RAD21: Quality Filtering:

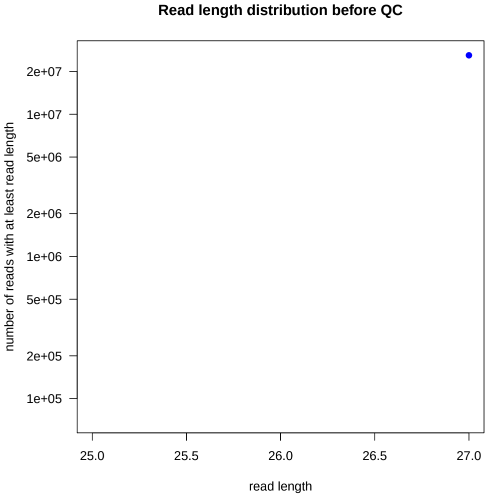

Figure 1: Read length reverse cumulative distribution before filtering

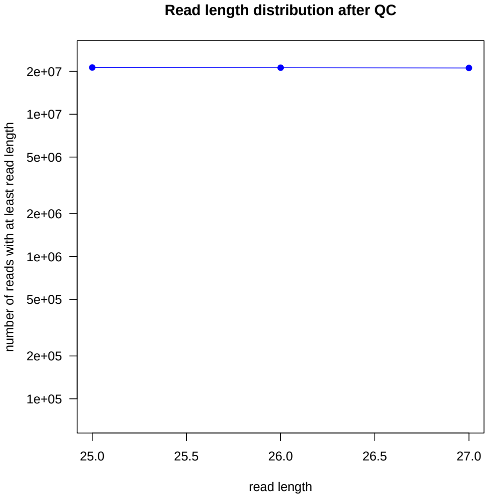

Figure 2: Read length reverse cumulative distribution after filtering

Adaptor Finding:

0.1

```
Chosen adaptor: GATCGGAAGAGCTCGTATGCCGTCTTCTGCTTG
```

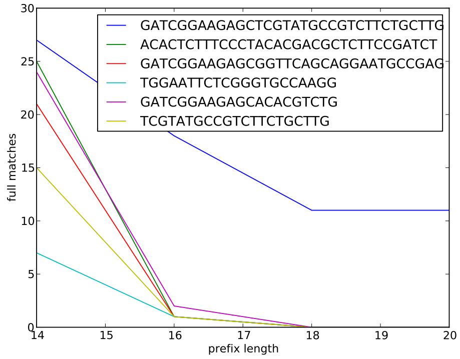

Figure 3: Plot for Adaptor Finding: Length of Adaptor versus Full Matches

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 19:10:17 2014: READING SEQUENCES
# Wed Nov 26 19:14:06 2014: done reading sequences (n=19561590)
# Wed Nov 26 19:14:06 2014: REMOVING ADAPTORS
# Wed Nov 26 19:30:53 2014: done removing adaptors (n=19559661)
# Wed Nov 26 19:30:53 2014: FILTERING AND OUTPUT
# Wed Nov 26 19:34:58 2014: done FILTERING AND OUTPUT
# SUMMARY:
# Number of sequences:
#                           input sequences :  21292540 (100.0%)
#                          unique sequences :  19561590 ( 91.9%)
#
# 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 :     25113 (   0.1%)
#     exact prefix matches to 3'-adaptor :   6263251 (  29.4%)
#
#            total matches to 5'-adaptor :         0 (   0.0%)
#            total matches to 3'-adaptor :   6288364 (  29.5%)
#
#        unique sequences (w/o adaptors) :  19559661 ( 91.9%)
#
# Sequence filtering:
#          5'-adaptor alignment too short :         0 (   0.0%)
```

```
#         3'-adaptor alignment too short :        0 (   0.0%)
#                            too short :     1522 (   0.0%)
#           too many non-base characters :        0 (   0.0%)
#                          low entropy :    92615 (   0.4%)
#
# Final sequences:
#                               passed :  21198403 (  99.6%)

Top 10 tags before adaptor removal:
>seq0    1291
AAAAAAAAAAAAAAAAAAAAAAAAAAA
>seq1    1275
GAATGGAATGGAATGGAATGGAATGGA
>seq2    1026
GGAATGGAATGGAATGGAATGGAATGG
>seq3    985
TGGAATGGAATGGAATGGAATGGAATG
>seq4    972
ATTCCATTCCATTCCATTCCATTCCAT
>seq5    852
TGTGTGTGTGTGTGTGTGTGTGTGTGT
>seq6    837
ACACACACACACACACACACACACACA
>seq7    761
ATGGAATGGAATGGAATGGAATGGAAT
>seq8    750
AATGGAATGGAATGGAATGGAATGGAA
>seq9    726
CACACACACACACACACACACACACAC
```

Mapping:

Number of mapped reads

```
Total reads mapped:     19212934
Total unmapped reads:   670858
Total overmapped reads (>100 hits):     1314611
```

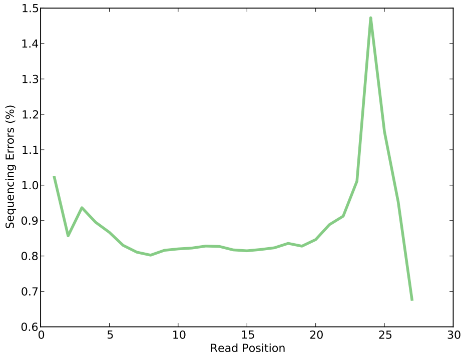

Figure 4: Fraction of all reads that had an error at read position

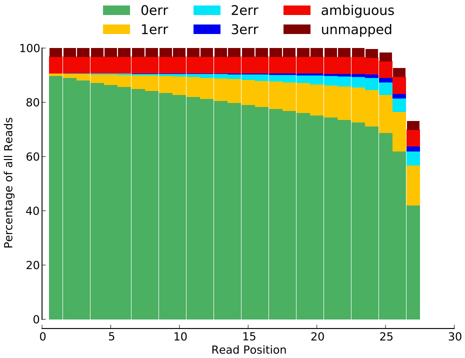

Figure 5: Fraction of all reads that have a certain mapping error up to read position

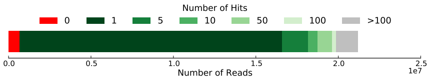

Figure 6: Number of reads that have a certain number of hits

Fragment Size Estimation:

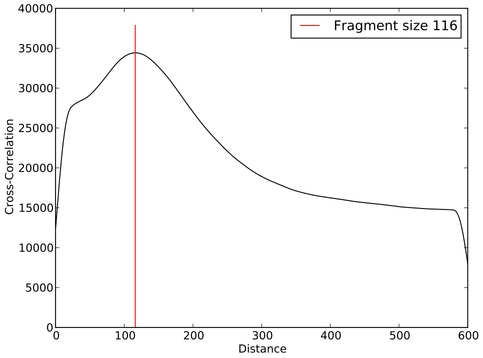

Figure 7: Fragment size correlation plot