QAA_report

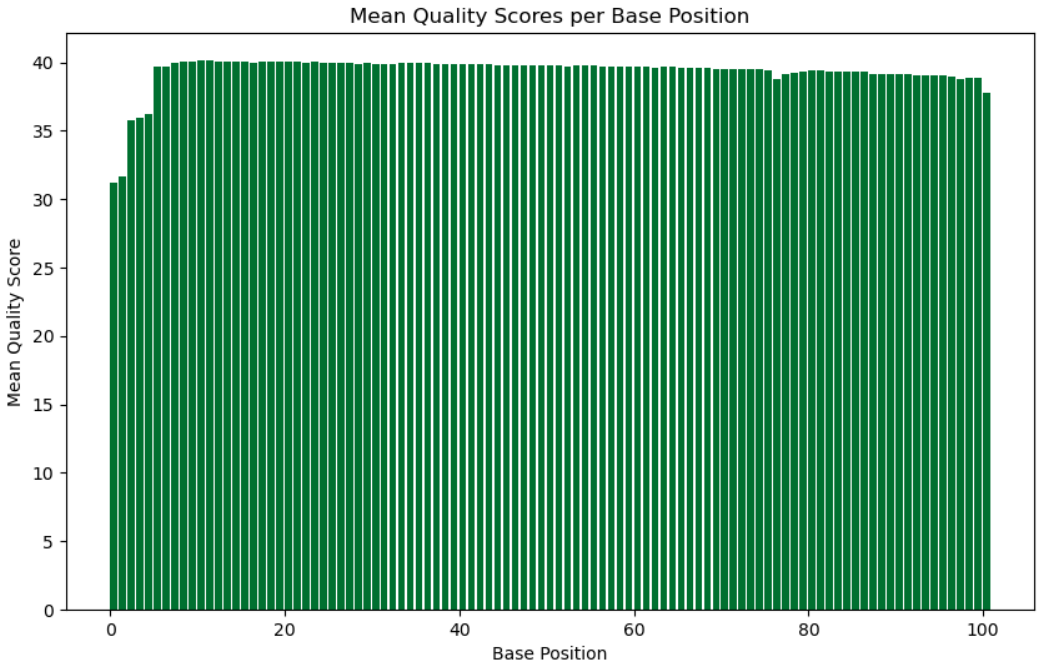
Jack Peplinski

2022-09-06

Plots

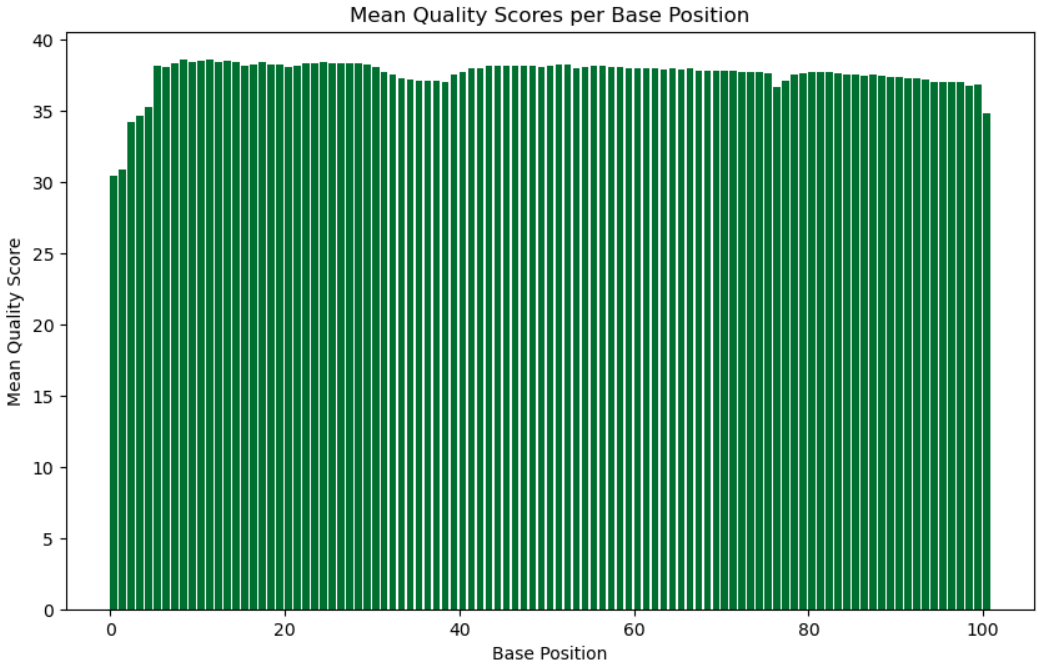
FASTQC Distribution for 27_4C_mbnl_S19_L008

Read 1



The FASTQC distribution for read 1 of 27_4C_mbnl_S19_L008, displaying the base position on the x-axis.

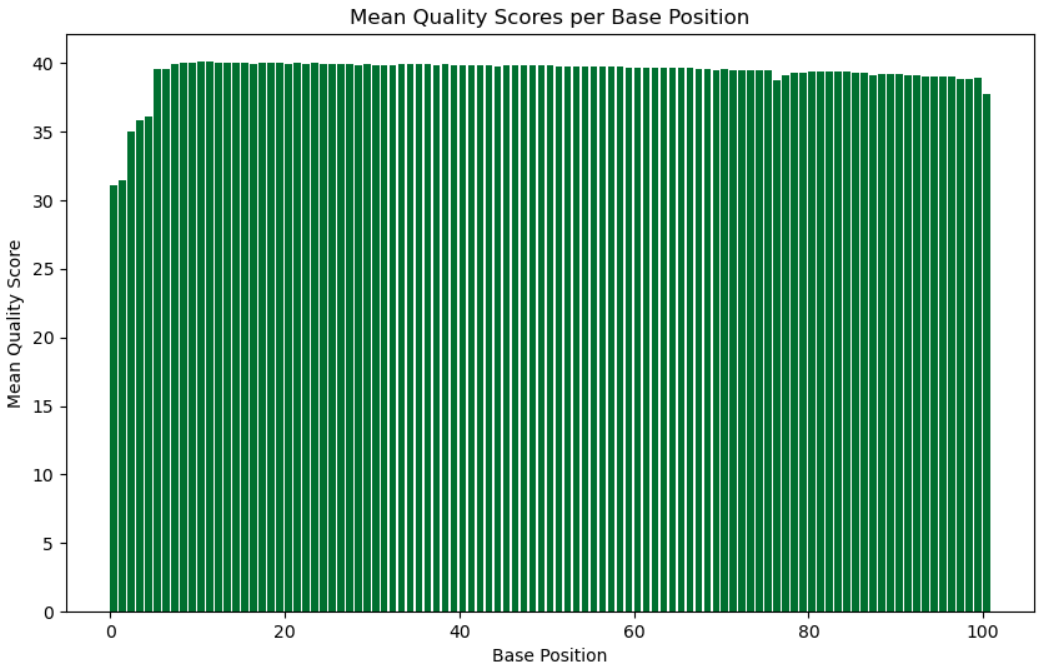
Read 2



The FASTQC distribution for read 2 of 27_4C_mbnl_S19_L008, displaying the base position on the x-axis.

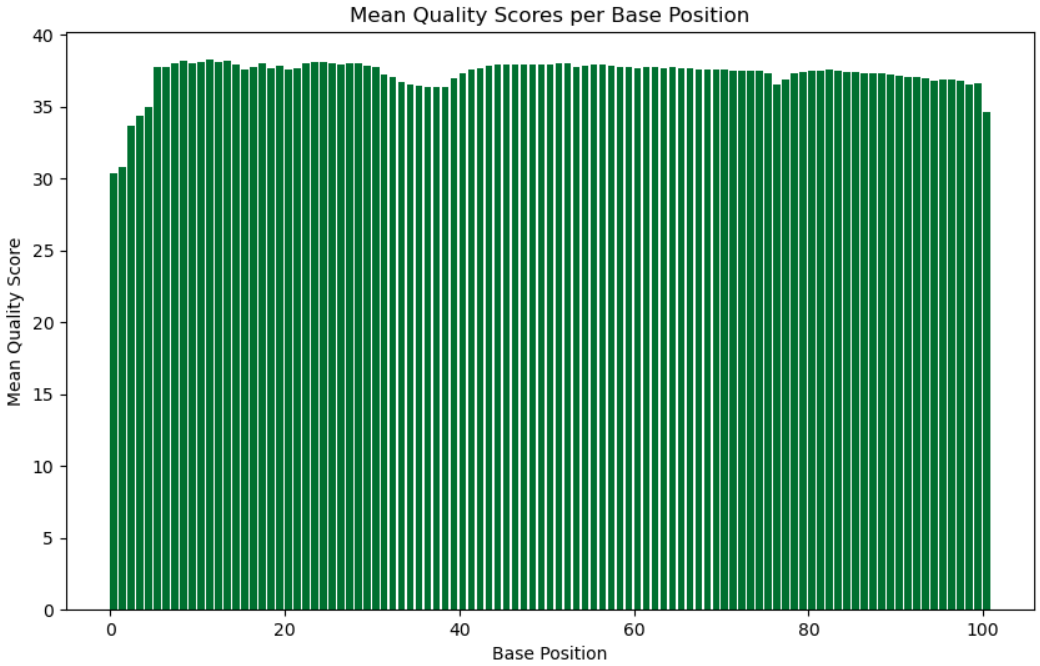
FASTQC Distribution for 28_4D_mbnl_S20_L008

Read 1



The FASTQC distribution for read 1 of 28_4D_mbnl_S20_L008, displaying the base position on the x-axis.

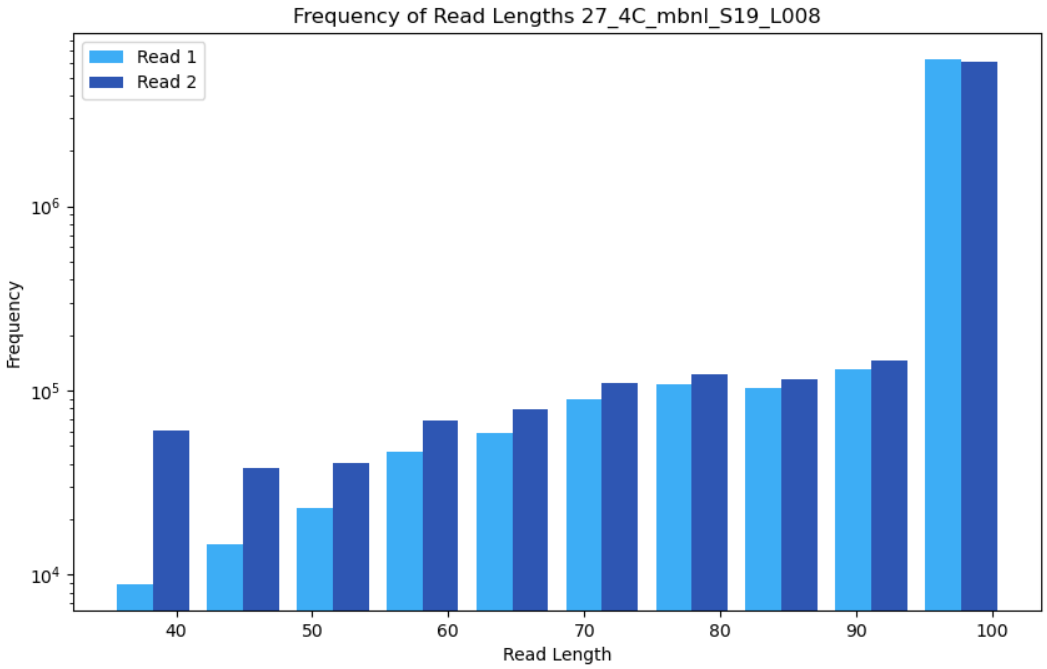
Read 2



The FASTQC distribution for read 1 of 28_4D_mbnl_S20_L008, displaying the base position on the x-axis.

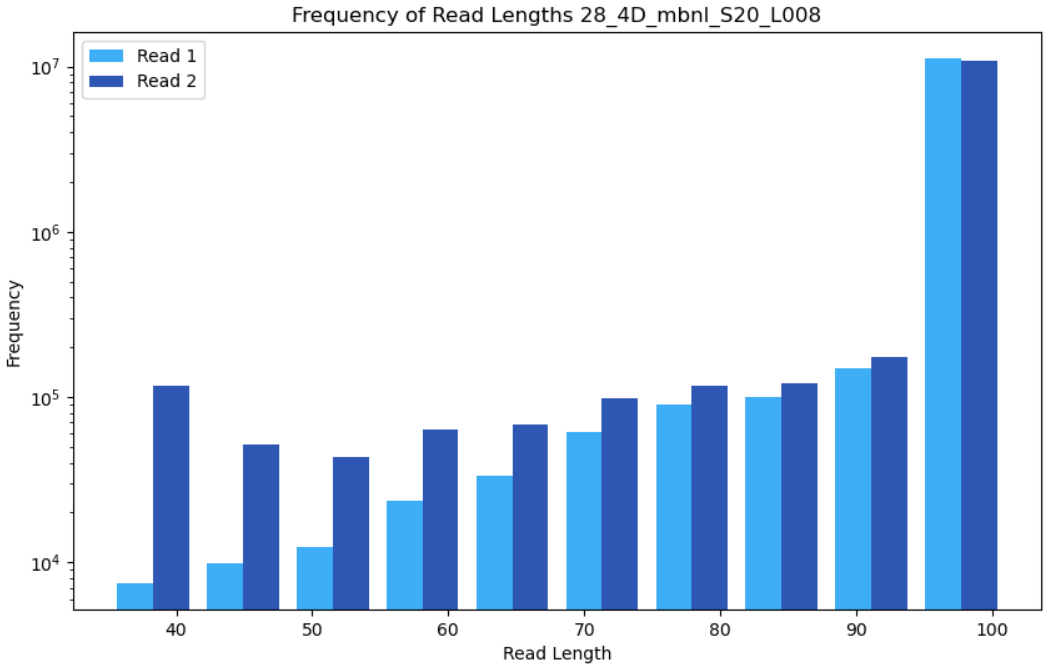
Trimmed Read Length Distributions

27_4C_mbnl_S19_L008



The trimmed read length distributions for 27_4C_mbnl_S19_L008, displaying read 1 in light blue and read 2 in dark blue.

28_4D_mbnl_S20_L008



The trimmed read length distributions for 28_4D_mbnl_S20_L008, displaying read 1 in light blue and read 2 in dark blue.

Answers

Part 1

3. The libraries are overall of high quality. Looking at the plots, R1 for both files are mostly above quality score of 35. For R2 on both files there is a slight dip in quality around the 35 bp and 75 bp regions. For all files, there is a dip in quality at the ends of the bp regions, which is to be expected.

Part 2

5. Adapters trimmed:

For the first file: 27_4C:

Read 1 with adapter: 751,117 (10.4%) Read 2 with adapter: 803,568 (11.1%)

For the second file: 28_4D:

Read 1 with adapter: 743,440 (6.0%) Read 2 with adapter: 841,389 (6.8%)

7. As shown, we should expect R1 and R2s to be trimmed at slightly different rates. This is due to variation in the sequencing process.

Part 3

12. cat 4Coutput.txt | awk '{sum+=$2} END{print sum}' 6876955 cat 4C_rev_output.txt | awk '{sum+=$2} END{print sum}' 6876955 cat 4Doutput.txt | awk '{sum+=$2} END{print sum}' 11725400 cat 4D_rev_output.txt | awk '{sum+=$2} END{print sum}' 11725400

The data are not from strand-specific libraries because the forward and reverse strands have the same number of mapped reads, making up about 50% of the mapped reads.

Read Counts

Read Type	27_4C_mbnl_S19_L008	28_4D_mbnl_S20_L008
Mapped	13320040	22657652
Unmapped	434803	795284
Mapped to Features (Forward Strand)	6876955	11725400
Mapped to Features (Reverse Strand)	6876955	11725400

Counts of mapped and unmapped reads, as well as reads mapped to features in the forward and reverse orientation. The equal read counts in forward and reverse orientation show a strand-unspecific dataset.