

CHD2: Quality Filtering:

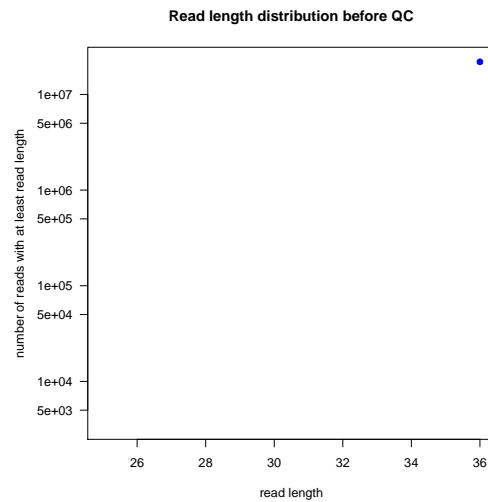


Figure 1: Read length reverse cumulative distribution before filtering

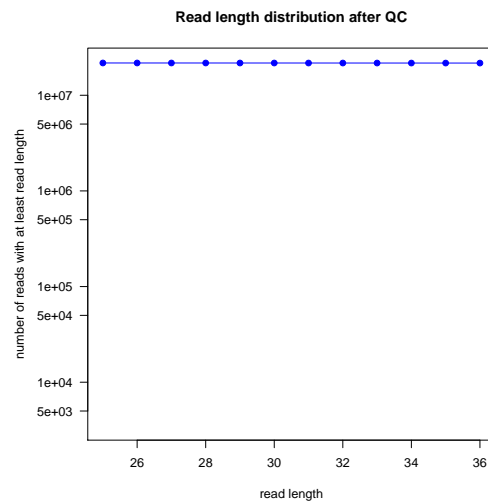


Figure 2: Read length reverse cumulative distribution after filtering

Adaptor Finding:

0.1

Chosen adaptor: GATCGGAAGAGCTCGTATGCGTCTTCTGCTTG

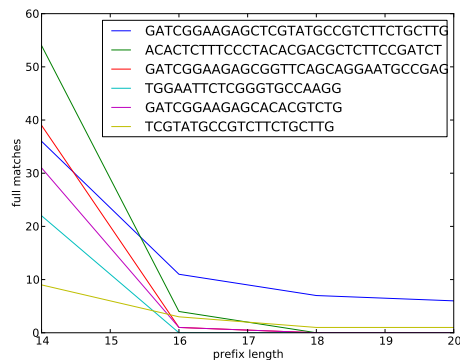


Figure 3: Plot for Adaptor Finding: Length of Adaptor versus Full Matches

Adaptor Removal and Quality Filtering:

0.2

```
# 5'/3' adaptors      : / GATCGGAAGAGCTCGTATGCCGCTTCTGCTTG
# 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 13:56:03 2014: READING SEQUENCES
# Sat Nov 29 14:00:00 2014: done reading sequences (n=19775709)
# Sat Nov 29 14:00:00 2014: REMOVING ADAPTORS
# Sat Nov 29 14:19:09 2014: done removing adaptors (n=19775325)
# Sat Nov 29 14:19:09 2014: FILTERING AND OUTPUT
# Sat Nov 29 14:24:20 2014: done FILTERING AND OUTPUT
# SUMMARY:
# Number of sequences:
#
#           input sequences : 21756670 (100.0%)
#           unique sequences : 19775709 ( 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 : 128 ( 0.0%)
#   inexact prefix matches to 3'-adaptor : 50865 ( 0.2%)
#     exact prefix matches to 3'-adaptor : 6868763 ( 31.6%)
#
#           total matches to 5'-adaptor : 0 ( 0.0%)
#           total matches to 3'-adaptor : 6919756 ( 31.8%)
#
#           unique sequences (w/o adaptors) : 19775325 ( 90.9%)
#
# Sequence filtering:
#           5'-adaptor alignment too short : 0 ( 0.0%)
```

```
#          3'-adaptor alignment too short :          0 (  0.0%)
#                                     too short :          499 (  0.0%)
#          too many non-base characters :          0 (  0.0%)
#                                     low entropy :       50010 (  0.2%)
#
# Final sequences:
#                                     passed :   21706161 ( 99.8%)
```

```
Top 10 tags before adaptor removal:
>seq0    2222
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
>seq1    532
CACACACACACACACACACACACACACACACACA
>seq2    474
GGAATGGAATGGAATGGAATGGAATGGAATGGAATG
>seq3    398
GAATGGAATGGAATGGAATGGAATGGAATGGAATGG
>seq4    370
ACACACACACACACACACACACACACACACACAC
>seq5    367
CCATTCATTCATTCATTCATTCATTCATTCATTC
>seq6    348
CATTCATTCATTCATTCATTCATTCATTCATTCATTC
>seq7    315
GATGIGIGGCTTCAACTCACAGAGTTTAAOCTTTCT
>seq8    314
ATTCCATTCATTCATTCATTCATTCATTCATTCATTC
>seq9    310
TGIGIGIGIGIGIGIGIGIGIGIGIGIGIGIGIG
```

Mapping:

Number of mapped reads

```
Total reads mapped:      20540046
Total unmapped reads:    347128
Total overmapped reads (>100 hits): 818987
```

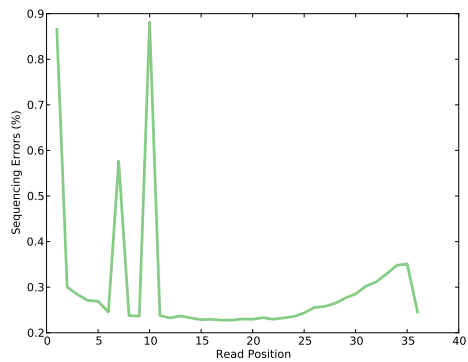


Figure 4: Fraction of all reads that had an error at read position

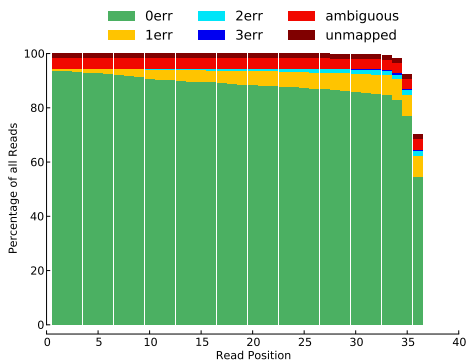


Figure 5: Fraction of all reads that have a certain mapping error up to read position

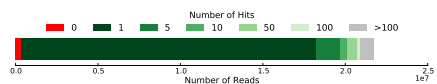


Figure 6: Number of reads that have a certain number of hits

Fragment Size Estimation:

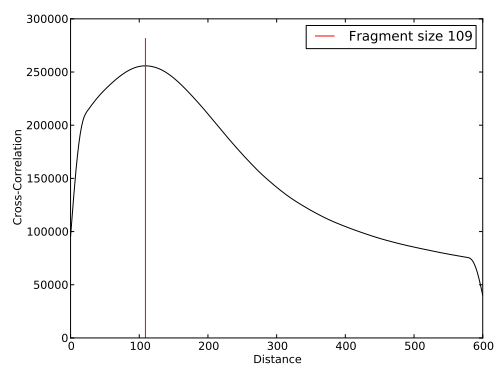


Figure 7: Fragment size correlation plot