Dr. Narendra Kumar

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT
T-3 EXAMINATION, Dec 2019
Ph.D. I Semester

COURSE CODE: 13M11BT112 **MAX MARKS: 35**
COURSE NAME: Advanced Bioinformatics **MAX. TIME: 2 Hrs.**

Note: All questions are compulsory. Attempt all questions of a particular section at one place. Answer each question to the point.

---

1. How do we describe the diversity of the microbiome? Explain various measures. (4 marks) (CO2)
2. Explain the various steps involved in the computational analysis of RNA-seq data, providing the name of the software. (5 marks) (CO3, CO4)
3. How do we perform loop modelling and side chain modelling for homology modelling? (3 marks) (CO)
4. Given the following gene expression data, answer the questions: (12 marks) (CO3, CO4)

Gene Symbol	Gene Name	Tissue (Liver)	Tissue (Pancreas)
CTRB2	chymotrypsinogen B2	1234	3761
CPA1	carboxypeptidase A1	1320	3541
PNLIPRP1	pancreatic lipase related protein 1	23	516
CTRB1	chymotrypsinogen B1	10	1908

   a. Construct a matrix in R containing the above data. Write code.
   b. Calculate the average expression of these genes in Liver. Write code.
   c. Calculate the average expression of these genes in Pancreas. Write code.
   d. Add a column to the matrix containing the average expression of gene. Write code.
   e. Add a row containing the average expression of tissue. Write code.
   f. Write the expression to get the values of CPA1 and CTRB1 in Liver tissue.

5. Why do we need the protein classification system? Write a note on the following: (2 + 2 + 2) (CO6)
   a. CATH
   b. SCOP
6. Describe Protein domains and how do we identify and study them? (5 marks) (CO5)