

Supplementary Figure 1: Output file generated by FaQCs. A) Summary of trimming statistics; B) Read length histogram; C) Nucleotide composition histogram for the reads; D) Per cycle nucleotide composition plot; E) k-mer rarefaction curve; F) k-mer frequency histogram; G) Average read quality histogram; H) Per cycle quality box plot; I) Per cycle, per score frequency plot; J) Average read quality histogram.

A)

QC stats

Before Trimming
Reads: 11458940
Total bases: 1730299940
Reads Length: 151.00

After Trimming
Reads: 11239533 (98.09 %)
Total bases: 1599703670 (92.45 %)
Mean Reads Length: 142.33
Paired Reads: 11041676 (98.24 %)
Paired total bases: 1573382551 (98.35 %)
Unpaired Reads: 197857 (1.76 %)
Unpaired total bases: 26321119 (1.65 %)

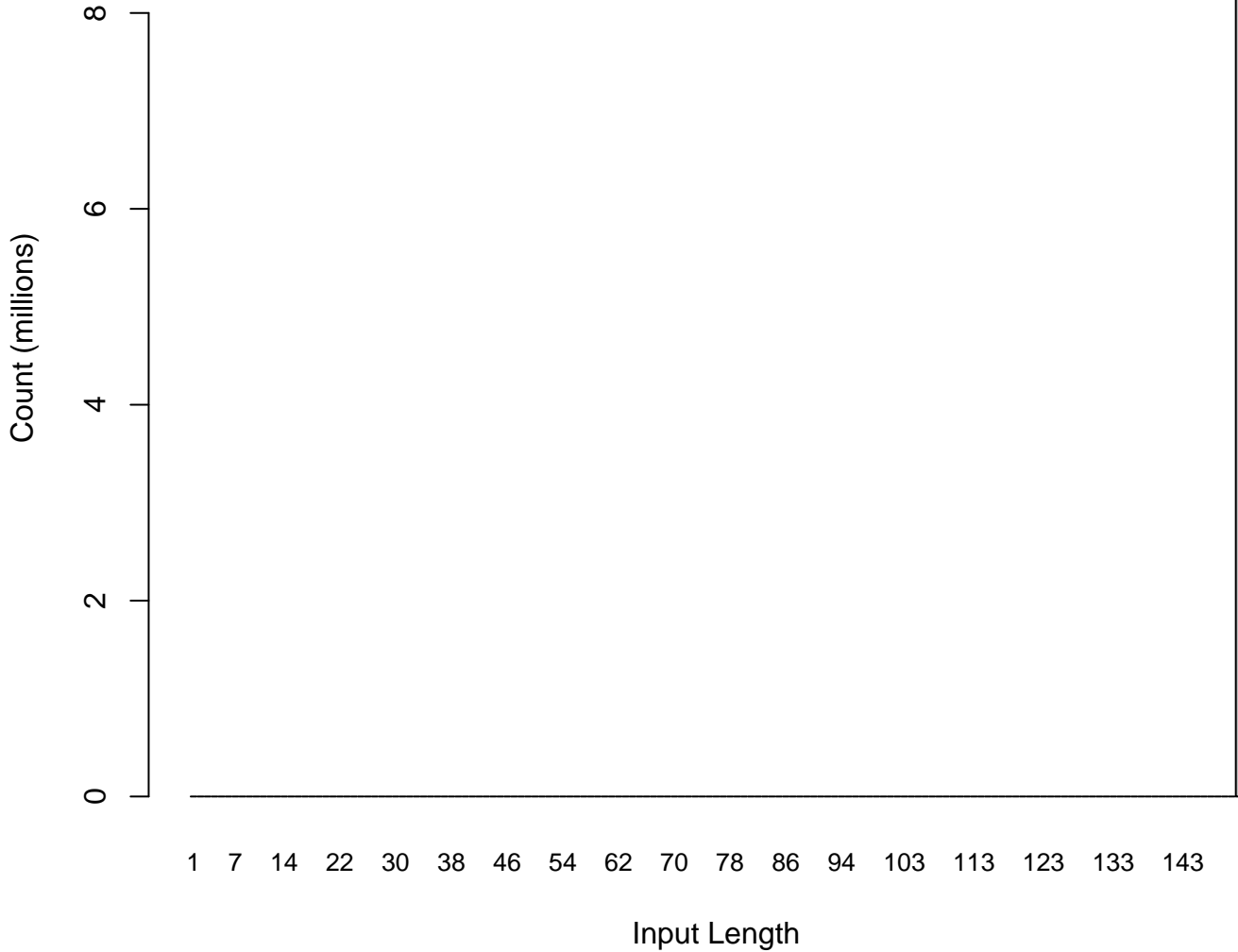
Discarded reads: 219407 (1.91 %)
Trimmed bases: 130596270 (7.55 %)
Reads Filtered by length cutoff (50 bp): 201882 (1.76 %)
Bases Filtered by length cutoff: 5907870 (0.34 %)
Reads Filtered by continuous base "N" (2): 0 (0.00 %)
Bases Filtered by continuous base "N": 0 (0.00 %)
Reads Filtered by low complexity ratio (0.8): 1805 (0.02 %)
Bases Filtered by low complexity ratio: 254307 (0.01 %)
Reads Filtered by avg quality (20.0): 15720 (0.14 %)
Bases Filtered by avg quality: 2331521 (0.13 %)
Reads Trimmed by quality (20.0): 6522200 (56.92 %)
Bases Trimmed by quality: 122102572 (7.06 %)

B)

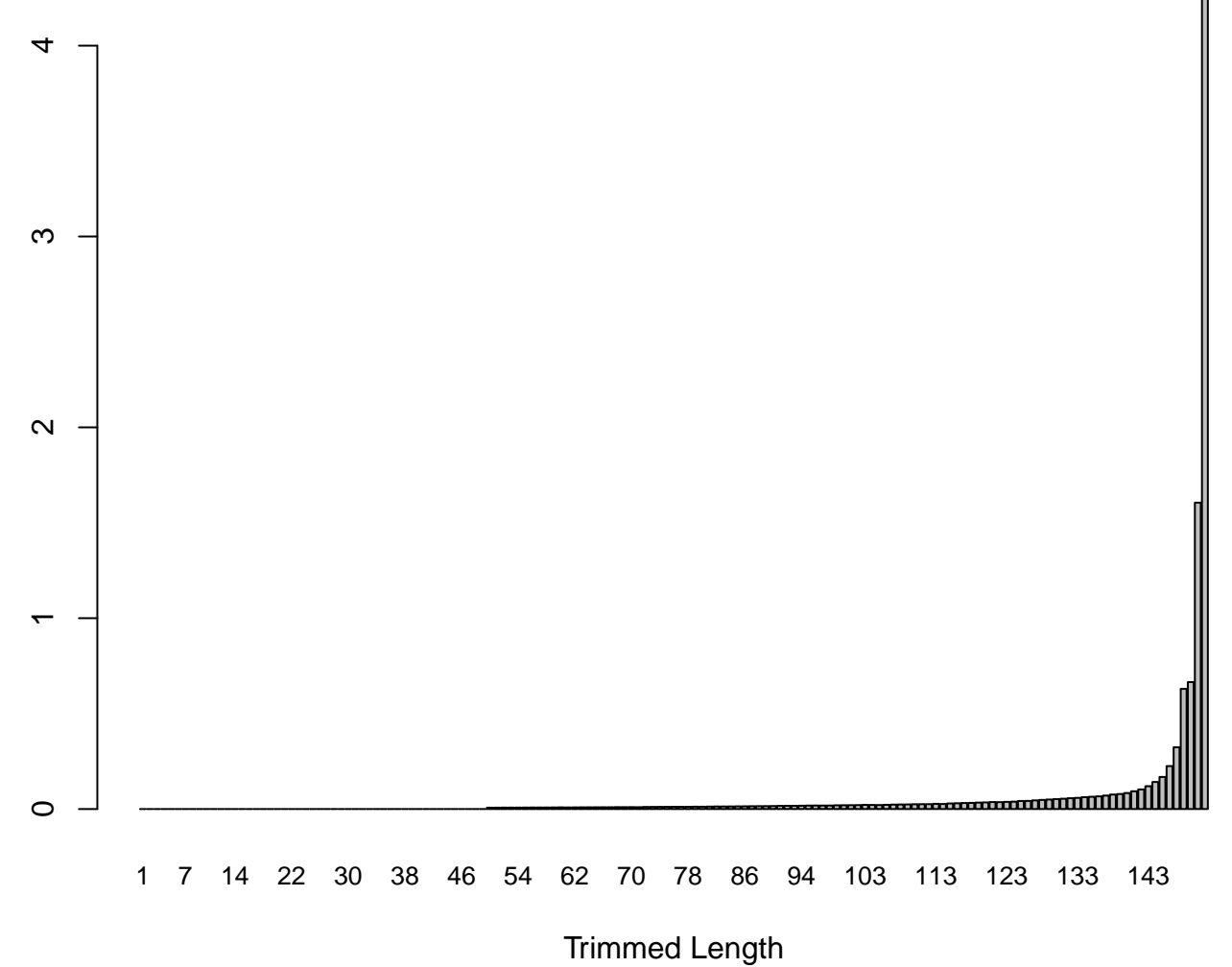
Reads Length Histogram

(Sampling 9.46 M Reads)

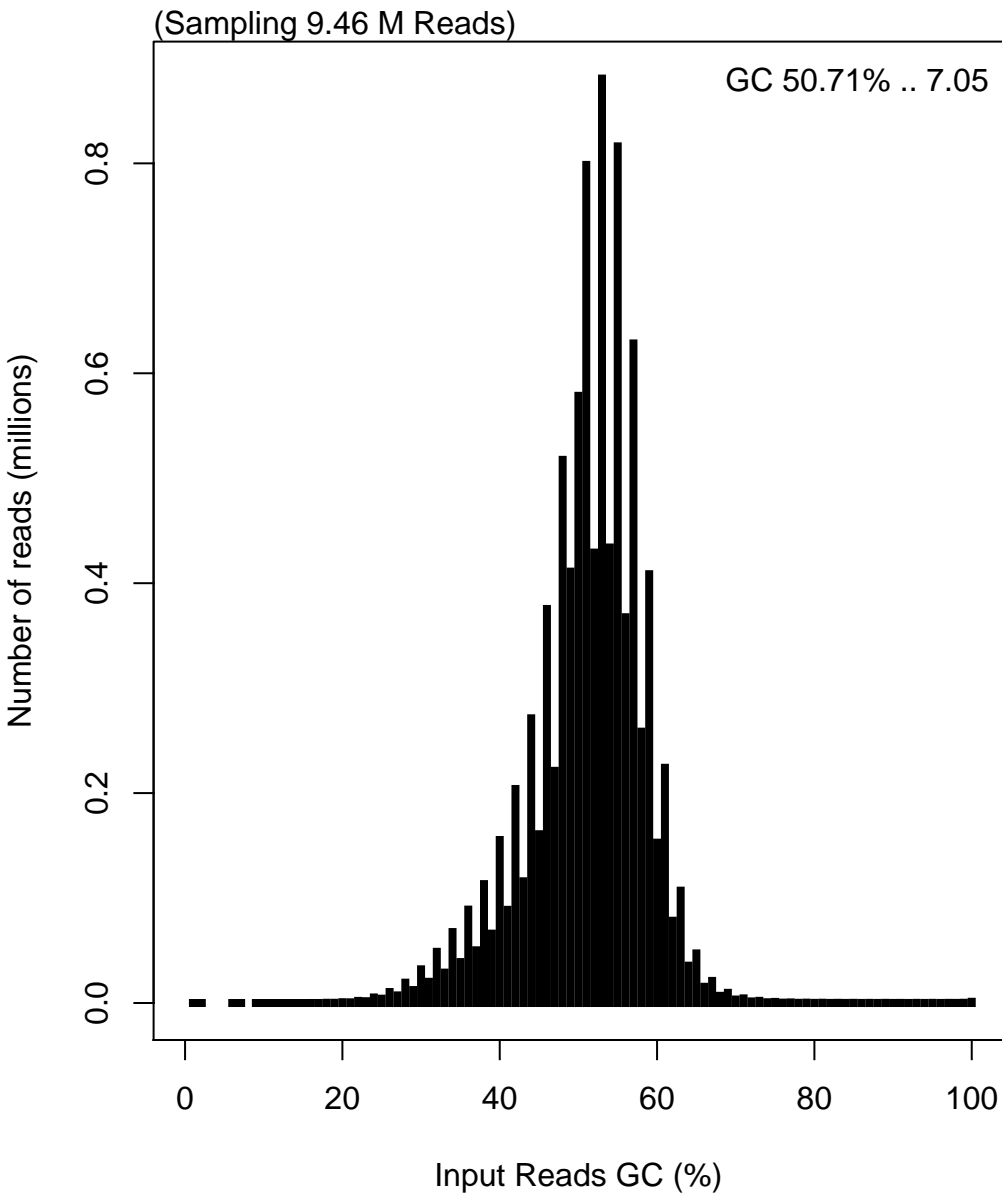
Mean 151.00 .. 0.00
Max 151
Min 151



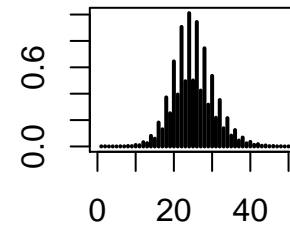
Mean 142.33 .. 18.26
Max 151
Min 50



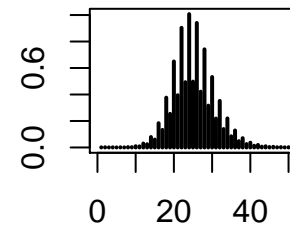
C)



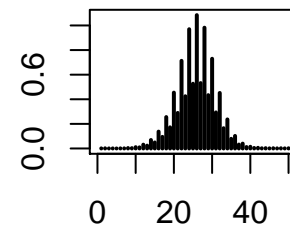
A 24.61% .. 5.41



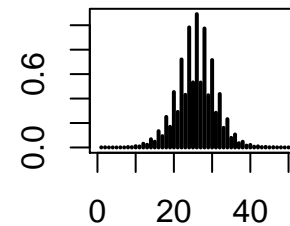
T 24.68% .. 5.41



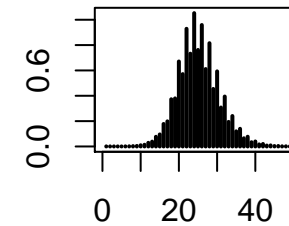
C 25.38% .. 5.17



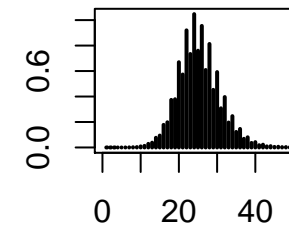
G 25.32% .. 5.04



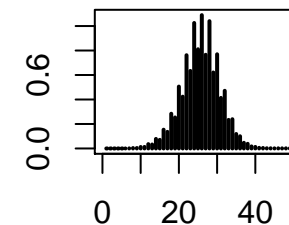
A 24.70% .. 5.37



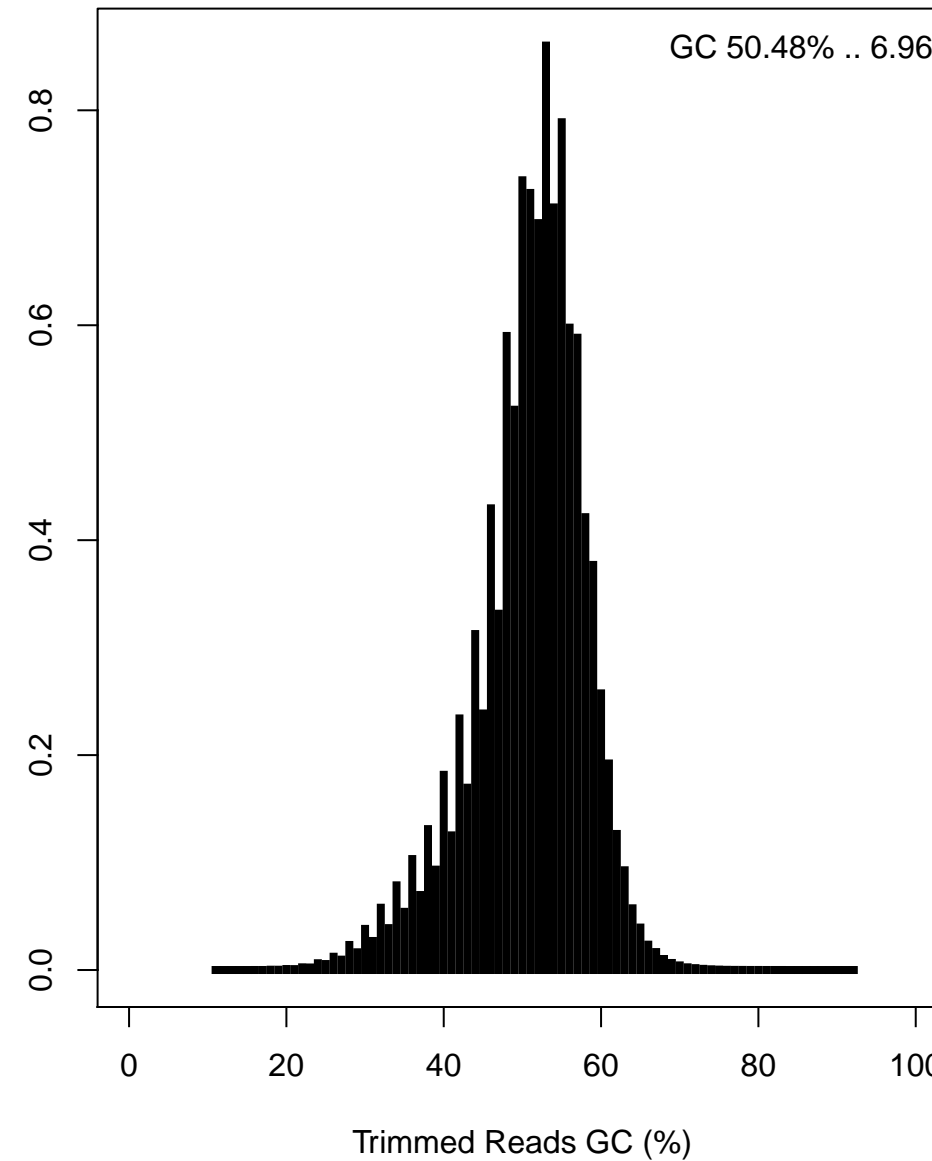
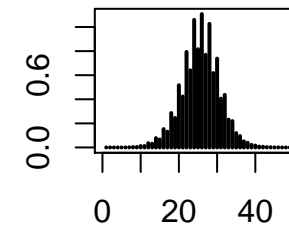
T 24.82% .. 5.40



C 25.30% .. 4.95



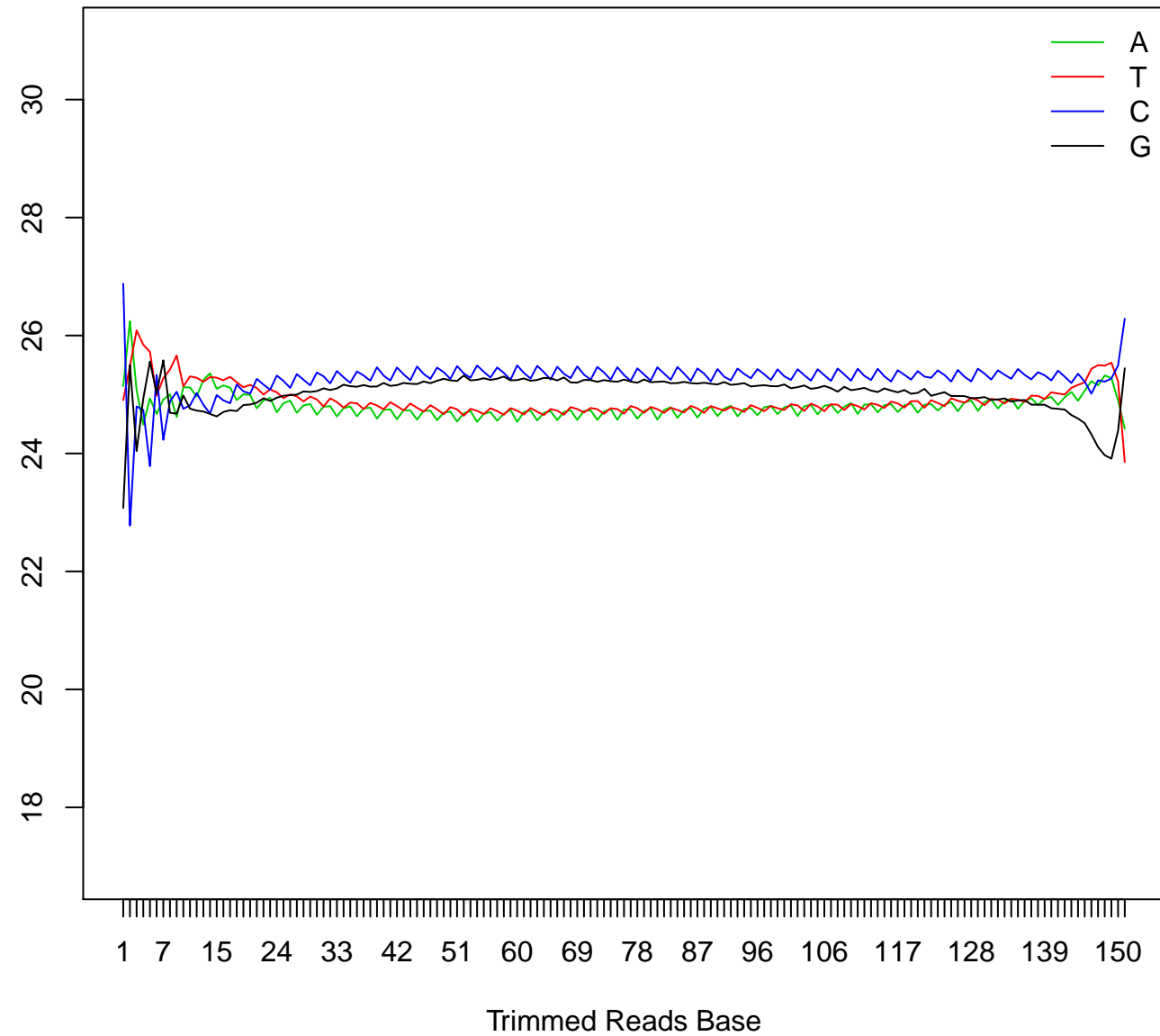
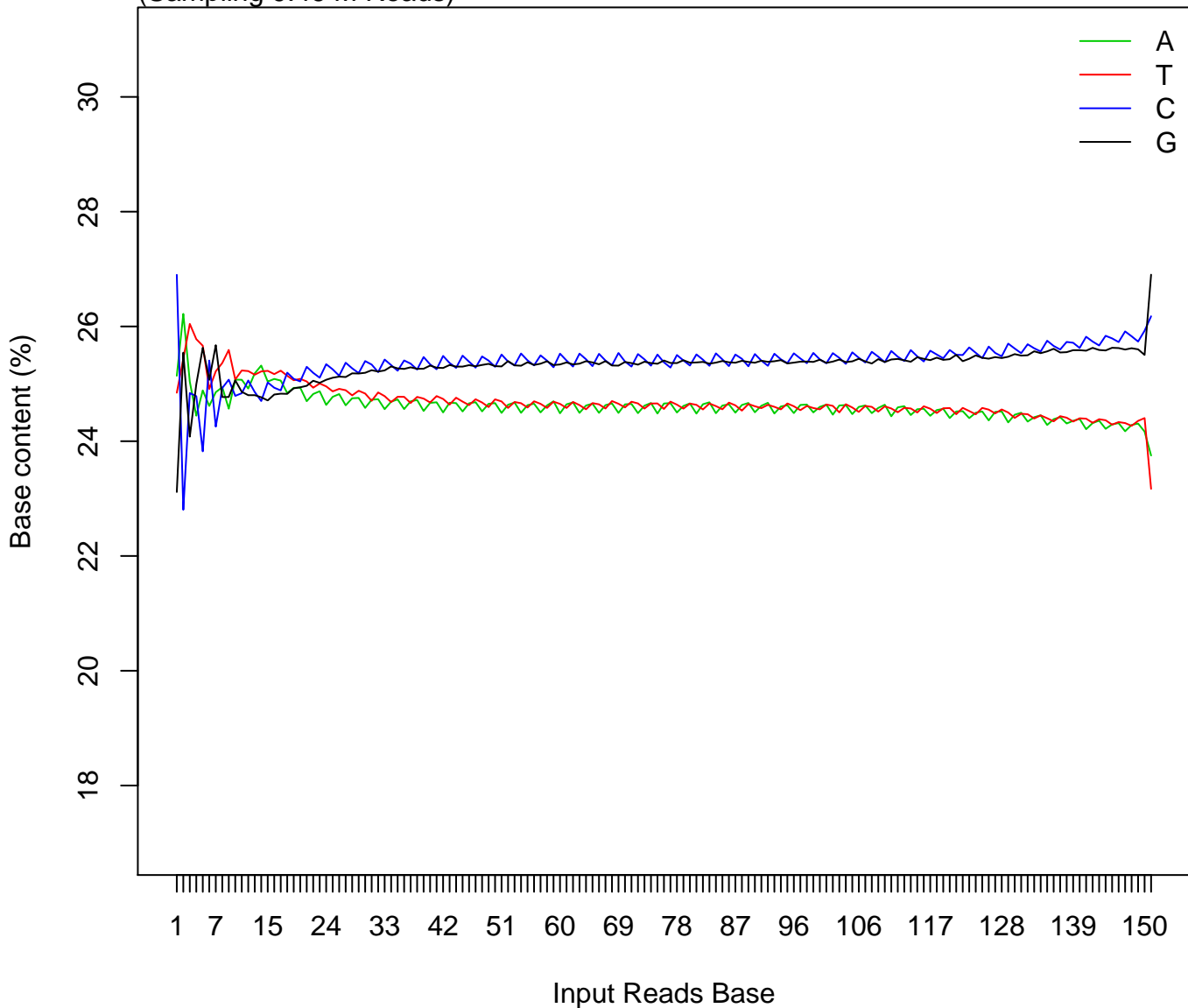
G 25.18% .. 4.93



D)

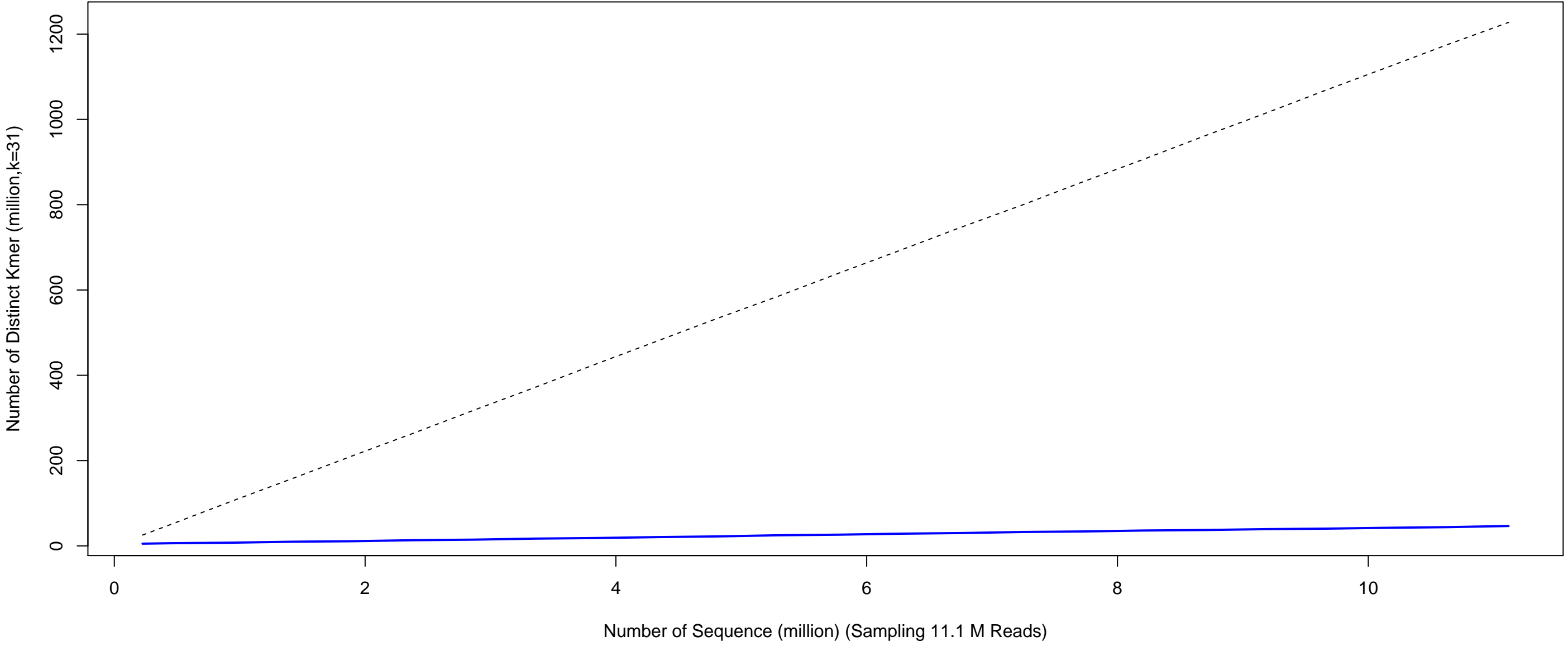
Nucleotide Content Per Cycle

(Sampling 9.46 M Reads)



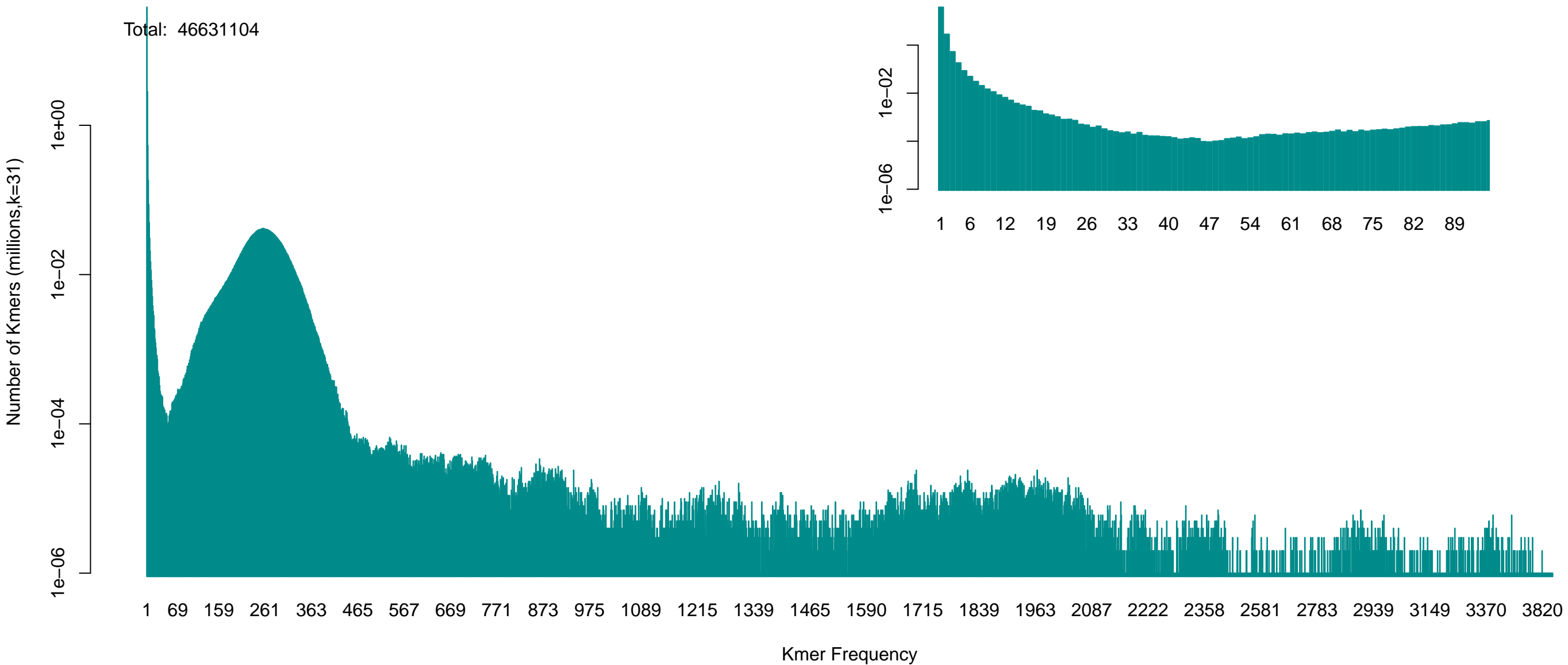
E)

Kmer Rarefaction Curve



F)

Kmer Frequency Histogram (Sampling 11.1 M Reads)

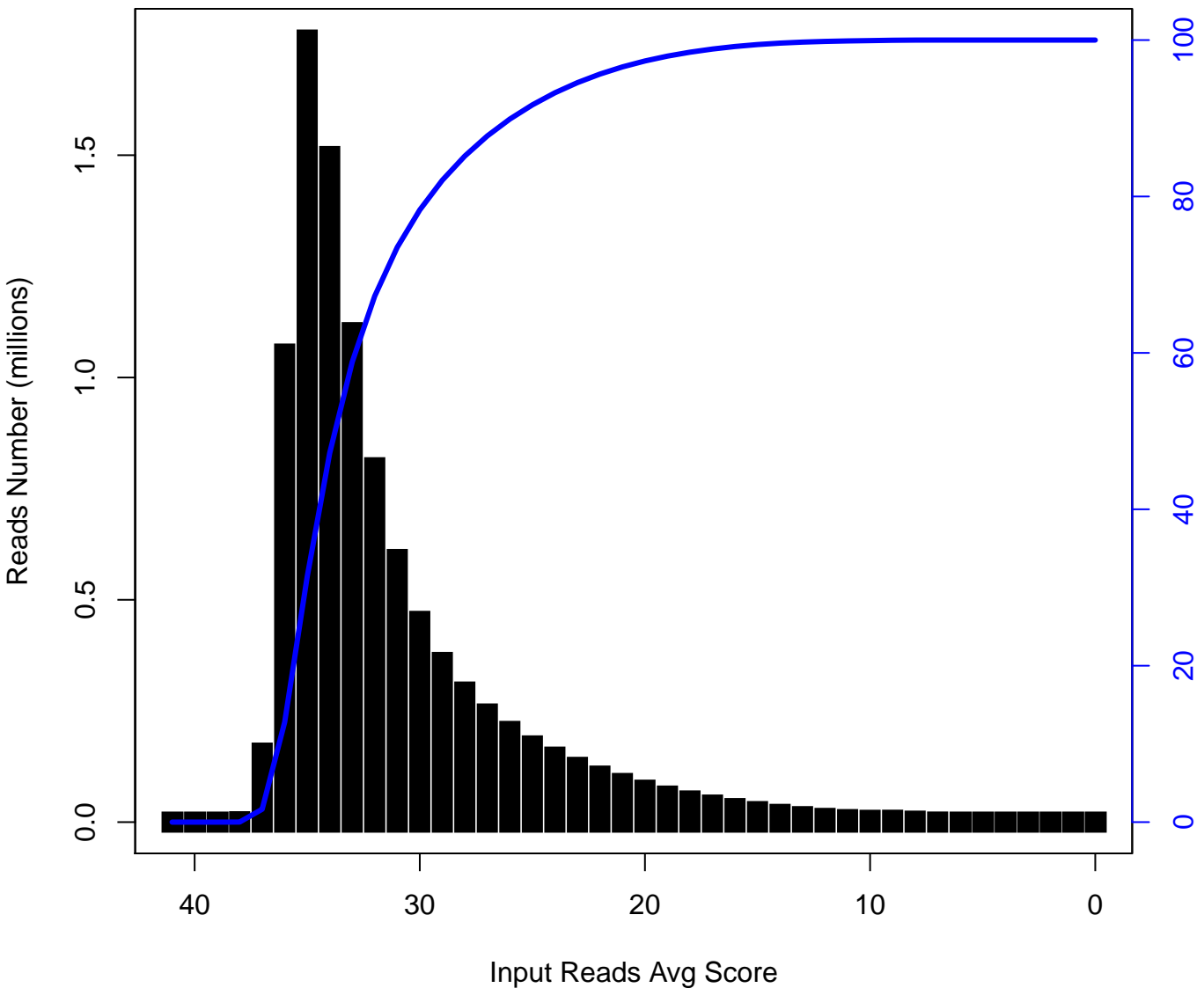


G)

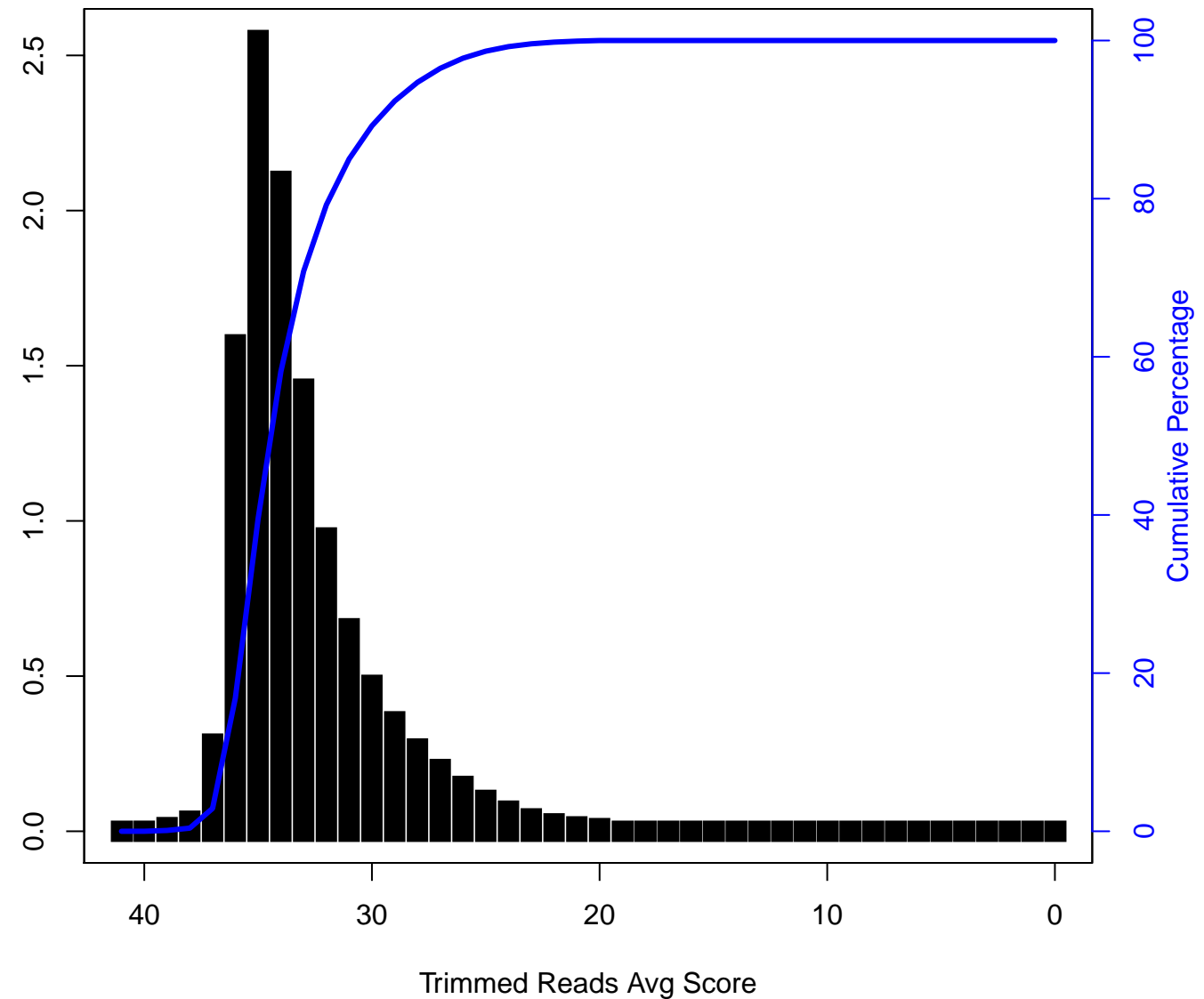
Reads Average Quality Histogram

(Sampling 9.46 M Reads)

Number of Q>=20 reads: 9,206,325 (97.33%) , mean Length: 151.00



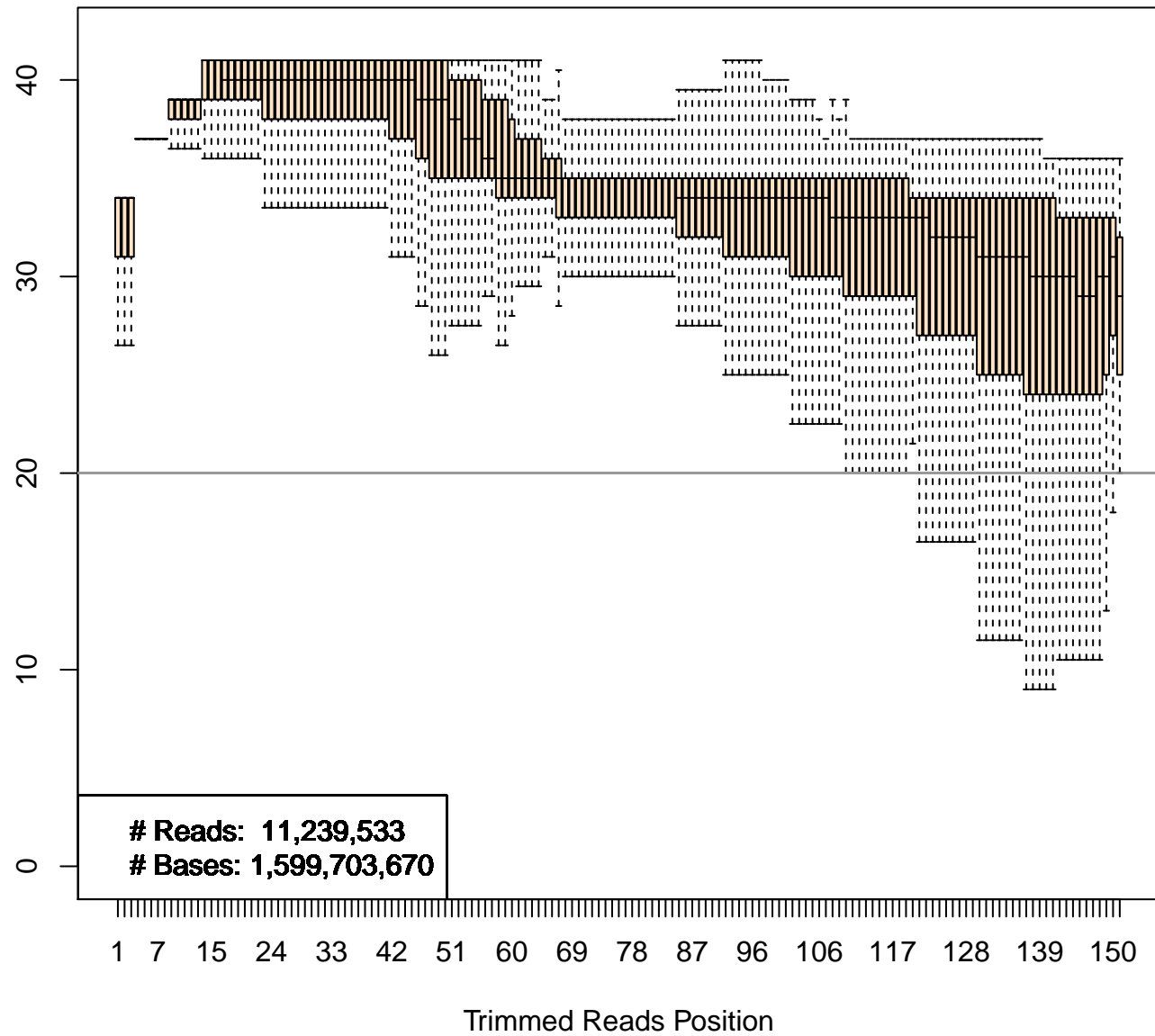
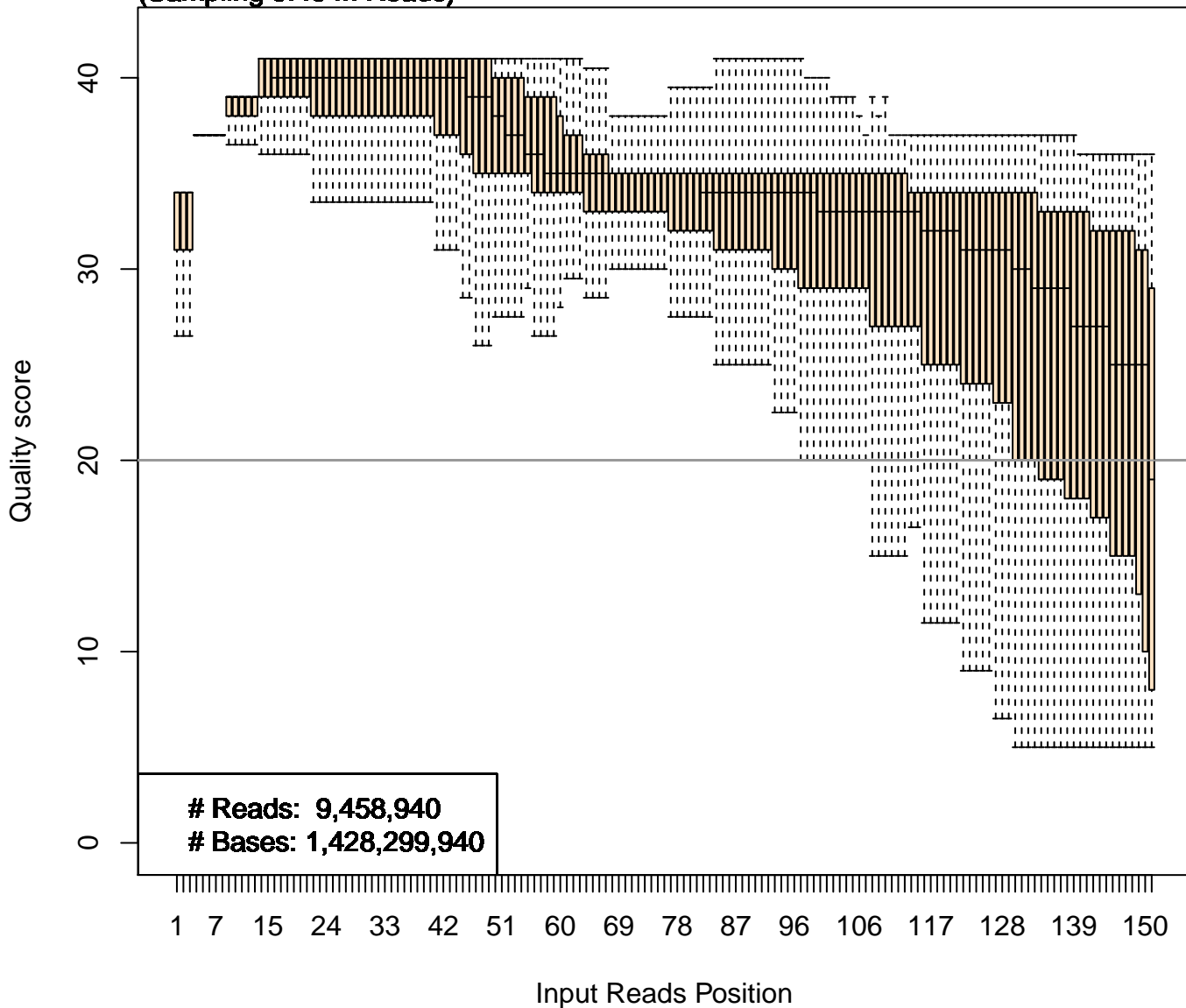
Number of Q>=20 reads: 11,239,533 (100.00%) , mean Length: 142.33



H)

Quality Boxplot Per Cycle

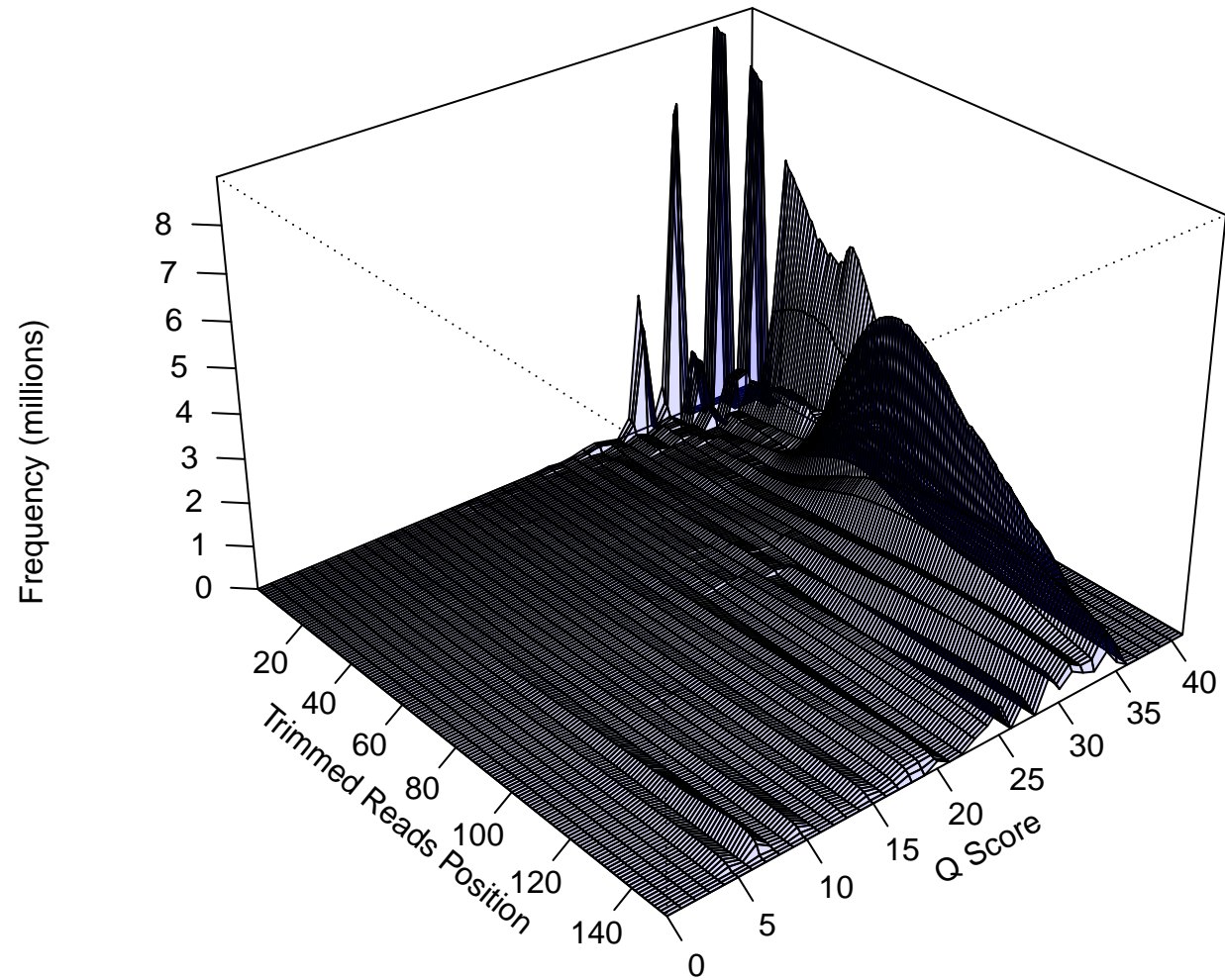
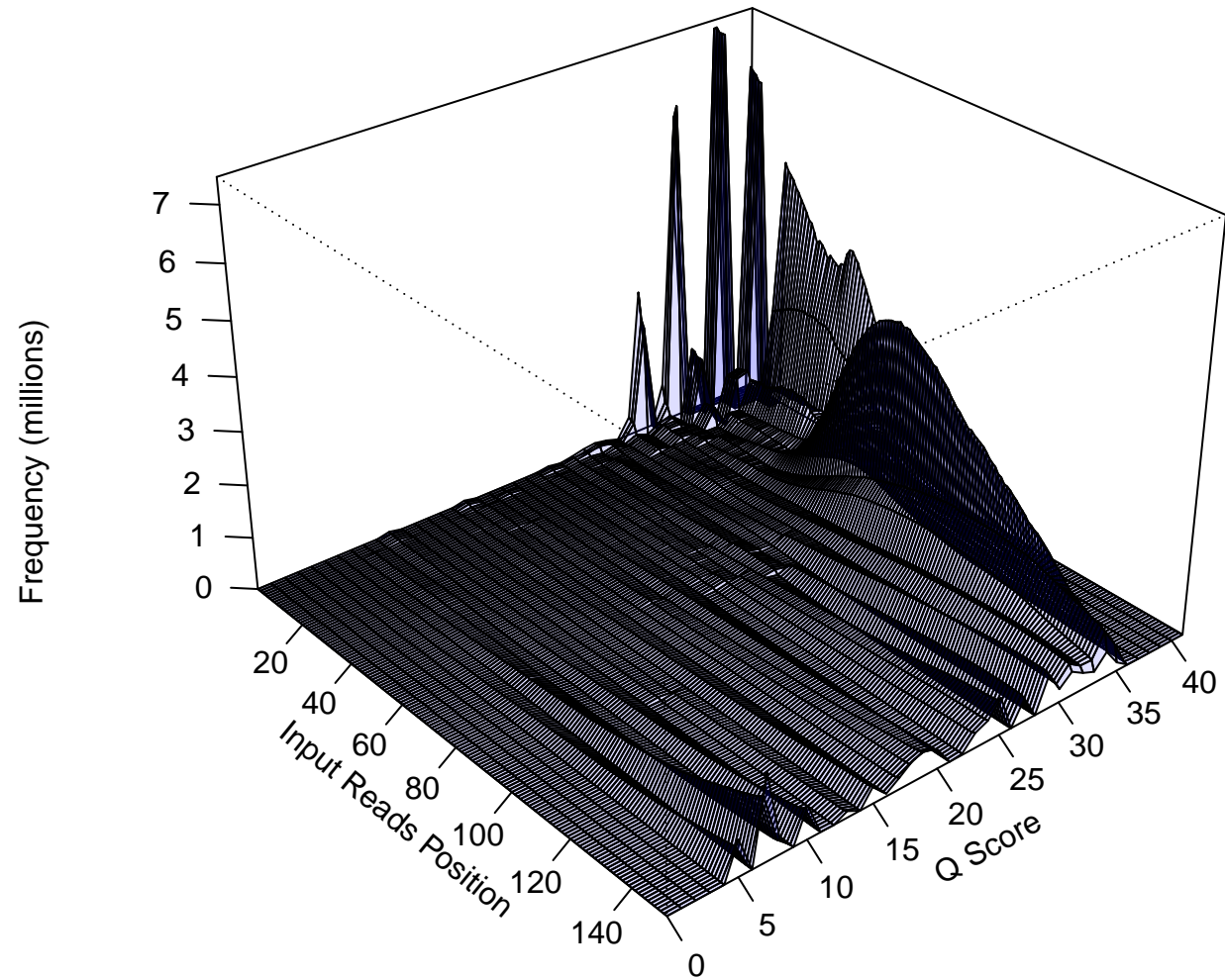
(Sampling 9.46 M Reads)



I)

Quality 3D plot. (Position vs. Score vs. Frequency)

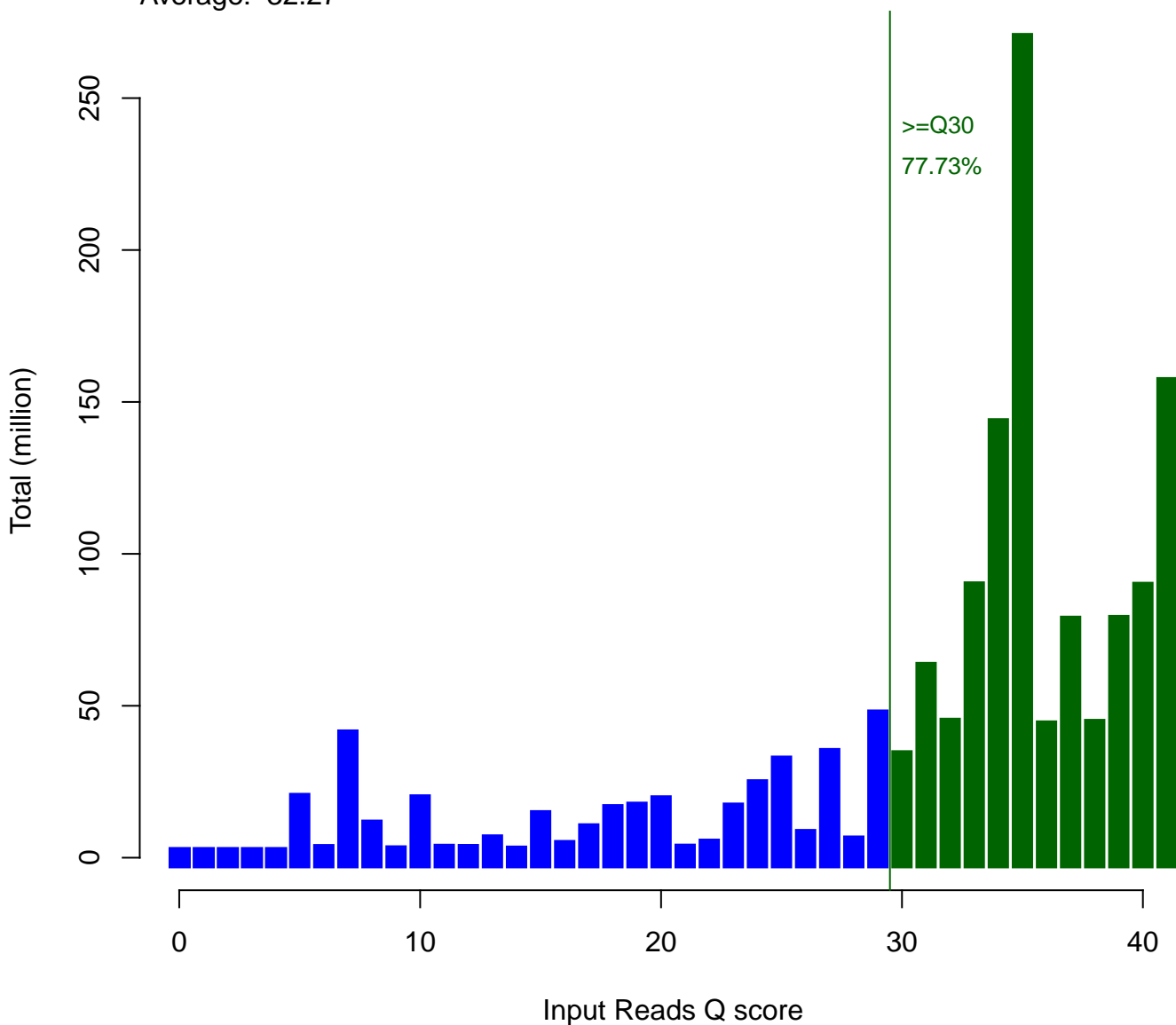
(Sampling 9.46 M Reads)



J)

(Sampling 9.46 M Reads)

Average: 32.27



Quality report

Average: 33.80

