BIRLA INSTITUTE OF TECHNOLOGY, MESRA, RANCHI
(END SEMESTER EXAMINATION)

CLASS: MSc BIOINFORMATICS & COMPUTATIONAL BIOLOGY   SEMESTER : III
BRANCH: BIO-ENGINEERING AND BIOTECHNOLOGY   SESSION : MO/2024

SUBJECT: BI303 NGS DATA ANALYSIS

TIME: 3 Hours   FULL MARKS: 50

INSTRUCTIONS:
1. The question paper contains 5 questions each of 10 marks and total 50 marks.
2. Attempt all questions.
3. The missing data, if any, may be assumed suitably.
4. Before attempting the question paper, be sure that you have got the correct question paper.
5. Tables/Data hand book/Graph paper etc. to be supplied to the candidates in the examination hall.

			CO	BL
Q.1(a)	Why has Next-Generation Sequencing (NGS) gained popularity over Microarray technology?	[5]	1	1
Q.1(b)	What are the key differences between Roche 454 sequencing and Illumina sequencing technologies?	[5]	1	2
Q.2(a)	Illustrate a schematic diagram of adapters in a library and describe the three main segments commonly found in an adapter sequence, along with their functions. Explain how Illumina output includes adapter sequences in FASTQ files	[5]	1	2
Q.2(b)	Find the Burrows-Wheeler Transform (BWT) of reference sequence "GACGTACG" and map read "GTA" onto the reference sequence using the BWT approach	[5]	3	2
Q.3(a)	Compare the CPM, FPKM and TPM methods for normalizing RNA-Seq data. Outline the steps involved in performing normalization using each of these methods.	[5]	2	3
Q.3(b)	Outline the best practice guidelines for generating high-quality RNA-Seq data	[5]	3	3
Q.4(a)	What is variant calling? What types of variants do you expect to find in the mapped reads when comparing them to the reference genome?	[5]	4	3
Q.4(b)	Draw a schematic diagram of the steps involved before variant calling. Why are local alignment, duplicate removal, and base quality score recalibration (BQSR) important in the pipeline? Write at least one tool name for performing local alignment, duplicate removal and BQSR steps.	[5]	4	2
Q.5(a)	Find the Shortest Common Superstring (SCS) from a directed graph generated from the following reads: ATAGGA, TAGGAC, AGGACT, GGACTA, and GACTAG. Draw the directed graph with updated edge weights for each step used to identify the SCS. The cutoff length for overlap is given as 3	[5]	3	3
Q.5(b)	Write all the steps involved in forming a de Bruijn graph. Explain what an Eulerian path is with an example	[5]	3	4

::::::21/11/2024::::::E