1. What are the main steps in second generation sequencing, from DNA to reads?

2. Which technologies uses emulsion PCR?

3. Which technology use bridge PCR?

4. What is the main base quality issue seen with Illumina data? Why?

5. What is the main read error in 454 and Ion Torrent data? Why?

6. How many lines is one read in fastq format? What are the lines?

7. What does it mean that a base in a read has a base quality of Q20?

8. Describe the difference between a single end and paired end read.

9. What does it mean to have sequenced a genome to 50X?

10. Mention two issues that can cause problems when aligning reads to a reference genome.

11. Briefly describe the principle of the Seed and Extend algorithm.

12. Create the Burrows-Wheeler Transformation of this sequence “TAGC$”. 
13. Why is alignment of paired end reads more accurate compared to using similar length single end reads?

14. Why do we perform Local Realignment after alignment with e.g. BWA?

15. Why do we perform Base Quality Score Recalibration?

16. Why are we using probabilistic methods (Bayesian) for genotyping?

17. Soft variant filtration (Variant Quality Score Recalibration) uses known polymorphic sites (e.g., HapMap) to?

18. How can we use the truth tranches to assess the quality of our variants?

19. Which method is normally used for assembling Illumina reads?

20. Create the *de Bruijn* graph of this sequence using k=3: ACGTTGGTCGTG

21. How do we create contigs and scaffolds from a *de Bruijn* graph?

22. Why is *de novo* assembly much harder for metagenomic data compared to single genome data?