Quality Filtering:

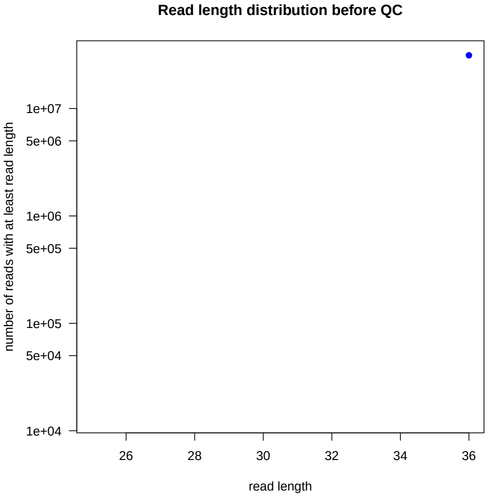

Figure 1: Read length reverse cumulative distribution before filtering

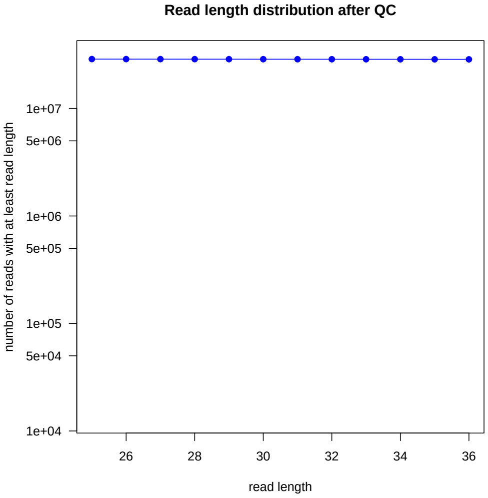

Figure 2: Read length reverse cumulative distribution after filtering

Adaptor Finding:

0.1

Chosen adaptor: GATCGGAAGAGCTCGTATGCCGTCTTCTGCTTG

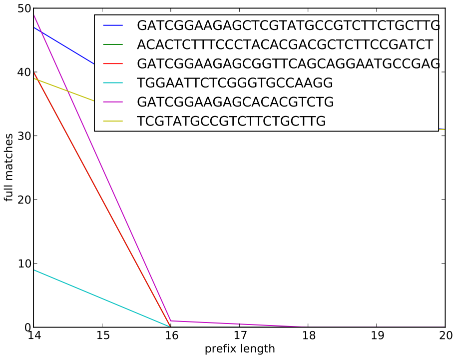

Figure 3: Plot for Adaptor Finding: Length of Adaptor versus Full Matches

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
# Sun Feb  1 22:40:36 2015: READING SEQUENCES
# Sun Feb  1 22:46:28 2015: done reading sequences (n=26474221)
# Sun Feb  1 22:46:28 2015: REMOVING ADAPTORS
# Sun Feb  1 23:19:49 2015: done removing adaptors (n=26473466)
# Sun Feb  1 23:19:49 2015: FILTERING AND OUTPUT
# Sun Feb  1 23:29:32 2015: done FILTERING AND OUTPUT
# SUMMARY:
# Number of sequences:
#                          input sequences :  28793339 (100.0%)
#                         unique sequences :  26474221 ( 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 :      3058 (   0.0%)
#    inexact prefix matches to 3'-adaptor :     30150 (   0.1%)
#      exact prefix matches to 3'-adaptor :   8341034 (  29.0%)
#
#             total matches to 5'-adaptor :         0 (   0.0%)
#             total matches to 3'-adaptor :   8374242 (  29.1%)
#
#         unique sequences (w/o adaptors) :  26473466 ( 91.9%)
#
# Sequence filtering:
#           5'-adaptor alignment too short :         0 (   0.0%)
```

```
#        3'-adaptor alignment too short :        0 (  0.0%)
#                            too short :     3273 (  0.0%)
#         too many non-base characters :        0 (  0.0%)
#                          low entropy :    81836 (  0.3%)
#
# Final sequences:
#                               passed :  28708230 ( 99.7%)

Top 10 tags before adaptor removal:
>seq0   2828
GATCGGAAGAGCTCGTATGCCGTCTTCTGCTTGAAA
>seq1   748
CTAACCCTAACCCTAACCCTAACCCTAACCCTAACC
>seq2   641
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
>seq3   549
ACACACACACACACACACACACACACACACACACAC
>seq4   542
CACACACACACACACACACACACACACACACACACA
>seq5   486
GGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAG
>seq6   461
CCTAACCCTAACCCTAACCCTAACCCTAACCCTAAC
>seq7   426
CCCTAACCCTAACCCTAACCCTAACCCTAACCCTAA
>seq8   423
GTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGG
>seq9   377
AGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTT
```

Mapping:

Number of mapped reads

```
Total reads mapped:     27061910
Total unmapped reads:   823115
Total overmapped reads (>100 hits):     823205
```

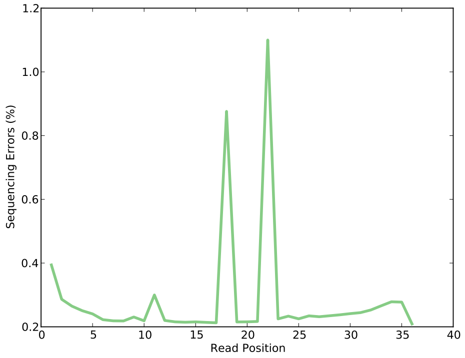

Figure 4: Fraction of all reads that had an error at read position

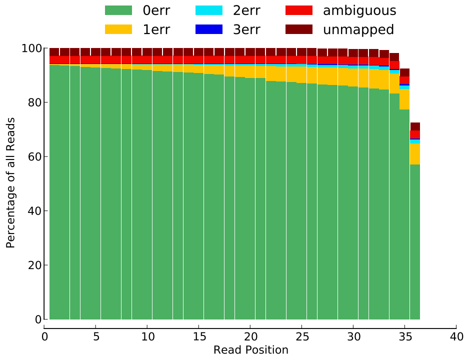

Figure 5: Fraction of all reads that have a certain mapping error up to read position

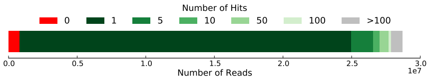

Figure 6: Number of reads that have a certain number of hits

Fragment Size Estimation:

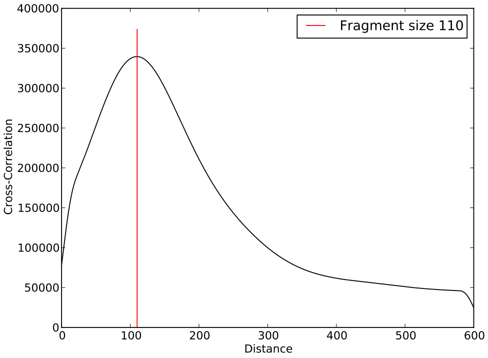

Figure 7: Fragment size correlation plot