Background: Quality Filtering:

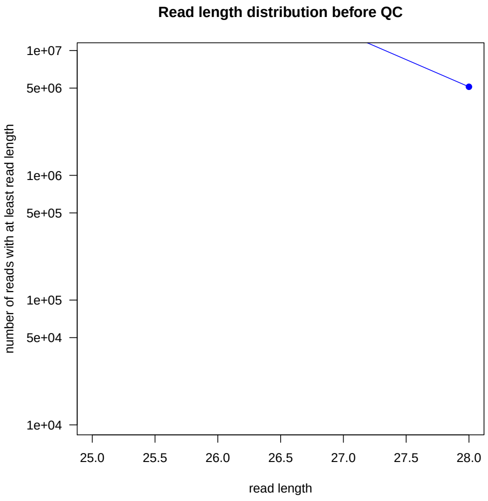

Figure 1: Read length reverse cumulative distribution before filtering

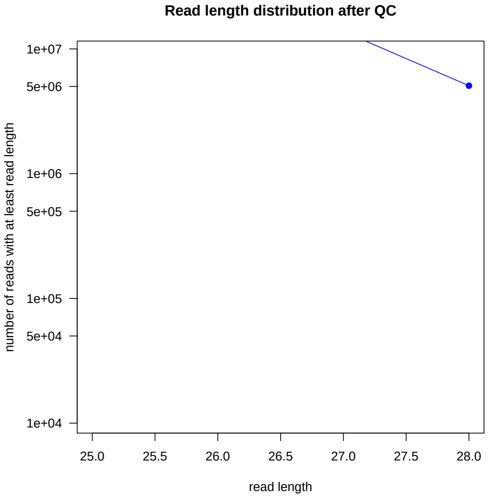

Figure 2: Read length reverse cumulative distribution after filtering

Adaptor Finding:

0.1

Chosen adaptor: ACACTCTTTCCCTACACGACGCTCTTCCGATCT

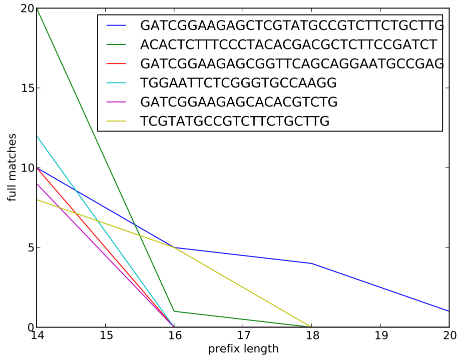

Figure 3: Plot for Adaptor Finding: Length of Adaptor versus Full Matches

Adaptor Removal and Quality Filtering:

0.2

```
# 5'/3' adaptors       :  / ACACTCTTTCCCTACACGACGCTCTTCCGATCT
# 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 Nov 30 18:51:35 2014: READING SEQUENCES
# Sun Nov 30 18:54:04 2014: done reading sequences (n=12406873)
# Sun Nov 30 18:54:04 2014: REMOVING ADAPTORS
# Sun Nov 30 19:04:39 2014: done removing adaptors (n=12348679)
# Sun Nov 30 19:04:39 2014: FILTERING AND OUTPUT
# Sun Nov 30 19:07:18 2014: done FILTERING AND OUTPUT
# SUMMARY:
# Number of sequences:
#                          input sequences :  13816004 (100.0%)
#                         unique sequences :  12406873 ( 89.8%)
#
# 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 :         0 (  0.0%)
#     inexact prefix matches to 3'-adaptor :     44171 (  0.3%)
#       exact prefix matches to 3'-adaptor :   4684456 ( 33.9%)
#
#              total matches to 5'-adaptor :         0 (  0.0%)
#              total matches to 3'-adaptor :   4728627 ( 34.2%)
#
#          unique sequences (w/o adaptors) :  12348679 ( 89.4%)
#
# Sequence filtering:
#           5'-adaptor alignment too short :         0 (  0.0%)
```

```
#         3'-adaptor alignment too short :         0 (   0.0%)
#                             too short :        83 (   0.0%)
#           too many non-base characters :         0 (   0.0%)
#                            low entropy :    277682 (   2.0%)
#
# Final sequences:
#                                 passed :  13538239 (  98.0%)
```

```
Top 10 tags before adaptor removal:
>seq0   172428
AAAAAAAAAAAAAAAAAAAAAAAAAAA
>seq1   30908
AAAAAAAAAAAAAAAAAAAAAAAAAAAA
>seq2   2454
CCCCCCCCCCCCCCCCCCCCCCCCCCC
>seq3   609
ACACACACACACACACACACACACACA
>seq4   584
AAAAAAAAAAAAAAAAAAAAAAAANAN
>seq5   581
CACACACACACACACACACACACACAC
>seq6   436
AAAAAAAAAAAAAAAAAAAAAAAAAAAC
>seq7   394
AAAAAAAAAAAAAAAAAAAAAAAAACA
>seq8   367
AAAAAAAAAAAAAAAAAAAAAAAAAACA
>seq9   365
GTGTGTGTGTGTGTGTGTGTGTGTGTG
```

Mapping:

Number of mapped reads

```
Total reads mapped:     12092970
Total unmapped reads:   638829
Total overmapped reads (>100 hits):     806440
```

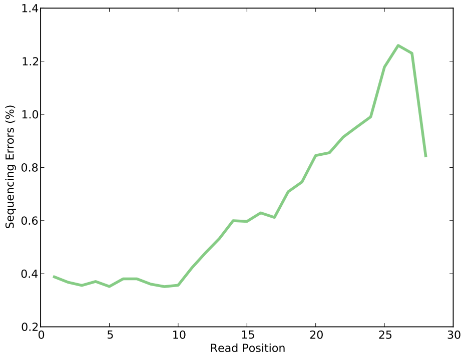

Figure 4: Fraction of all reads that had an error at read position

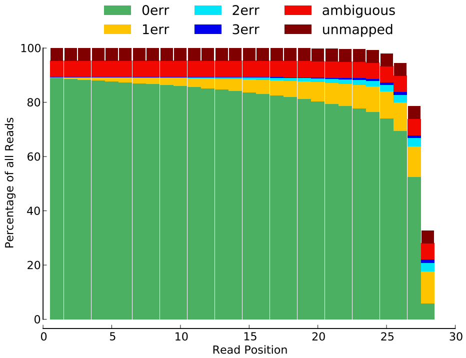

Figure 5: Fraction of all reads that have a certain mapping error up to read position

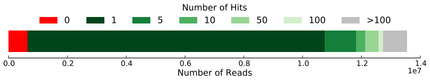

Figure 6: Number of reads that have a certain number of hits

Fragment Size Estimation:

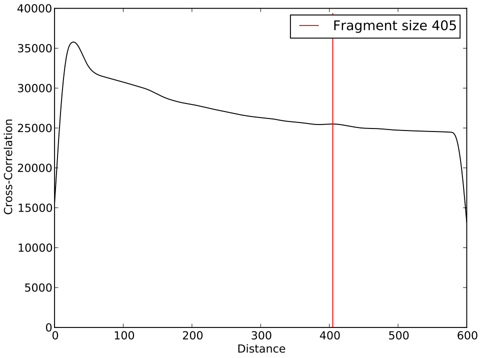

Figure 7: Fragment size correlation plot