Applied Genomics – Multiple Choice Questions (MCQs)

Martina Castellucci

University of Bologna – Master in Bioinformatics

Contents

Instructions

Each question has four options (A–D). The correct answer is marked with **(*)** and a brief explanation is provided. References point to the course notes or lecture slides.

1 Classical Genetics and Mendelian Principles

Q1. Mendel's law of segregation states that:

(A) Each allele from a parent remains linked in the offspring

(B) Alleles segregate randomly during gamete formation **(*)**

(C) Genes on the same chromosome segregate independently

(D) Phenotype is always dominant over genotype

Explanation: Alleles of a gene separate during meiosis so that each gamete carries only one allele. *Ref:* Lecture 1, Slide 15.

Q2. In a dihybrid cross of two heterozygous individuals (AaBb x AaBb), the expected phenotypic ratio is:

(A) 3:1

(B) 9:3:3:1 **(*)**

(C) 1:2:1

(D) 2:1

Explanation: Independent assortment of two genes gives the 9:3:3:1 phenotypic ratio. *Ref:* AG_sbobins p. 21.

Q3. In a pedigree chart, a half-filled circle represents:

(A) A healthy male

(B) A healthy female

(C) A female carrier **(*)**

(D) An affected female

Explanation: By convention, circles = females; half-filled = carrier (heterozygote). *Ref:* Lecture 1, Pedigree slides.

2 Population Genetics

Q4. Hardy-Weinberg equilibrium assumes all of the following EXCEPT:

(A) Random mating

(B) Infinite population size

(C) Natural selection **(*)**

(D) No mutation

Explanation: H-W assumes no selection, mutation, migration, or drift. *Ref:* AG_sbobins, Population genetics section.

Q5. If allele A has a frequency of 0.6, what is the expected frequency of heterozygotes under H-W?

(A) 0.16

(B) 0.24

(C) 0.48 (*)

(D) 0.36

Explanation: Heterozygote frequency = 2pq = 2*0.6*0.4 = 0.48. *Ref:* Lecture 1, H-W example.

3 Molecular Genetics and Sequencing

Q6. The first-generation sequencing method is:

(A) Illumina sequencing

(B) Nanopore sequencing

(C) Sanger dideoxy sequencing (*)

(D) SOLiD sequencing

Explanation: Sanger sequencing was the classical 1st-generation method using chain terminators. *Ref:* Lecture 2, Sanger slides.

Q7. Which NGS technology detects changes in pH?

(A) Illumina

(B) PacBio

(C) Ion Torrent (*)

(D) Nanopore

Explanation: Ion Torrent detects H^+ release (pH change) during nucleotide incorporation. *Ref:* AG_sbobins p. 112.

Q8. A Phred score of 30 corresponds approximately to:

(A) 1/10 error rate

(B) 1/100 error rate

(C) 1/1000 error rate (*)

(D) 99% error rate

Explanation: Q = -10 $\log_{10}$P $\rightarrow$ Q30 = 0.001 error probability = 99.9% accuracy. *Ref:* Lecture 4, FASTQ slide.

4 NGS Data Analysis and QC

Q9. Which tool is most commonly used for initial quality control of FASTQ files?

(A) BWA

(B) FastQC **(*)**

(C) GATK

(D) SAMtools

Explanation: FastQC provides modular per-base and per-sequence quality plots. *Ref:* Lecture 4, QC module.

Q10. The correct order in a simple variant discovery pipeline is:

(A) Variant calling → Alignment → QC

(B) QC → Alignment → Variant calling **(*)**

(C) Alignment → QC → Variant calling

(D) QC → Variant calling → Alignment

Explanation: First check data quality, then align, then call variants. *Ref:* AG_sbobins, NGS pipeline.

5 Genome Assembly and Comparative Genomics

Q11. De Bruijn graphs in genome assembly use:

(A) Reads as nodes

(B) K-mers as edges **(*)**

(C) Reads as edges

(D) Contigs as nodes

Explanation: In a de Bruijn graph, (k-1)-mers are nodes and k-mers form edges. *Ref:* Lecture 6, assembly section.

Q12. Which strategy can improve assembly contiguity in repetitive regions?

(A) Using only short Illumina reads

(B) Increasing coverage with short reads

(C) Using long reads (PacBio/ONT) **(*)**

(D) Ignoring repeats

Explanation: Long reads span repeats and reduce fragmentation. *Ref:* AG_sbobins genome assembly notes.

6 Transcriptomics and Functional Genomics

Q13. Which RNA fraction is typically enriched for mRNA-Seq in eukaryotes?

(A) rRNA

(B) Poly-A RNA (*)

(C) tRNA

(D) snRNA

Explanation: Poly-A selection enriches mature mRNAs and removes most rRNA. *Ref:* Lecture 7, RNA-Seq library prep.

Q14. What is the main challenge in RNA-Seq read alignment?

(A) Sequencing errors

(B) Short read length

(C) Splice junctions (*)

(D) GC content

Explanation: Reads often span exon-exon junctions requiring splice-aware mappers. *Ref:* AG_sbobins transcriptomics.

Q15. Which tool is typically used for differential gene expression analysis?

(A) GATK

(B) DESeq2 (*)

(C) MAUVE

(D) OrthoFinder

Explanation: DESeq2 and edgeR are standard R packages for RNA-Seq DEGs. *Ref:* Lecture 8, RNA-Seq analysis.

7 GWAS, CNV, and Population Genomics

Q16. In a case-control GWAS, which plot is typically used to visualize significant associations?

(A) Synteny plot

(B) Manhattan plot (*)

(C) Phylogenetic tree

(D) PCA plot

Explanation: Manhattan plots show SNP p-values along the genome. *Ref:* Lecture 10, GWAS slide.

Q17. The fixation index (F_{ST}) measures:

(A) Allele frequency in one population

(B) Differentiation between populations (*)

(C) Inbreeding within a single individual

(D) Number of heterozygotes

Explanation: F_{ST} compares variance between vs. within populations. *Ref:* AG_sbobins, Pop Genomics.

Q18. Which technique is best to detect copy number variations genome-wide?

(A) ChIP-Seq

(B) aCGH (*)

(C) RNA-Seq

(D) ATAC-Seq

Explanation: Array-CGH compares hybridization signals to detect gains/losses. *Ref:* Lecture 11, CNV slides.

8 Applied Genomics and Bioinformatics Tools

Q19. Which tool is used for orthology and comparative genomics?

(A) STAR

(B) OrthoFinder (*)

(C) BUSCO

(D) DESeq2

Explanation: OrthoFinder clusters orthologs and infers species trees. *Ref:* Lecture 9, Comparative Genomics.

Q20. BUSCO is primarily used to:

(A) Detect SNPs

(B) Assess genome completeness (*)

(C) Identify CNVs

(D) Perform de novo assembly

Explanation: BUSCO searches for universal single-copy orthologs. *Ref:* Lecture 6, Assembly QC.

Q21. AntiSMASH is specialized for:

(A) Variant calling

(B) Biosynthetic gene cluster prediction (*)

(C) RNA-Seq alignment

(D) Contamination removal

Explanation: AntiSMASH annotates secondary metabolite clusters (BGCs). *Ref:* Lecture 9, Functional genomics.

9 Epigenomics and Chromatin Analysis

Q22. The main principle of ChIP-Seq is:

(A) Sequencing cDNA molecules

(B) Capturing DNA fragments bound by proteins of interest (*)

(C) Detecting methylated cytosines

(D) Measuring RNA expression

Explanation: ChIP-Seq uses antibodies to immunoprecipitate DNA-protein complexes and sequences the bound DNA. *Ref:* Lecture 10, ChIP-Seq slides.

Q23. Which chemical treatment is used in bisulfite sequencing to detect methylation?

(A) Formaldehyde

(B) Sodium bisulfite (*)

(C) Proteinase K

(D) DNase I

Explanation: Sodium bisulfite converts unmethylated cytosines to uracils, leaving 5-mC unchanged. *Ref:* AG_sbobins, Methyl-Seq notes.

Q24. In methyl-seq analysis, a C that remains a C after bisulfite treatment means:

(A) Conversion failed

(B) It was methylated (*)

(C) It was unmethylated

(D) It is a sequencing error

Explanation: Methylated cytosines resist bisulfite conversion. *Ref:* Lecture 10, DNA methylation slides.

10 Metagenomics and Environmental Genomics

Q25. The main advantage of shotgun metagenomics over 16S rRNA sequencing is:

(A) Lower cost

(B) It only targets bacteria

(C) Provides functional and taxonomic information (*)

(D) Requires pure cultures

Explanation: Shotgun metagenomics sequences all DNA to identify both species and gene functions. *Ref:* AG_sbobins, Metagenomics section.

Q26. Which step is NOT typical in a metagenomics workflow?

(A) DNA extraction from mixed sample

(B) Library preparation

(C) Taxonomic profiling

(D) Sanger capillary electrophoresis (*)

Explanation: Modern metagenomics uses NGS, not first-generation Sanger. *Ref:* Lecture 12, Environmental genomics.

Q27. The term "microbiome" refers to:

(A) Only bacterial DNA in soil

(B) The set of genes of a microbial community (*)

(C) Only culturable microorganisms

(D) Fungal spores in the environment

Explanation: Microbiome = collective genomes of all microbes in a niche. *Ref:* AG_sbobins p. 256.

11 Hybrid Sequencing and Assembly Statistics

Q28. The main advantage of hybrid genome assembly is:

(A) It avoids sequencing errors

(B) Combines long reads for contiguity and short reads for accuracy (*)

(C) Requires no polishing

(D) It only uses Illumina data

Explanation: Hybrid assembly leverages ONT/PacBio long reads to span repeats and Illumina reads to polish errors. *Ref:* Lecture 6, Assembly slides.

Q29. N50 is defined as:

(A) The number of reads covering 50% of the genome

(B) The length at which 50% of the assembly is contained in contigs of that size or longer (*)

(C) The average read length

(D) The coverage of longest contig

Explanation: N50 is a measure of assembly contiguity: 50% of genome is in contigs N50 length. *Ref:* AG_sbobins assembly metrics.

Q30. L50 represents:

(A) 50% GC content

(B) The minimum number of contigs covering 50% of genome (*)

(C) Number of long reads over 50 kb

(D) Lowest base quality score

Explanation: L50 counts the fewest contigs needed to cover half the total genome size. *Ref:* Lecture 6, QC metrics.

12 k-mer Concepts and NGS Pipelines

Q31. In a de Bruijn graph, k-mer size affects:

(A) GC content

(B) Assembly resolution and repeat handling (*)

(C) Sequencing chemistry

(D) Only read length

Explanation: Larger k reduces ambiguity but requires higher coverage; smaller k can join repeats incorrectly. *Ref:* AG_sbobins genome assembly.

Q32. Which is the correct simplified order for an NGS variant calling pipeline?

(A) Alignment → Variant calling → QC

(B) QC → Alignment → Variant calling (*)

(C) QC → Variant annotation → Alignment

(D) Variant calling → Annotation → Alignment

Explanation: Raw data → QC → Align reads → Call variants → Annotate. *Ref:* Lecture 4, Variant analysis slide.

Q33. BUSCO evaluates:

(A) Structural variants

(B) Genome completeness using single-copy orthologs (*)

(C) Read duplication rate

(D) Gene expression levels

Explanation: BUSCO finds expected conserved genes to assess completeness of assemblies or annotations. *Ref:* Lecture 6, BUSCO.

13 Population and Comparative Genomics

Q34. ROH (Runs of Homozygosity) indicate:

(A) Recombination hotspots

(B) Segments identical by descent (*)

(C) Only heterozygous regions

(D) CNV duplications

Explanation: ROH = long homozygous stretches in the genome, linked to inbreeding. *Ref:* Lecture 11, Inbreeding section.

Q35. Which parameter measures population differentiation?

(A) F_{ST} (*)

(B) Heterozygosity

(C) Linkage disequilibrium

(D) LOD score

Explanation: F_{ST} compares genetic variance within vs. between populations. *Ref:* AG_sbobins, Population genomics.

Q36. A Manhattan plot in GWAS displays:

(A) Chromosome synteny

(B) SNP positions vs. –log10(p-value) (*)

(C) Gene expression levels

(D) Recombination frequency

Explanation: Peaks indicate loci significantly associated with traits. *Ref:* Lecture 10, GWAS example.