20128B50063 Shrajay

Midsem

BBL735 (Genomics and Proteomics)

Maximum marks: 30 Time: 60 mins

Write your Name and ID number on the answer sheet

Q1. Calculate the number of 100 bp DNA fragments in 1 ng of fragmented DNA (3)

OR

Q1. How can you use carboxylated beads to select DNA fragments only of size between 100 bps and 300 bps (3)

Q2. Calculate the probability of misassignment of DNA fragments when multiplexing two samples in an Illumina sequencing reaction using two barcodes AAATTT and AATTTT and assuming a FastQC cut-off of 30 (3)

Q3. Make the Graph for the following: (3)

Distribution of DNA fragments in a well if the library is diluted and if it is not-diluted

OR

Error vs Phred score for Illumina sequencing and Nanopore sequencing

Q4. Make the DeBrujn Graph of the sequence AATTCGCCGC with a read length of 3 and coverage of 6x. How will the graph change if you have 3x reads from both AATTCGCCGC and AATACGCCGC? (3+2=5)

Q5. Which sequencing technology (Illumina, Nanopore, etc) and sequencing assay (DNA sequencing, RNA sequencing, Exome sequencing, etc) will you use for the following use-cases and why? (Pick any 3; 1 x 3 = 3)

a. To detect a genetic disease in a new-born baby of no-immediate concern
b. To detect an infection in a patient suspected of sepsis (death expected within 48 hrs)
c. To detect mutations associated with 25 genes associated with cancer.
d. To identify fake mRNA vaccines (mRNA vaccines have modified RNA)

Q6. Design a genomics assay to calculate the number of RNA-molecules that are being translated. *Hint: This experiment will be sequencing two samples* (3)

OR

Design a genomics assay to assign transcriptional factories (actively transcribed regions) (3)

Q7. You run a 16S rRNA sequencing experiment after mixing bacterial DNA from 3 species in equal ratio. Explain why you may get the following results : (1 x 3 =3)

a. You get 4 species

b. One of the species is not detected

c. You find 50% of reads from one specie and 25% reads from the other two

Q8. The following data depicts the mapping statistics of paired-end reads: (1+1 = 2)

```
50000 reads; of these:
  50000 (100.00%) were paired; of these:
    1880 (3.76%) aligned concordantly 0 times
    44731 (89.46%) aligned concordantly exactly 1 time
    3389 (6.78%) aligned concordantly >1 times
    ----
    1880 pairs aligned concordantly 0 times; of these:
      275 (14.63%) aligned discordantly 1 time
    ----
    1605 pairs aligned 0 times concordantly or discordantly; of these:
      3210 mates make up the pairs; of these:
        1882 (58.63%) aligned 0 times
        947 (29.50%) aligned exactly 1 time
        381 (11.87%) aligned >1 times
98.12% overall alignment rate
```

What does it mean to align concordantly ? Why can a read pair align concordantly >1 times ?

Q9. Which of the following statements are TRUE or FALSE. Give explanation: (1x3 =3)

a. You will always get split reads in Bacterial RNA sequencing

b. The genes in the following two samples can be directly compared

Gene	Sample 1	Sample 2
A	30	235
B	24	196
C	0	0
D	361	0
E	5	39
F	15	102
Total	435	435

c. The number of lines in two paired end sequencing files (R1 or R2) files is never equal

Q10. What can happen during the sequencing experiment if you (Explain any one): (2)

Underestimate the size of the library of DNA fragments before clustering flowcell

(You think it is 200 bp but it is 400 bp in reality)

Or

Perform too many PCR cycles after adapter ligation

(You do 20 PCR cycles extra)