ZNF143: Quality Filtering:

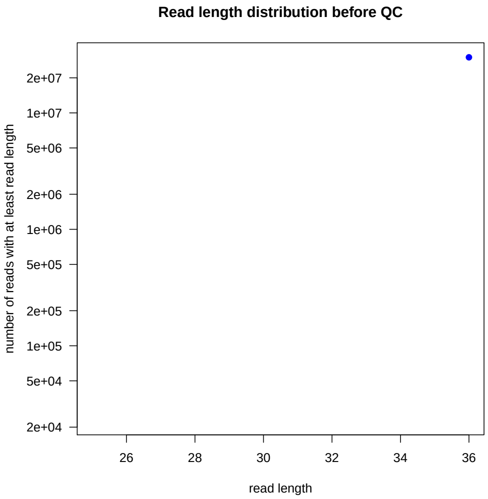

Figure 1: Read length reverse cumulative distribution before filtering

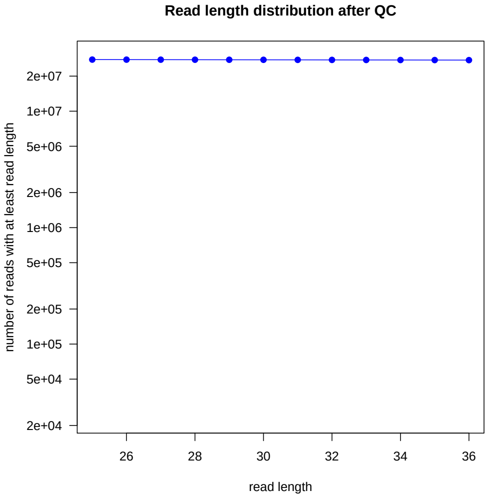

Figure 2: Read length reverse cumulative distribution after filtering

Adaptor Finding:

0.1

```
Chosen adaptor: TCGTATGCCGTCTTCTGCTTG
```

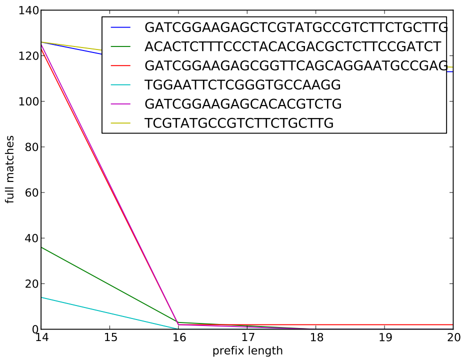

Figure 3: Plot for Adaptor Finding: Length of Adaptor versus Full Matches

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 Nov 30 15:09:39 2014: READING SEQUENCES
# Sun Nov 30 15:14:51 2014: done reading sequences (n=23450557)
# Sun Nov 30 15:14:51 2014: REMOVING ADAPTORS
# Sun Nov 30 15:41:59 2014: done removing adaptors (n=23448875)
# Sun Nov 30 15:41:59 2014: FILTERING AND OUTPUT
# Sun Nov 30 15:47:55 2014: done FILTERING AND OUTPUT
# SUMMARY:
# Number of sequences:
#                         input sequences :  27745544 (100.0%)
#                        unique sequences :  23450557 ( 84.5%)
#
# 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 :     12075 (  0.0%)
#    inexact prefix matches to 3'-adaptor :     22853 (  0.1%)
#      exact prefix matches to 3'-adaptor :   9082347 ( 32.7%)
#
#             total matches to 5'-adaptor :         0 (  0.0%)
#             total matches to 3'-adaptor :   9117275 ( 32.9%)
#
#        unique sequences (w/o adaptors) :  23448875 ( 84.5%)
#
# Sequence filtering:
#          5'-adaptor alignment too short :         0 (  0.0%)
```

```
#          3'-adaptor alignment too short :        0 (   0.0%)
#                               too short :    12236 (   0.0%)
#            too many non-base characters :        0 (   0.0%)
#                             low entropy :    56018 (   0.2%)
#
# Final sequences:
#                                  passed :  27677290 (  99.8%)
```

```
Top 10 tags before adaptor removal:
>seq0   11025
GATCGGAAGAGCTCGTATGCCGTCTTCTGCTTGAAA
>seq1   1621
CTGCTCTATGAAAAGAAAGGTTAAACTCTGTGAGTT
>seq2   888
TCTGCTCTGTGTAAAGGATCGTTCAACTCTGTGAGT
>seq3   827
CTTCCTTGTGTTGTGTGTATTCAACTCACAGAGTTG
>seq4   707
GGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAG
>seq5   676
CACACACACACACACACACACACACACACACACACA
>seq6   635
ACACACACACACACACACACACACACACACACACAC
>seq7   627
GAAACACTCTGTTTGTAAAGTCTGCAAGTGGATATT
>seq8   586
CATTGACCTCAAAGCGGCTGAAATCTCCACTTGCAA
>seq9   529
AAAATGTGGCACATATACACCATGGAATACTATGCA
```

Mapping:

Number of mapped reads

```
Total reads mapped:     26096018
Total unmapped reads:   832371
Total overmapped reads (>100 hits):     748901
```

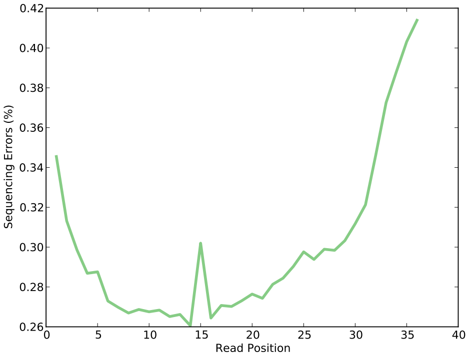

Figure 4: Fraction of all reads that had an error at read position

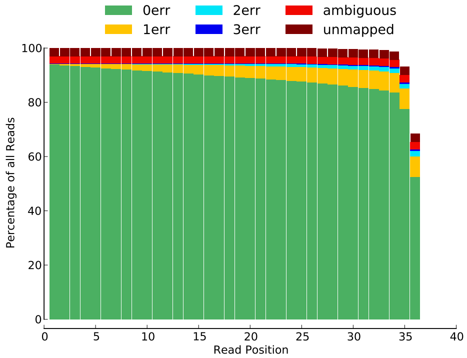

Figure 5: Fraction of all reads that have a certain mapping error up to read position

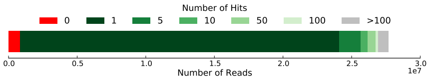

Figure 6: Number of reads that have a certain number of hits

Fragment Size Estimation:

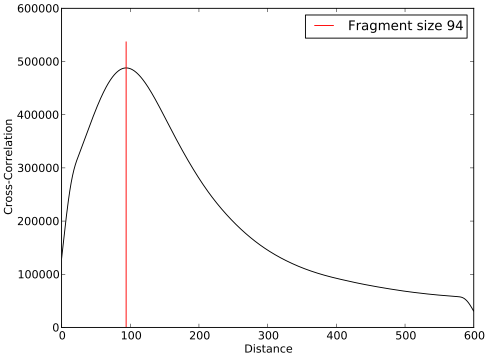

Figure 7: Fragment size correlation plot