1. A skin cell and brain cell from the same organism have the same genome but differ wildly in form and function. This is possible because these two cell types differ from each other molecularly in all of the following ways EXCEPT: (5 points)
2. Different 3D genome conformations
3. Different histone modifications
4. Different methylation patterns
5. Different promoter sequences
6. Different transcription factor activities
7. Protein-protein interaction networks are of importance in biomedical science for all of the following reasons EXCEPT: (6 points)
8. They allow inference of guilt-by-association in disease
9. They allow pharmaceutical companies to design a drug to affect the interaction partner of a protein that cannot be directly inhibited
10. They are a starting point for elucidating the mechanism of interactions
11. They are part of a complete survey of the cell
12. They help to define biological pathways
13. They teach us about the mass to charge ratio of peptides
14. Jim Grey considers data science to be a new, fourth way of doing science, in which computers play a central role in unifying the three existing ways of doing science: observation, theory, and \_\_\_. (2 points)
15. Deduction
16. Practice
17. Simulation
18. Visualization
19. What is the typical number of single nucleotide variants in one person’s (germline) genome with respect to the human reference genome? (2 points)
20. 30
21. 300
22. 3,000
23. 30,000
24. 300,000
25. 3,000,000
26. 30,000,000
27. For determining the structure of a large complex, flexible molecule, the best argument in favor of using cryo-EM over x-ray-based methods is the fact that cryo-EM: (4 points)
28. Can still operate at colder temperatures
29. Does not require crystallization
30. Is cheaper
31. Offers higher resolution
32. Use cases for sequencing technologies. Match the goals to their guiding priority (2 points each) and the technology to whichever priority you selected (1 point each). Only list each priority once and each technology once. (9 points total)

GOAL1: Bridging gaps from the repetitive regions in the assembly of a new genome

GOAL2: Cataloguing single nucleotide variants in a research cohort of 100 patients

GOAL3: Validation of a clinically significant germline single nucleotide polymorphism

PRIORITY1: Achieving long read lengths

PRIORITY2: Having the highest possible base calling accuracy

PRIORITY3: Keeping costs affordable

TECHNOLOGY1: PacBio single molecule sequencing

TECHNOLOGY2: Standard Illumina next-generation sequencing

TECHNOLOGY3: Traditional Sanger sequencing

For GOAL1, we care most about PRIORITY\_\_, which of the options, is best served

by TECHNOLOGY\_\_.

For GOAL2, we care most about PRIORITY\_\_, which of the options, is best served by TECHNOLOGY\_\_.

For GOAL3, we care most about PRIORITY\_\_, which of the options, is best served by TECHNOLOGY\_\_.

1. Suppose Carl has recently moved zip codes from 20015 to 06511. The registrar made ***some*** update in the “Lecturer information” excel spreadsheet to reflect this move, but failed to make a complete update to the “Lecturer information” spreadsheet.

What is the error below in the “Lecturer information” spreadsheet, and what concept/s from relational databases could have prevented it? (4 points)

Spreadsheet 1: Lecturer information

|  |  |  |  |
| --- | --- | --- | --- |
| lecturer | Lecture\_duration | City | Zip code |
| Kai | 60 | New Haven | 06511 |
| Mark | 75 | New Haven | 06511 |
| Carl | 40 | Washington | 06511 |

Spreadsheet 2: Zip code information

|  |  |
| --- | --- |
| Zip code | City |
| 06511 | New Haven |
| 20015 | Washington |

Error:

Preventative concept:

1. List two benefits proponents of clinical genomics hope we will achieve if we link patient genomes to their electronic health records. (4 points)
2. List one limitation of ChIP-seq or difficulty in its interpretation that was mentioned in class. (3 points)
3. Use the Needleman-Wunsch algorithm to compute the similarity matrix and sum matrix of the two peptide sequences ASVAVB and EAVABD, find the traceback, and report the optimal alignment. Use a mismatch score of 0, a match score of 1, a gap initiation penalty of 0.5, and a gap extension penalty of 0. Be sure to check your work since careless mistakes are common.(18 points)

Tabulate the similarity (dot) matrix. (2 points)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | A | S | V | A | V | B |
| E |  |  |  |  |  |  |
| A |  |  |  |  |  |  |
| V |  |  |  |  |  |  |
| A |  |  |  |  |  |  |
| B |  |  |  |  |  |  |
| D |  |  |  |  |  |  |

Compute the sum matrix (8 points) and draw the optimal trace-back (4 points)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | A | S | V | A | V | B |
| E |  |  |  |  |  |  |
| A |  |  |  |  |  |  |
| V |  |  |  |  |  |  |
| A |  |  |  |  |  |  |
| B |  |  |  |  |  |  |
| D |  |  |  |  |  |  |

Write the optimal alignment with dashes for gaps (4 points)