

## Read alignment

| Question   | Answer 1   | Answer 2  | Answer 3  | Answer 4  | answer(s) - choose at |
|--|--|---|---|---|-----------------------|
| 1 Who maintains the human and other reference genomes?   | NCBI   | Genome Reference Consortium                                     | Does not exist  | UCSC  | 2                     |
| 2 The reference file is a large Fasta file?  | TRUE   | FALSE   |   |   | 1                     |
| 3 What is the optimal alignment?   | An alignment with a MAPQ=60                                    | An alignment with the most differences                          | An alignment with the highest mapping score                               | An alignment with the most matches                                  | 3                     |
| 4 CA-CG is the optimal Smith-Waterman alignment for aligning CACG to CATCG. If the penalty matrix for scoring is a | 1  | 2   | 3   | 4   | 3                     |
| 5 How many 150bp Illumina 10X reads do you expect in a 30X human sequencing project?                               | 800x10^6   | 700x10^6  | 600x10^6  | 500x10^6  | 3                     |
| 6 What is an example of a hash-based aligner?  | MAQ  | Bowtie  | Bowtie2   | bwa   | 1                     |
| 7 What is an example of a BWT aligner?   | DRAGEN Aligner   | MAQ   | Eland   | bwa   | 4                     |
| 8 What is a MAPQ of 30?  | It is a 1/1000 probability that the read is incorrectly mapped | It is a 1/30 probability that the read is incorrectly mapped    | It is a 1/100 probability that the read is incorrectly mapped             | It is a 1/10000 probability that the read is incorrectly mapped     | 1                     |
| 9 What is the difference between SAM, BAM and CRAM?  | There is not difference all are the same alignment format      | alignment format, BAM is the format of the SAM and CRAM is the  | format, BAM is the format of the SAM and CRAM is the encrypted version of | SAM and BAM are alignment formats. CRAM is not an alignment format. | 2                     |
| 10 What factors affect alignment?  | Read Length and complexity of the genome                       | Alignment around indels   | High density SNP regions  | All of the above  | 4                     |
| 11 What is the MAPQ score of an alignment that is placed equally on the reference elsewhere?                       | 30   | No score  | 0   |   | 1 3                   |
| 12 What type of sequencing is preferred in cases where you have repeats in the reference?                          | single end sequencing  | paired end sequencing   | none of the above   |   | 2                     |
| 13 A 60Mbp bwa alignment takes 6 minutes. 1 human genome sequenced at 45x = 150 Gbp per HiSeqX10 lane. How         | 1.25 hours   | 1 hour  | .5 hours  | 2 hours   | 2                     |
| 14 What important steps are involved in BAM improvement?   | Marking duplicates and base quality recalibration              | Alignment around indels   | All of the above  | None of the above   | 1                     |
| 15 What is needed in the base quality score recalibration process?   | The reference  | A list of known SNP sites                                       | The alignments to be recalibrated   | All of the above  | 4                     |
| 16 What is the cause of library duplicates?  | Misalignment   | Library merging   | Regions of SNPs near INDELs   | The PCR amplification step in library preparation                   | 4                     |
| 17 Which of these commands creates necessary BWA indexes?  | bwa index<br>GRCm38.68.dna.toplevel.chr7.fa.gz                 | samtools index md5638.sorted.bam                                | ../ref/GRCm38.68.dna.toplevel.chr7.fa.gz md5638a_7_87000000_R1.fastq.gz   | None of the above   | 1                     |
| 18 What does this command do? bwa mem ../ref/GRCm38.68.dna.toplevel.chr7.fa.gz md5638a_7_87000000_R1.fastq         | format and then sorted by coordinates                          | format, then converted to BAM and then sorted by coordinates    | format and then converted to an unsorted BAM                              | None of the above   | 2                     |
| 19 What does the -R flag specify in the bwa command?   | It specifies the reference                                     | It specifies the ReadGroup information                          | It specifies the output format  | None of the above   | 2                     |
| 20 Which of these commands merge lanes and then mark duplicates on them?   | merge-lanes lane1/lane1.sorted.bam, lane2/lane2.sorted.bam -O  | merge-lanes lane1/lane1.sorted.bam -l lane2/lane2.sorted.bam -O | merge-lanes lane1/lane1.sorted.bam -l lane2/lane2.sorted.bam -O           | None of the above   | 2                     |
| 21   |  |   |   |   |                       |