ZNF143: Quality Filtering:

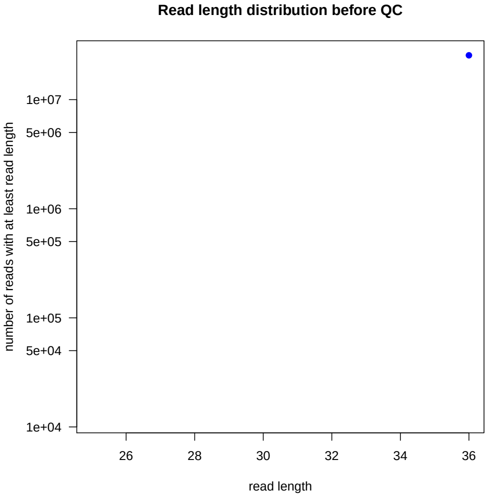

Figure 1: Read length reverse cumulative distribution before filtering

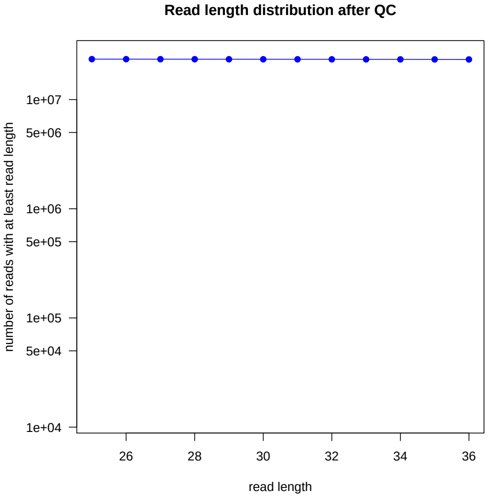

Figure 2: Read length reverse cumulative distribution after filtering

Adaptor Finding:

0.1

Chosen adaptor: GATCGGAAGAGCTCGTATGCCGTCTTCTGCTTG

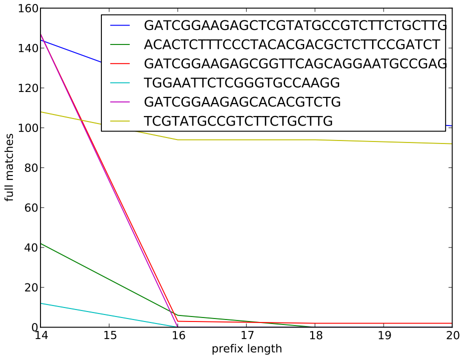

Figure 3: Plot for Adaptor Finding: Length of Adaptor versus Full Matches

Adaptor Removal and Quality Filtering:

0.2

```
# 5'/3' adaptors       :  / GATCGGAAGAGCTCGTATGCCGTCTTCTGCTTG
# 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 14:59:00 2014: READING SEQUENCES
# Sun Nov 30 15:03:34 2014: done reading sequences (n=20267249)
# Sun Nov 30 15:03:34 2014: REMOVING ADAPTORS
# Sun Nov 30 15:24:54 2014: done removing adaptors (n=20265919)
# Sun Nov 30 15:24:54 2014: FILTERING AND OUTPUT
# Sun Nov 30 15:30:08 2014: done FILTERING AND OUTPUT
# SUMMARY:
# Number of sequences:
#                          input sequences :  23462173 (100.0%)
#                         unique sequences :  20267249 ( 86.4%)
#
# 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 :      9243 (   0.0%)
#   inexact prefix matches to 3'-adaptor :     74843 (   0.3%)
#     exact prefix matches to 3'-adaptor :   7497730 (  32.0%)
#
#            total matches to 5'-adaptor :         0 (   0.0%)
#            total matches to 3'-adaptor :   7581816 (  32.3%)
#
#        unique sequences (w/o adaptors) :  20265919 ( 86.4%)
#
# Sequence filtering:
#         5'-adaptor alignment too short :         0 (   0.0%)
```

```
#          3'-adaptor alignment too short :         0 (   0.0%)
#                               too short :      9948 (   0.0%)
#           too many non-base characters :         0 (   0.0%)
#                             low entropy :     63642 (   0.3%)
#
# Final sequences:
#                                  passed :  23388583 (  99.7%)

Top 10 tags before adaptor removal:
>seq0    8765
GATCGGAAGAGCTCGTATGCCGTCTTCTGCTTGAAA
>seq1    1068
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
>seq2    652
CACACACACACACACACACACACACACACACACACA
>seq3    649
CTAACCCTAACCCTAACCCTAACCCTAACCCTAACC
>seq4    602
ACACACACACACACACACACACACACACACACACAC
>seq5    551
GGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAG
>seq6    516
GTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGG
>seq7    489
CCCTAACCCTAACCCTAACCCTAACCCTAACCCTAA
>seq8    408
GGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTA
>seq9    405
GAATGGAATGGAATGGAATGGAATGGAATGGAATGG
```

Mapping:

Number of mapped reads

```
Total reads mapped:      21123856
Total unmapped reads:    1449184
Total overmapped reads (>100 hits):      815543
```

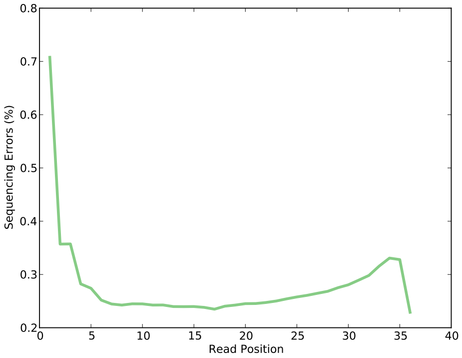

Figure 4: Fraction of all reads that had an error at read position

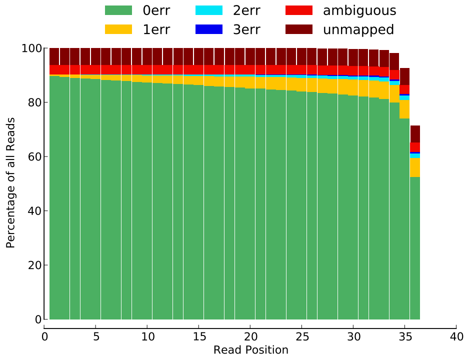

Figure 5: Fraction of all reads that have a certain mapping error up to read position

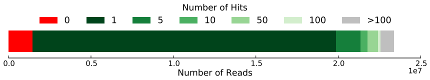

Figure 6: Number of reads that have a certain number of hits

Fragment Size Estimation:

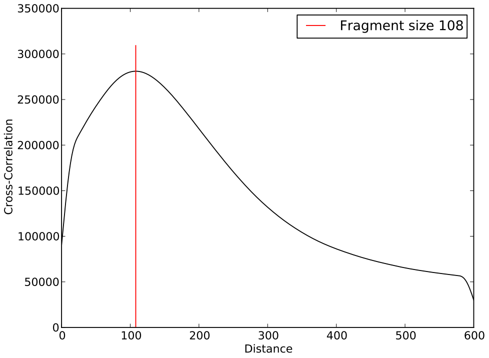

Figure 7: Fragment size correlation plot