RAD21: Quality Filtering:

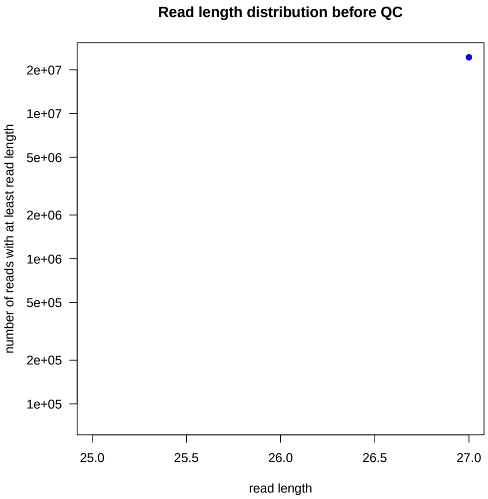

Figure 1: Read length reverse cumulative distribution before filtering

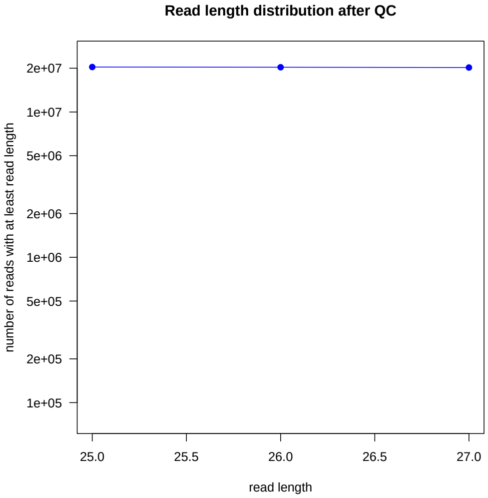

Figure 2: Read length reverse cumulative distribution after filtering

Adaptor Finding:

0.1

Chosen adaptor: GATCGGAAGAGCTCGTATGCCGTCTTCTGCTTG

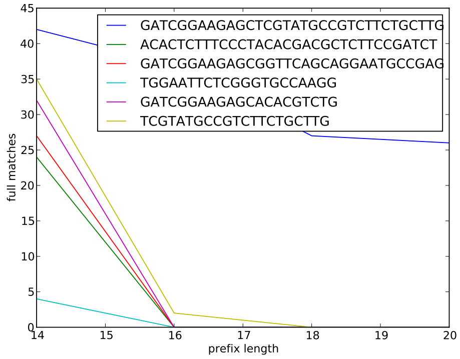

Figure 3: Plot for Adaptor Finding: Length of Adaptor versus Full Matches

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
# Wed Nov 26 19:09:52 2014: READING SEQUENCES
# Wed Nov 26 19:13:27 2014: done reading sequences (n=18527501)
# Wed Nov 26 19:13:27 2014: REMOVING ADAPTORS
# Wed Nov 26 19:28:28 2014: done removing adaptors (n=18525251)
# Wed Nov 26 19:28:28 2014: FILTERING AND OUTPUT
# Wed Nov 26 19:32:10 2014: done FILTERING AND OUTPUT
# SUMMARY:
# Number of sequences:
#                          input sequences :   20391660 (100.0%)
#                         unique sequences :   18527501 ( 90.9%)
#
# 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 :      24061 (   0.1%)
#      exact prefix matches to 3'-adaptor :    5880653 ( 28.8%)
#
#             total matches to 5'-adaptor :          0 (   0.0%)
#             total matches to 3'-adaptor :    5904714 ( 29.0%)
#
#         unique sequences (w/o adaptors) :   18525251 ( 90.8%)
#
# Sequence filtering:
#           5'-adaptor alignment too short :          0 (   0.0%)
```

```
#          3'-adaptor alignment too short :        0 (   0.0%)
#                               too short :     3352 (   0.0%)
#            too many non-base characters :        0 (   0.0%)
#                             low entropy :    80886 (   0.4%)
#
# Final sequences:
#                                  passed :  20307422 (  99.6%)
```

```
Top 10 tags before adaptor removal:
>seq0   1573
GATCGGAAGAGCTCGTATGCCGTCTTC
>seq1   1260
AAAAAAAAAAAAAAAAAAAAAAAAAAA
>seq2   882
ACACACACACACACACACACACACACA
>seq3   777
CACACACACACACACACACACACACAC
>seq4   670
TGTGTGTGTGTGTGTGTGTGTGTGTGT
>seq5   621
GAATGGAATGGAATGGAATGGAATGGA
>seq6   567
ATTCCATTCCATTCCATTCCATTCCAT
>seq7   501
GGAATGGAATGGAATGGAATGGAATGG
>seq8   478
TGGAATGGAATGGAATGGAATGGAATG
>seq9   465
CATTCCATTCCATTCCATTCCATTCCA
```

Mapping:

Number of mapped reads

```
Total reads mapped:     18853317
Total unmapped reads:   579919
Total overmapped reads (>100 hits):     874186
```

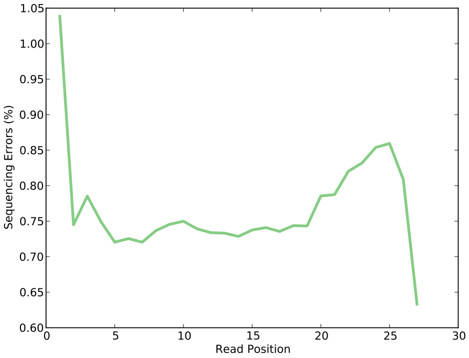

Figure 4: Fraction of all reads that had an error at read position

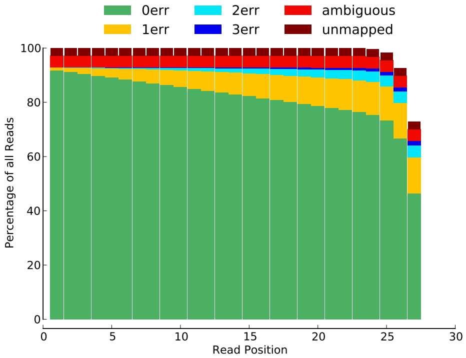

Figure 5: Fraction of all reads that have a certain mapping error up to read position

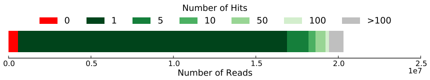

Figure 6: Number of reads that have a certain number of hits

Fragment Size Estimation:

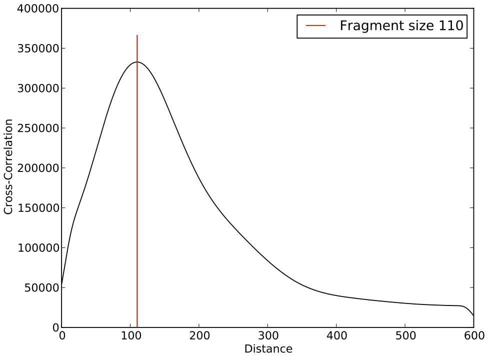

Figure 7: Fragment size correlation plot