22126 — NGS Analysis

Dummy / Recap Exam (Practice Only)

ANSWERS IN BOLD

Q1

Which statement best describes why next-generation sequencing (NGS) was revolutionary compared to Sanger sequencing?

A. NGS produces much longer reads
B. NGS sequences single molecules without amplification
C. NGS massively parallelises sequencing reactions
D. NGS eliminates sequencing errors

Q2

Which feature distinguishes third-generation sequencing technologies from Illumina-style short-read sequencing?

A. Use of fluorescently labelled nucleotides
B. Sequencing by synthesis on a flow cell
C. Ability to generate very long reads from single molecules
D. Reliance on PCR amplification for cluster generation

Q3

Which sequencing error type is most commonly observed in Illumina data?

A. Large insertions
B. Homopolymer-length errors
C. Base substitutions
D. Chromosomal rearrangements

Q4

What is the main purpose of paired-end sequencing?

A. To reduce sequencing cost
B. To infer base quality scores
C. To provide information about fragment orientation and distance
D. To eliminate PCR duplicates

Q5

Which file format typically stores raw sequencing reads and base quality scores?

A. SAM
B. BAM
C. FASTA
D. FASTQ

Q6

What does a PHRED quality score measure?

A. Probability a read is incorrectly mapped
B. Probability a base call is wrong
C. Probability a variant is false
D. Probability a sample is contaminated

Q7

Which factor(s) most strongly influences genome-wide sequencing coverage?

A. Read mapping quality
B. Read length, read count, and genome size
C. GC content alone
D. Duplicate rate alone

Q8

Why is quality control (QC) typically performed before read alignment?

A. Alignment improves base quality
B. QC reduces computational cost and downstream artefacts
C. QC assigns genomic coordinates
D. QC performs variant calling

Q9

Which problem is most likely caused by adapter contamination?

A. Incorrect variant genotypes
B. Artificially long reads
C. Poor alignment at read ends
D. Low mapping quality across the entire genome

Q10

What is the primary role of a read aligner?

A. Assemble genomes
B. Identify sequencing errors
C. Determine where reads most likely originate in a reference genome
D. Predict gene function

Q11

Why are repetitive regions challenging for short-read alignment?

A. Reads from repeats have lower base quality
B. Reads can map equally well to multiple locations
C. Repeats are not sequenced efficiently
D. Repeats increase sequencing depth

Q12

Which quantity reflects confidence in where a read is placed in the genome?

A. Base quality
B. Coverage
C. Mapping quality (MAPQ)
D. GC content

Q13

Which step typically comes first in a standard variant-calling workflow?

A. Variant filtering
B. Genotyping
C. Read alignment
D. Annotation

Q14

Why does low sequencing depth reduce confidence in variant calls?

A. Base qualities decrease
B. Mapping qualities decrease
C. There are fewer independent observations supporting alleles
D. Reads become shorter

Q15

What is the key difference between alignment and de novo assembly?

A. Alignment uses long reads only
B. Assembly requires a reference genome
C. Alignment places reads on an existing reference; assembly reconstructs

sequences without one
D. Assembly cannot be used for microbes

Q16

Which assembly metric describes the contig length at which half of the genome assembly is contained?

A. Coverage
B. GC content
C. N50
D. BUSCO

Q17

Why do raw variant call files (VCFs) typically require post-processing / filtering?

A. Raw VCFs contain only coding variants
B. Raw VCFs often include false positives from sequencing or alignment artefacts
C. Filtering increases sequencing coverage
D. Filtering converts FASTQ files into BAM files

Q18

Which statement best describes what Hi-C data measures?

A. DNA sequence variation
B. Gene expression levels
C. Physical contacts between genomic regions
D. DNA methylation

Q19

Which characteristic is typical of ancient DNA samples?

A. Long fragment lengths
B. High sequencing depth
C. Chemical damage such as cytosine deamination
D. Low GC content

Q20

Why is it generally unsafe to conclude genotype from a single read at a genomic position?

A. Sequencing always produces errors
B. One read provides insufficient evidence for genotype inference
C. Mapping quality is always zero
D. Variant callers ignore single reads