Name: **Enrolment No:**	

UPES

End Semester Examination, May 2024

Course: Bioinformatics and Computational Biology **Semester : VI**

Program: B.Tech Biotechnology **Duration : 3 Hours**

Course Code: HSBT3008 **Max. Marks: 100**

Instructions:

S. No.	Section A Short answer questions/ MCQ/T&F (20Qx1.5M= 30 Marks)	Marks	COs
Q 1	Which of the following databases is commonly used for storing DNA and RNA sequence data? a) UniProt b) GenBank c) PDB d) STRING	**1.5**	**CO1**
Q 2	Which of the following is a literature database commonly used to search for scientific articles? a) GenBank b) PubMed c) UniProt d) Ensembl	**1.5**	**CO1**
Q 3	Which of the following is NOT a common technique used in proteomics studies? a) Mass spectrometry b) Western blotting c) Polymerase chain reaction (PCR) d) Two-dimensional gel electrophoresis	**1.5**	**CO1**
Q 4	Which of the following is a type of high-throughput sequencing technology? a) PCR (Polymerase Chain Reaction) b) Sanger sequencing c) Next-generation sequencing (NGS) d) Western blotting	**1.5**	**CO1**
Q 5	The Ensembl database provides: a) Information about protein structures b) Genome annotation and gene prediction for various species c) Pathway information for metabolic processes d) Protein-protein interaction data	**1.5**	**CO1**

Q 6	Which scoring matrix is commonly used for aligning closely related protein sequences? a) PAM (Point Accepted Mutation) b) BLOSUM (BLOcks SUbstitution Matrix) c) PWM (Position Weight Matrix) d) JTT (Jones, Taylor, and Thornton)	**1.5**	**CO2**
Q 7	What does the term "FASTA" represent in the context of bioinformatics? a) Fast Algorithm for Sequence Transformation and Analysis b) Fast Automated Sequence Transfer and Alignment c) Fast Alignment of Sequence and Transcription Analysis d) Fast All Seq Tool for Alignment	**1.5**	**CO2**
Q 8	Which algorithm is commonly used for optimal global sequence alignment? a) Needleman-Wunsch algorithm b) Smith-Waterman algorithm c) BLAST algorithm d) FASTA algorithm	**1.5**	**CO2**
Q 9	In multiple sequence alignment, what is the purpose of gap opening and gap extension penalties? a) To align sequences globally b) To align sequences locally c) To penalize the creation of new gaps and extension of existing gaps d) To penalize mismatches between sequences	**1.5**	**CO2**
Q 10	What is the main purpose of a scoring matrix in sequence alignment? a) To determine the length of the sequences b) To identify the positions of gaps in the alignment c) To assign scores to matches, mismatches, and gaps d) To calculate the identity percentage between sequences	**1.5**	**CO2**
Q 11	Which algorithm is commonly used for motif discovery in protein sequences? a) Gibbs sampler b) BLAST c) Smith-Waterman algorithm d) Needleman-Wunsch algorithm	**1.5**	**CO3**
Q 12	What is the main advantage of using a Hidden Markov Model (HMM) for sequence analysis? a) It is faster and more efficient than other algorithms b) It can predict protein structures with high accuracy c) It can model dependencies between sequence positions d) It is specifically designed for pairwise sequence alignment	**1.5**	**CO3**

Q 13	Gene prediction algorithms rely primarily on sequence similarity to identify protein-coding genes in genomic sequences. Justify your choice. a) True b) False	**1.5**	**CO3**
Q 14	In phylogenetics, what does a phylogenetic tree represent? a) Protein-protein interactions b) Evolutionary relationships between species or genes c) Protein structures d) Conserved motifs in protein sequences	**1.5**	**CO3**
Q 15	What is the primary goal of motif discovery in biological sequences? a) Predicting gene expression patterns b) Identifying conserved patterns or functional domains c) Analyzing protein-protein interactions d) Estimating population genetics parameters	**1.5**	**CO3**
Q 16	What is the primary purpose of analyzing gene expression using microarray or sequencing datasets? a) Identifying protein-protein interactions b) Determining protein structures c) Understanding the expression levels of genes under different conditions d) Predicting gene functions	**1.5**	**CO4**
Q 17	Gene ontology (GO) is a standardized terminology used for: a) Analyzing protein structures b) Predicting protein-protein interactions c) Describing gene functions and attributes d) Estimating population genetics parameters	**1.5**	**CO4**
Q 18	Which of the following is a statistical method used to identify overrepresented functional categories in a gene list? a) Principal Component Analysis (PCA) b) Gene set enrichment analysis c) BLAST (Basic Local Alignment Search Tool) d) Southern blotting	**1.5**	**CO4**
Q 19	Which of the following is NOT a common step in analyzing gene expression using microarray data? a) Normalization b) Clustering analysis c) PCR amplification d) Differential expression analysis	**1.5**	**CO4**
Q 20	What is the main goal of studying protein-protein interactions? a) Predicting gene functions b) Identifying post-translational modifications c) Understanding cellular processes and pathways	**1.5**	**CO4**

	d) Analyzing gene expression patterns		
Section B (4Qx5M=20 Marks)			
Q 1	Discuss the field of metabolomics and its role in studying small molecule metabolites in biological systems. Provide examples of metabolomics applications in health and disease.	**5**	**CO1**
Q 2	Explain the significance of scoring matrices such as PAM and BLOSUM in sequence alignment. Describe how these matrices are constructed and their role in assessing sequence similarity.	**5**	**CO2**
Q 3	Explain the principle of the Gibbs sampler algorithm. Discuss its application in identifying conserved motifs or patterns within biological sequences.	**5**	**CO3**
Q 4	Describe the process of gene set enrichment analysis (GSEA) and its role in interpreting gene expression data.	**5**	**CO4**
Section C (2Qx15M=30 Marks)			
Q 1	Case study: The study involves two individuals: Patient A, who has been diagnosed with a hereditary disease, and Control Individual B, who does not exhibit any symptoms of the disease. Both individuals have volunteered to participate in genetic analysis to understand the underlying genetic factors contributing to the disease. Based on your understanding of sequence alignment, answer the following questions: A) Discuss how sequence alignment can be employed to study genetic variations associated with disease susceptibility. B) Provide a step-by-step flow chart for the methodology followed in identifying the genetic variations through sequence alignment. C) Discuss the tools and databases commonly used for functional annotation and their role in prioritizing candidate variants for further investigation. D) Discuss the ethical considerations involved in conducting genetic analysis studies.	**3+5+5+2**	**CO2**
Q 2	Case Study: The study involves the analysis of proteomics data obtained from two experimental conditions: Treatment Condition and Control Condition. The Treatment Condition represents cells or tissues subjected to a specific treatment or manipulation, while the Control Condition serves as an untreated reference. The objective is to	**5+5+3+2**	**CO4**

	identify differentially expressed proteins and gain insights into the biological mechanisms underlying the treatment response. Based on your understanding of data analysis, answer the following questions: A) What is "Proteome" and how proteomics data can be helpful in identifying key players in cellular processes? B) What are the common analytical techniques helpful for proteomics data acquisition. Discuss the principles of any two. C) What do you understand by the term "Differential expression analysis"? D) What are the common tools used to perform functional annotation of the differentially expressed proteins?		
Section D **(2Qx10M=20 Marks)**			
Q 1	Describe the steps involved in pattern finding using the Gibbs sampler algorithm. Discuss how the algorithm iteratively samples sequences to identify statistically significant motifs. **OR** Explain the concept of gene prediction and its importance in genome annotation. Discuss the advantages and limitations of using Hidden Markov Models (HMMs) for sequence analysis.	**10**	**CO3**
Q 2	A) Explain the workflow of analyzing gene expression data obtained from microarray or RNA sequencing experiments. B) Discuss the challenges associated with each step and strategies to address them.	**10**	**CO4**