Quality Filtering:

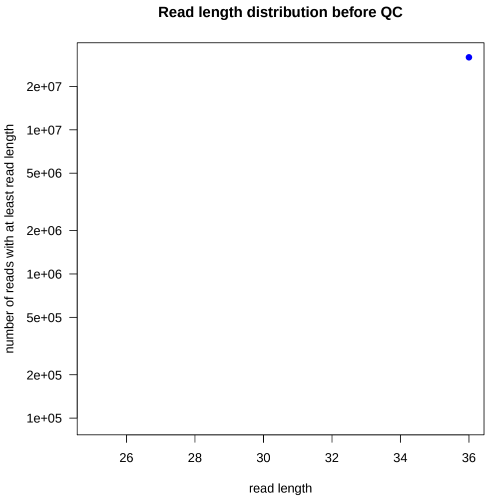

Figure 1: Read length reverse cumulative distribution before filtering

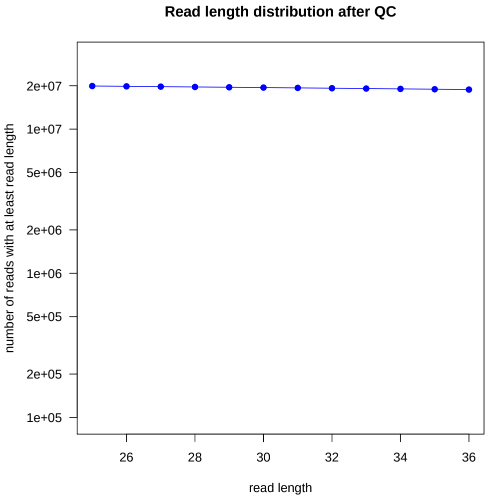

Figure 2: Read length reverse cumulative distribution after filtering

Adaptor Finding:

0.1

Chosen adaptor: ACACTCTTTCCCTACACGACGCTCTTCCGATCT

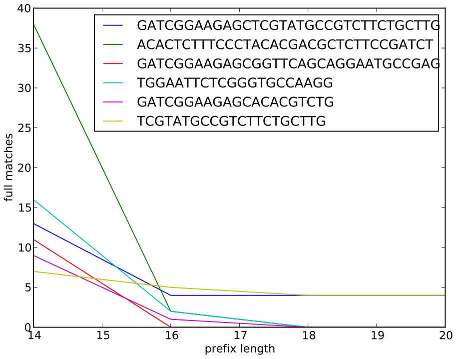

Figure 3: Plot for Adaptor Finding: Length of Adaptor versus Full Matches

Adaptor Removal and Quality Filtering:

0.2

```
# 5'/3' adaptors       :  / ACACTCTTTCCCTACACGACGCTCTTCCGATCT
# 5'/3' min aln. len.  : 0 / 0
# min len (1 mm)       : 7
# min len (2 mm)       : 10
# max mm (full)        : 2
# sequence filter      : on
# Sun Feb  1 21:37:23 2015: READING SEQUENCES
# Sun Feb  1 21:41:29 2015: done reading sequences (n=17759699)
# Sun Feb  1 21:41:29 2015: REMOVING ADAPTORS
# Sun Feb  1 22:01:40 2015: done removing adaptors (n=17756188)
# Sun Feb  1 22:01:40 2015: FILTERING AND OUTPUT
# Sun Feb  1 22:07:50 2015: done FILTERING AND OUTPUT
# SUMMARY:
# Number of sequences:
#                         input sequences :  19910032 (100.0%)
#                        unique sequences :  17759699 ( 89.2%)
#
# 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 :    63693 (   0.3%)
#     exact prefix matches to 3'-adaptor :  7311924 (  36.7%)
#
#           total matches to 5'-adaptor :        0 (   0.0%)
#           total matches to 3'-adaptor :  7375617 (  37.0%)
#
#        unique sequences (w/o adaptors) :  17756188 ( 89.2%)
#
# Sequence filtering:
#         5'-adaptor alignment too short :        0 (   0.0%)
```

```
#          3'-adaptor alignment too short :        0 (   0.0%)
#                               too short :       15 (   0.0%)
#          too many non-base characters :        0 (   0.0%)
#                             low entropy :    54878 (   0.3%)
#
# Final sequences:
#                                  passed :  19855139 (  99.7%)

Top 10 tags before adaptor removal:
>seq0    558
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
>seq1    400
AGTTTATATATCACAGTGAGAATTCTATGATGGACC
>seq2    333
TGTCCTACAAGCATTAATTAATTAACACACTTTAGT
>seq3    293
GTCCTACAAGCATTAATTAATTAACACACTTTAGTA
>seq4    249
AGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA
>seq5    246
TGTTCGCCTGTAATATTGAACGTAGGTGCGATAAAT
>seq6    246
GAATGGAATGGAATGGAATGGAATGGAATGGAATGG
>seq7    240
GAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAA
>seq8    234
TGGAATGGAATGGAATGGAATGGAATGGAATGGAAT
>seq9    228
GTTGTGGTAAACTTTAATAGTGTAGGAAGCTGAATA
```

Mapping:

Number of mapped reads

```
Total reads mapped:      18601134
Total unmapped reads:    855265
Total overmapped reads (>100 hits):      398740
```

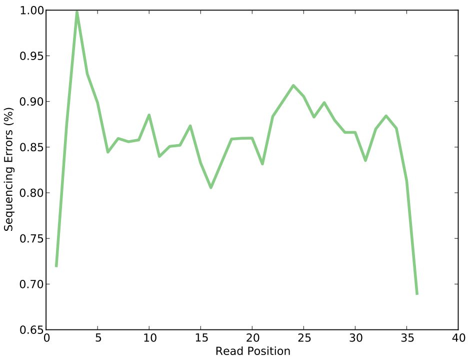

Figure 4: Fraction of all reads that had an error at read position

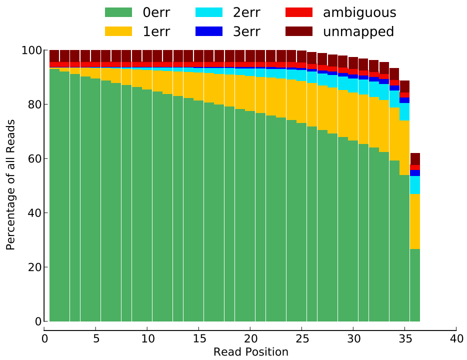

Figure 5: Fraction of all reads that have a certain mapping error up to read position

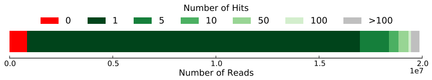

Figure 6: Number of reads that have a certain number of hits

Fragment Size Estimation:

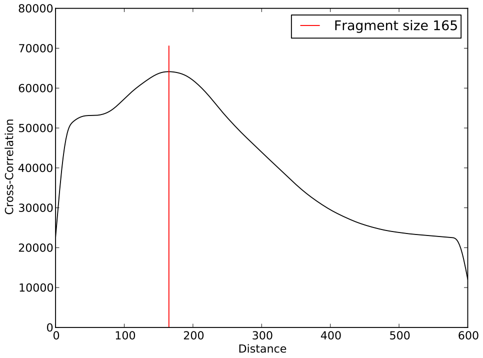

Figure 7: Fragment size correlation plot