGAGCTCGTATGCOGTCTTC
>seq9    1189
CACACACACACACACACACACACACAC

```

Mapping:

Number of mapped reads

```

Total reads mapped:           39862320
Total unmapped reads:        2126253
Total overmapped reads (>100 hits):    2211869

```

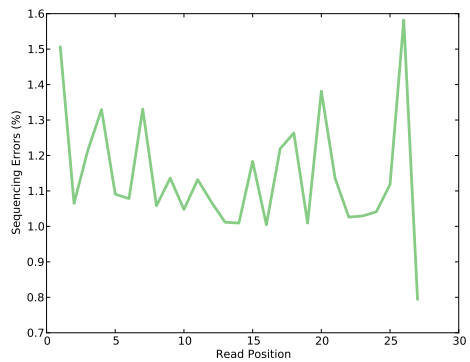


Figure 4: Fraction of all reads that had an error at read position

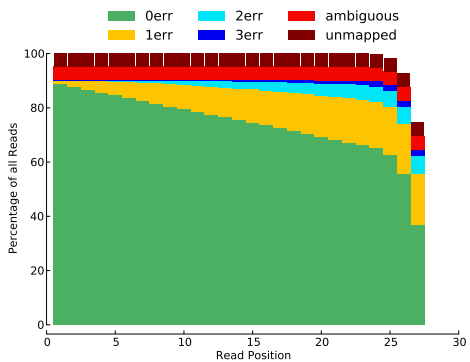


Figure 5: Fraction of all reads that have a certain mapping error up to read position

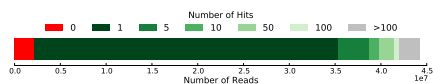


Figure 6: Number of reads that have a certain number of hits

Fragment Size Estimation:

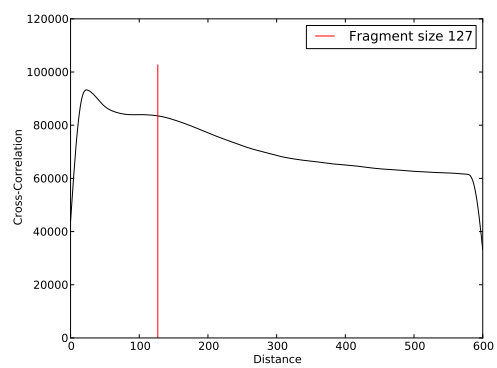


Figure 7: Fragment size correlation plot