Quality Filtering:

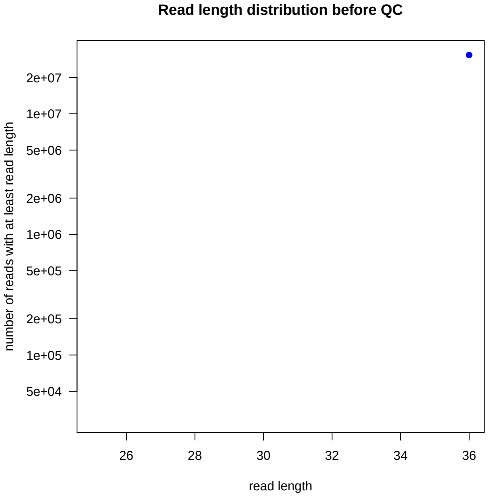

Figure 1: Read length reverse cumulative distribution before filtering

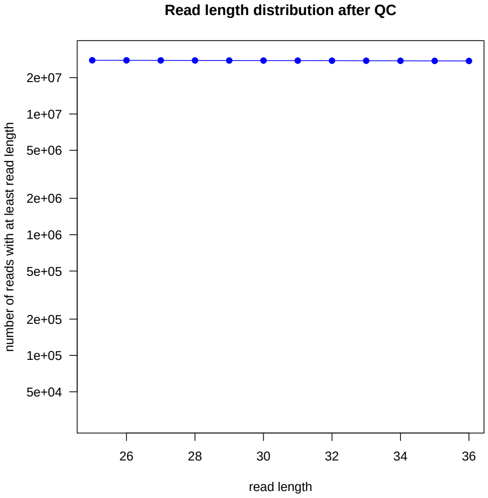

Figure 2: Read length reverse cumulative distribution after filtering

Adaptor Finding:

0.1

Chosen adaptor: TCGTATGCCGTCTTCTGCTTG

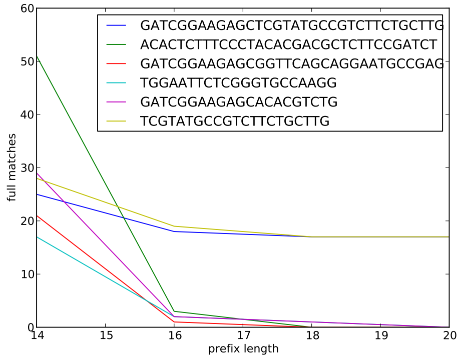

Figure 3: Plot for Adaptor Finding: Length of Adaptor versus Full Matches

Adaptor Removal and Quality Filtering:

0.2

```
# 5'/3' adaptors       :  / TCGTATGCCGTCTTCTGCTTG
# 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 22:26:02 2015: READING SEQUENCES
# Sun Feb  1 22:31:37 2015: done reading sequences (n=25214068)
# Sun Feb  1 22:31:37 2015: REMOVING ADAPTORS
# Sun Feb  1 23:04:32 2015: done removing adaptors (n=25212364)
# Sun Feb  1 23:04:32 2015: FILTERING AND OUTPUT
# Sun Feb  1 23:13:28 2015: done FILTERING AND OUTPUT
# SUMMARY:
# Number of sequences:
#                          input sequences :  27838418 (100.0%)
#                         unique sequences :  25214068 ( 90.6%)
#
# 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 :      2222 (  0.0%)
#   inexact prefix matches to 3'-adaptor :     25101 (  0.1%)
#     exact prefix matches to 3'-adaptor :  9642839 ( 34.6%)
#
#             total matches to 5'-adaptor :         0 (  0.0%)
#             total matches to 3'-adaptor :  9670162 ( 34.7%)
#
#        unique sequences (w/o adaptors) :  25212364 ( 90.6%)
#
# Sequence filtering:
#         5'-adaptor alignment too short :         0 (  0.0%)
```

```
#         3'-adaptor alignment too short :        0 (  0.0%)
#                              too short :     2289 (  0.0%)
#           too many non-base characters :        0 (  0.0%)
#                            low entropy :    58179 (  0.2%)
#
# Final sequences:
#                                 passed : 27777950 ( 99.8%)
```

```
Top 10 tags before adaptor removal:
>seq0   1901
GATCGGAAGAGCTCGTATGCCGTCTTCTGCTTGAAA
>seq1   1758
CTGCTCTATGAAAAGAAAGGTTAAACTCTGTGAGTT
>seq2   1070
CTTCCTTGTGTTGTGTGTATTCAACTCACAGAGTTG
>seq3   961
CATTGACCTCAAAGCGGCTGAAATCTCCACTTGCAA
>seq4   922
GAAACACTCTGTTTGTAAAGTCTGCAAGTGGATATT
>seq5   907
TCTGCTCTGTGTAAAGGATCGTTCAACTCTGTGAGT
>seq6   770
TGGAATGGAATGGAATGGAATGGAATGGAATGGAAT
>seq7   751
AATGGAATGGAATGGAATGGAATGGAATGGAATGGA
>seq8   736
CATTCCATTCCATTCCATTCCATTCCATTCCATTCC
>seq9   719
AAAATGTGGCACATATACACCATGGAATACTATGCA
```

Mapping:

Number of mapped reads

```
Total reads mapped:     25857877
Total unmapped reads:   890135
Total overmapped reads (>100 hits):     1029938
```

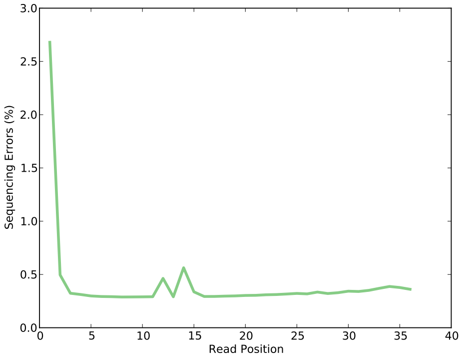

Figure 4: Fraction of all reads that had an error at read position

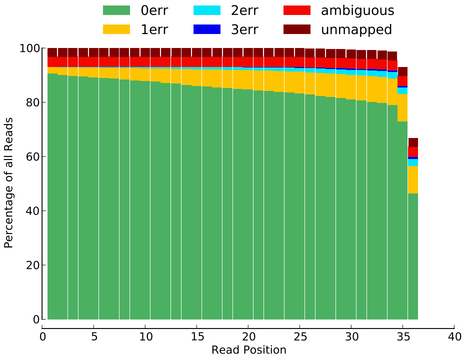

Figure 5: Fraction of all reads that have a certain mapping error up to read position

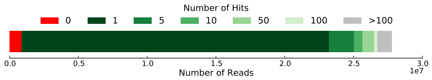

Figure 6: Number of reads that have a certain number of hits

Fragment Size Estimation:

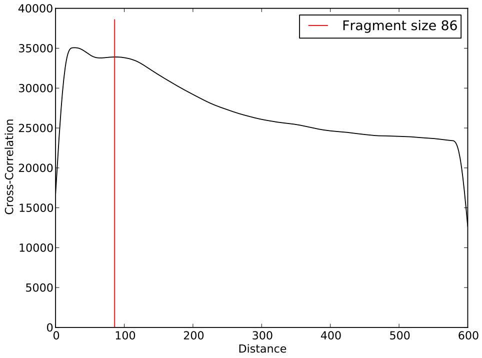

Figure 7: Fragment size correlation plot