CHD2: Quality Filtering:

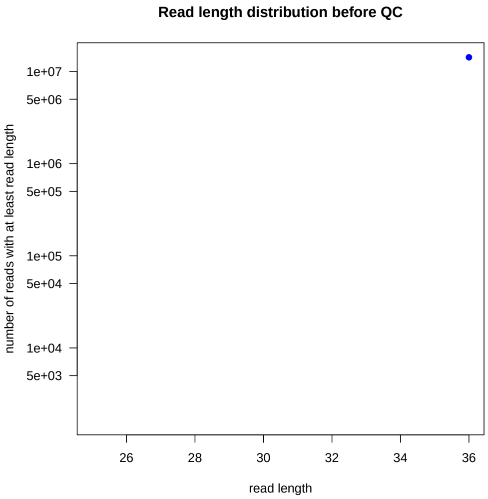

Figure 1: Read length reverse cumulative distribution before filtering

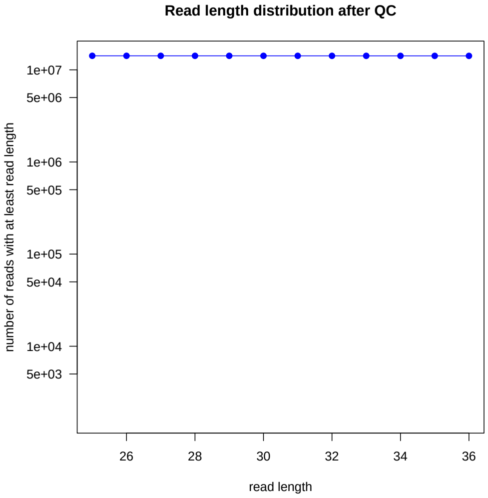

Figure 2: Read length reverse cumulative distribution after filtering

Adaptor Finding:

0.1

```
Chosen adaptor: GATCGGAAGAGCTCGTATGCCGTCTTCTGCTTG
```

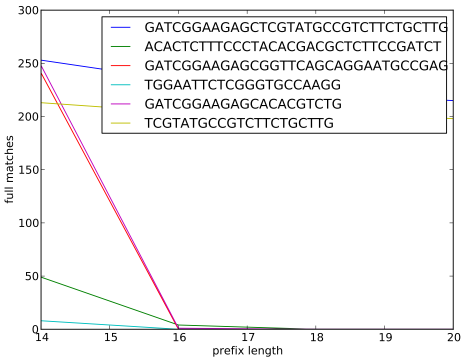

Figure 3: Plot for Adaptor Finding: Length of Adaptor versus Full Matches

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
# Sat Nov 29 16:19:29 2014: READING SEQUENCES
# Sat Nov 29 16:22:01 2014: done reading sequences (n=13254489)
# Sat Nov 29 16:22:01 2014: REMOVING ADAPTORS
# Sat Nov 29 16:36:00 2014: done removing adaptors (n=13254223)
# Sat Nov 29 16:36:00 2014: FILTERING AND OUTPUT
# Sat Nov 29 16:39:18 2014: done FILTERING AND OUTPUT
# SUMMARY:
# Number of sequences:
#                           input sequences :  14196800 (100.0%)
#                          unique sequences :  13254489 ( 93.4%)
#
# 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 :    10165 (   0.1%)
#    inexact prefix matches to 3'-adaptor :    26887 (   0.2%)
#      exact prefix matches to 3'-adaptor :  4502051 (  31.7%)
#
#             total matches to 5'-adaptor :        0 (   0.0%)
#             total matches to 3'-adaptor :  4539103 (  32.0%)
#
#         unique sequences (w/o adaptors) :  13254223 ( 93.4%)
#
# Sequence filtering:
#          5'-adaptor alignment too short :        0 (   0.0%)
```

```
#          3'-adaptor alignment too short :         0 (   0.0%)
#                               too short :     11283 (   0.1%)
#           too many non-base characters :         0 (   0.0%)
#                             low entropy :     35759 (   0.3%)
#
# Final sequences:
#                                  passed :  14149758 (  99.7%)
```

```
Top 10 tags before adaptor removal:
>seq0    9508
GATCGGAAGAGCTCGTATGCCGTCTTCTGCTTGAAA
>seq1    3655
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
>seq2    574
CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
>seq3    340
GGAATGGAATGGAATGGAATGGAATGGAATGGAATG
>seq4    339
GAATGGAATGGAATGGAATGGAATGGAATGGAATGG
>seq5    314
TGTAGTCCCAGCTACTCGGGAGGCTGAGGCAGGAGA
>seq6    304
CACACACACACACACACACACACACACACACACACA
>seq7    302
CCATTCCATTCCATTCCATTCCATTCCATTCCATTC
>seq8    281
CGATGATTGCATTCAATTCATTCGATGACGATTCCA
>seq9    263
TTCCATTCCATTCCATTCCATTCCATTCCATTCCAT
```

Mapping:

Number of mapped reads

```
Total reads mapped:     13237522
Total unmapped reads:   258651
Total overmapped reads (>100 hits):     653585
```

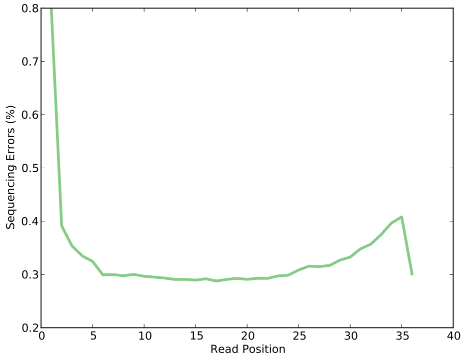

Figure 4: Fraction of all reads that had an error at read position

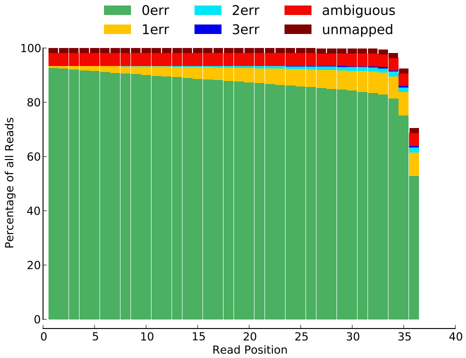

Figure 5: Fraction of all reads that have a certain mapping error up to read position

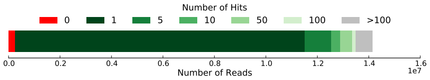

Figure 6: Number of reads that have a certain number of hits

Fragment Size Estimation:

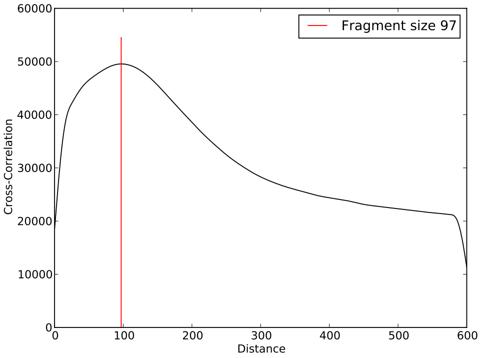

Figure 7: Fragment size correlation plot