

ISLAMIC UNIVERSITY OF TECHNOLOGY (IUT)
ORGANISATION OF ISLAMIC COOPERATION (OIC)
Department of Computer Science and Engineering (CSE)

MID SEMESTER EXAMINATION
DURATION: 1 HOUR 30 MINUTES

WINTER SEMESTER, 2022-2023
FULL MARKS: 75

CSE 4753: Bioinformatics

Programmable calculators are not allowed. Do not write anything on the question paper.
 Answer **all 3 (three)** questions. Figures in the right margin indicate full marks of questions whereas corresponding CO and PO are written within parentheses.

1. a) Write down the steps to prepare RNA-Seq library to measure gene expressions along with short details. 10
(CO1)
(PO1)
- b) Explain the observed post processing operations of Transcription phase during Central-Dogma. 8
(CO2)
(PO1)
- c) There are 64 codon combinations possible but only 20 amino acids are there. Nature adopts a many-to-one mapping (codon to amino acid) during translation phase of Central-Dogma. The mapping table is shown in Figure 1. 7
(CO2)
(PO1)

		Second Letter															
		U			C			A			G						
1st letter	U	UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys	3rd letter	U	U	UUC	UCC	UAC	UGC	C
		UUA	Leu	UCA		UAA	Stop	UGA	Stop		A						
		UUG		UCG		UAG	Stop	UGG	Trp		G						
		C		CUU		CCU	CAU	His	CGU		Arg		U				
CUC	CCC	CAC	Gln	CGC	C												
CUA	CCA	CAA	Glu	CGA	A												
CUG	CCG	CAG	Arg	CGG	G												
1st letter	A	AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser	3rd letter	U						
		AUC		ACC		AAC	AGC	C									
		AUA		ACA		AAA	Lys	AGA	A								
		AUG		ACG		AAG	Arg	AGG	G								
1st letter	G	GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly	U							
		GUC		GCC		GAC	GGC	C									
		GUA		GCA		GAA	Glu	GGA		A							
		GUG		GCG		GAG	Gly	GGG		G							

Figure 1: Codon to amino acid mapping table for Question 1.c)

Discuss the significance of the mapping scheme.

2. a) Show various gene products by drawing a diagram showing their dependency relationship. 5
(CO1)
(PO1)
- b) Nussinov Folding Algorithm detects best secondary structure for an RNA sequence. Describe how the algorithm works. 10
(CO3)
(PO1)

Table 1: Gene expression data for Question 2.c)

	Sample 1	Sample 2	Sample 3
Gene 1	34	24	64
Gene 2	120	140	20
Gene 3	80	95	160
Gene 4	40	110	10

c) Table 1 represents gene expression data for 4 genes across 3 samples. Normalize the gene expression data using quantile method.

10
(CO2)
(PO1)

3. a) Write a short note on scoring matrix.

5
(CO2)
(PO1)

b) A set of sequences S1, S2, S3, S4, and S5 is needed to be aligned. Distance matrix for the sequences is shown in Table 2. Construct a guided tree for ClustalW method using Unweighted Pair Group Method with Arithmetic mean (UPGMA) approach.

10
(CO2)
(PO1)

Table 2: Distance matrix for Question 3.b)

	S1	S2	S3	S4	S5
S1	0				
S2	14	0			
S3	12	8	0		
S4	10	15	12	0	
S5	15	18	16	20	0

c) A biostatistician wants to apply Center Start Method to align sequences mentioned in Question 3.b). Find the reference sequence for him/her.

10
(CO2)
(PO1)