

Quality Filtering:

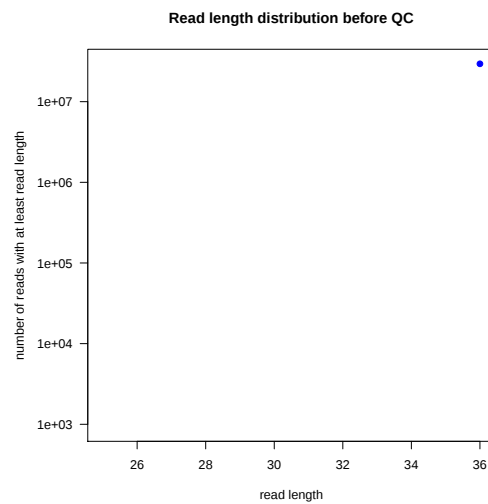


Figure 1: Read length reverse cumulative distribution before filtering

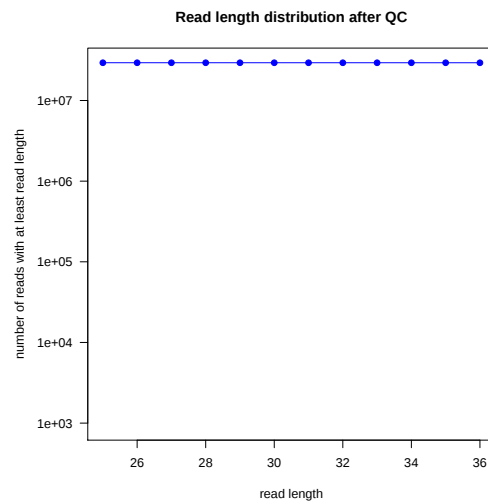


Figure 2: Read length reverse cumulative distribution after filtering

Adaptor Finding:

0.1

Chosen adaptor: TCGTATGCCGICTTCTGCTTG

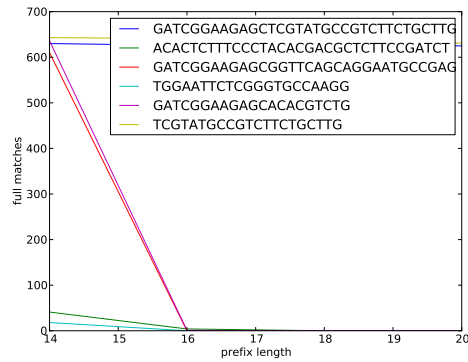


Figure 3: Plot for Adaptor Finding: Length of Adaptor versus Full Matches

Adaptor Removal and Quality Filtering:

0.2

```
# 5'/3' adaptors      : / TCGTATGCCGCTTCTGCTTG
# 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 22:26:12 2015: READING SEQUENCES
# Sun Feb  1 22:32:03 2015: done reading sequences (n=24336862)
# Sun Feb  1 22:32:03 2015: REMOVING ADAPTORS
# Sun Feb  1 23:03:15 2015: done removing adaptors (n=24336127)
# Sun Feb  1 23:03:15 2015: FILTERING AND OUTPUT
# Sun Feb  1 23:11:51 2015: done FILTERING AND OUTPUT
# SUMMARY:
# Number of sequences:
#
#           input sequences : 29389998 (100.0%)
#           unique sequences : 24336862 ( 82.8%)
#
# 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 : 73199 ( 0.2%)
#   inexact prefix matches to 3'-adaptor : 25704 ( 0.1%)
#     exact prefix matches to 3'-adaptor : 10221150 ( 34.8%)
#
#           total matches to 5'-adaptor : 0 ( 0.0%)
#           total matches to 3'-adaptor : 10320053 ( 35.1%)
#
#           unique sequences (w/o adaptors) : 24336127 ( 82.8%)
#
# Sequence filtering:
#           5'-adaptor alignment too short : 0 ( 0.0%)
```

```
#          3'-adaptor alignment too short :          0 (  0.0%)
#                                     too short :       73188 (  0.2%)
#          too many non-base characters :          0 (  0.0%)
#                                     low entropy :    115813 (  0.4%)
#
# Final sequences:
#                                     passed :   29200997 ( 99.4%)
```

Top 10 tags before adaptor removal:

```
>seq0    126205
AGCTTACTGTTAATCTTATTTTGCCTGCTGGTTG
>seq1    116418
GATCCAAATTTCCAACAAAAGGTGGTAGATGGCTG
>seq2    111884
AGCTTGATCCCCGCCCTCCCGTAGAGGAGCTCCAC
>seq3    97096
GATCCCTTCCCTGTGCGGGCCGCGGACGCTTTTCGCC
>seq4    66924
GATCGGAAGAGCTCGTATGCGGTCTTCTGCTTGAAA
>seq5    53868
TCGAGTGGGTTTTTGCTCCAGCGCGCTTACAAAGTCC
>seq6    52854
AGCTTTAACAACCACGTGCGCACGCTTGTGGCGCGC
>seq7    40082
AGCTTAAAAGATGACACCGAACAGGGCGGGGATGGC
>seq8    37430
GATCCCTTCTGCGGCTGCTCGGCTCAGCGCGCTT
>seq9    12206
TCGAGACTCCGGTAGCCATGGGCGCGATGGGACGGG
```

Mapping:

Number of mapped reads

```
Total reads mapped:      26029452
Total unmapped reads:    2150856
Total overmapped reads (>100 hits): 1020689
```

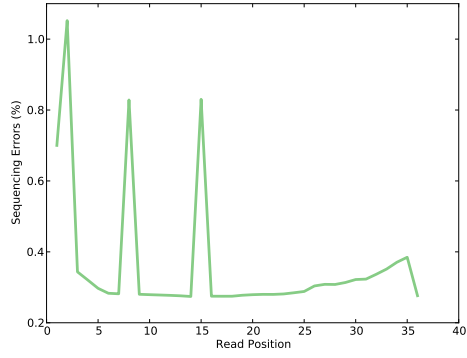


Figure 4: Fraction of all reads that had an error at read position

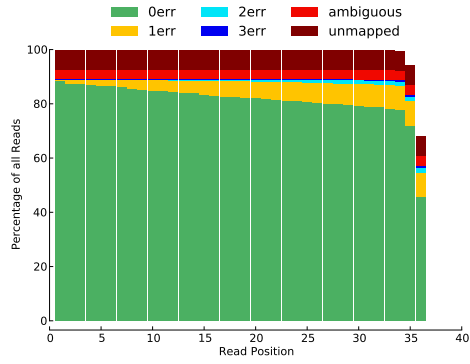


Figure 5: Fraction of all reads that have a certain mapping error up to read position

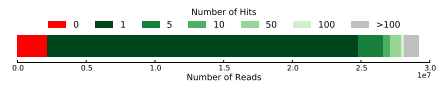


Figure 6: Number of reads that have a certain number of hits

Fragment Size Estimation:

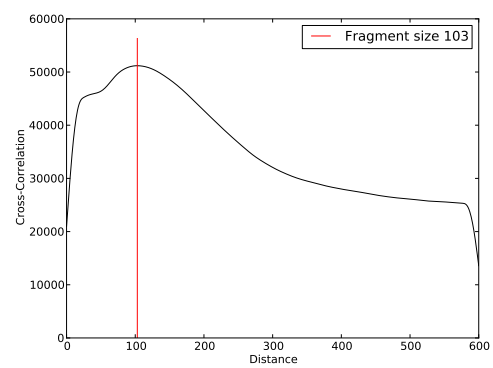


Figure 7: Fragment size correlation plot