B.Tech DEGREE EXAMINATION, DECEMBER 2023

Fifth Semester

18GEO104T - COMPUTATIONAL GENOMICS

(For the candidates admitted during the academic year 2020 - 2021 & 2021 - 2022)

Note:
i. **Part - A** should be answered in OMR sheet within first 40 minutes and OMR sheet should be handed over to hall invigilator at the end of 40th minute.
ii. **Part - B** and **Part - C** should be answered in answer booklet.

Time: 3 Hours   **Max. Marks: 100**

PART - A (20 × 1 = 20 Marks)
Answer **all** Questions

		Marks	BL	CO
1.	Which of the following is not present in DNA?	1	1	1
	(A) Adenine (B) Guanine			
	(C) Cytosine (D) Uracil			
2.	Which of the following is not a part of central dogma of molecular biology?	1	1	1
	(A) RNA (B) Protein			
	(C) enzyme (D) DNA			
3.	Study of total gene expression of a cell is called as	1.	1	1
	(A) Lipidomics (B) Genomics			
	(C) Transcriptomics (D) Proteomics			
4.	Which of the following is a start codon?	1	1	1
	(A) AUC (B) AUU			
	(C) AUG (D) UAC			
5.	Which of the following is not a nucleotide sequence database?	1	1	2
	(A) Genbank (B) EMBL			
	(C) DDBJ (D) Uniprot			
6.	Which of the following is not a sequence file format related to genomic data?	1	1	2
	(A) PDB (B) Fastq			
	(C) BAM (D) Genbank			
7.	Which of the following is a metabolic pathway database?	1	2	2
	(A) KEGG (B) Genbank			
	(C) TCGA (D) CEGG			
8.	Which of the following is an important constituent of the Fastq file format?	1	2	2
	(A) @ (B) >			
	(C) ^ (D) #			
9.	The data stored in the R data frame cannot be a _____ type.	1	2	3
	(A) numeric (B) factor			
	(C) character (D) special function			
10.	Which of the following is not a built-in datatype in python under the numeric category?	1	2	3
	(A) int (B) float			
	(C) complex (D) decimal			
11.	R was created by Ross Ihaka and _____	1	1	3
	(A) Rose Mary (B) Robert Gentleman			
	(C) Rocky Maivia (D) Ross Taylor			

12. The R language has a good similarity with which of the following languages   1   2   3
    (A) Python   (B) Perl
    (C) C   (D) Java

13. The content of a package is only available when a package is loaded using _____   1   3   4
    (A) install()   (B) library()
    (C) update()   (D) search()

14. Differential gene expression analysis can NOT be done using _____   1   2   4
    (A) RSEM   (B) LIMMA
    (C) DeSeq   (D) EdgeR

15. Which of the following languages is used for statistical data analysis and   1   2   4
    visualization in bioinformatics?
    (A) Python   (B) Java
    (C) C   (D) R

16. What is the name of the package that we use in bioinformatics for sequence analysis?   1   2   4
    (A) Bioconductor   (B) Bigconductor
    (C) Autoconductor   (D) Caret

17. The primary data structure that is used in biopython is   1   1   5
    (A) Lists   (B) Tuples
    (C) Dictionaries   (D) Sets

18. Name the biopython module that is used for sequence alignment   1   1   5
    (A) Bio.Align   (B) AlignIO
    (C) Seq.Align   (D) SeqIO

19. The _____ biopython object stores sequence annotations and metadata.   1   1   5
    (A) BioRecord   (B) SeqRecord
    (C) RecSeq   (D) Bio.Annotate

20. Which of the following is the mostly preferred language for bioinformatics analysis?   1   1   5
    (A) Python   (B) SAS
    (C) HTML   (D) C++

PART - B (5 × 4 = 20 Marks)
Answer any 5 Questions

	Marks	BL	CO
21. Contrast the structure of DNA and RNA with reference to the ribose	4	1	1
22. Relate the importance of genetic code to the protein synthesis process.	4	2	1
23. What is the significance of sequencing identity in choosing the optimal alignment that has biological significance?	4	3	2
24. Classify operators in R programming language.	4	4	3
25. List the different data types in Python with syntax.	4	4	3
26. Compare the GRanges and IRanges packages of R.	4	5	4
27. How can biopython improvise the biological data analysis and further help in inferring the results?	4	6	5

PART - C (5 × 12 = 60 Marks)
Answer all Questions

	Marks	BL	CO
28. (a) What are aromatic amino acids? Write their physical and chemical properties and the single letter and three letter codes?	12	2	1

(OR)

(b) Name the purine bases and draw their structures.

29. (a) Illustrate the importance of ASCII in the calculation of Phred score during the sequencing experiments.   12   3   2

(OR)

(b) Explain the types and the importance of nucleotide sequence databases and their common INSDC portal.

30. (a) Distinguish between Matrices and Data frames in R programming language.   12   4   3

(OR)

(b) Distinguish between file and sort functions of python.

31. (a) Compare EdgeR and LIMMA packages of Bioconductor.   12   5   4

(OR)

(b) Explain in detail about the GenomicRanges package.

32. (a) Elaborate the steps involved in the parsing of Fasta sequence file formats using biopython.   12   6   5

(OR)

(b) Elaborate the steps involved in the parsing of Genbank sequence file formats using biopython

* * * * *