QC – Answers to questions

1. Which positions in the reads have a median phred-score above 28 (very good quality calls) in each sample?

Sample	Read 1	Read 2
HG00097	2-101	2-101
HG00100	1-107	1-107
HG00101	1-101	1-101

2. Do any of the samples have warnings or failures in the Per Base Sequence Quality module?

HG00097 Read 1 and Read 2

3. Why?

This module will raise a failure if the lower quartile for any base is less than 5 or if the median for any base is less than 20.

In sample HG00097, the first base have median quality of 15, which is less than the threshold for raising a failure. An option could be to trim off the first base, but since the quality values are taken into account in variant calling it is OK to keep this one base with lower quality in the reads.

4. How do quality scores change for HG00101 after quality trimming?

R1: Higher mean quality overall; fewer low-quality bases from about position 30 and onwards.
R2: Higher quality from about position 85 and onwards.

5. How long are the reads?

Sample	Read 1	Read 2
HG00097	101	101
HG00100	100-108	100-108
HG00101	101	101

6. Do any of the samples have warnings or failures in the Sequence Length Distribution module?

Yes, HG00100

7. Why?

This module will raise a warning if all sequences are not the same length.

For HG00100 most reads are 100 bp, but a subset of the reads are 108 bp.

8. How does the length distribution change for HG00101 after quality trimming?

R1: Lengths 0-101 bp exist. Most are still 101 bp, but some reads even get trimmed to length 0. FastQC warns (but this can be ignored).
R2: Most reads are still 101 bp, but other lengths exist down to 9 bp.

9. What is the first read whose quality was trimmed in the `HG00101_trimmed_2.fq` file?

The read with identifier `ERR229776.12974779/2`.