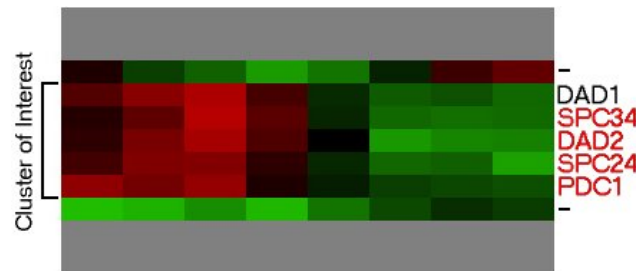


1. Liquid chromatography-mass spectrometry (LC-MS) is a common technique for the analysis of peptides in mixtures. These data are three-dimensional with one dimension being intensity and two dimensions of separation. What are the two dimensions of separation?



2. Analyze the section of microarray data on the right. The data are from *Saccharomyces cerevisiae* and measure mRNA expression under different experimental conditions. According to the clustering algorithm, these data indicate the five genes are in the same pathway (arbitrarily called the DAD1 pathway). Using the tools given to you during this quarter's quiz sections, validate whether or not the four genes with red lettering are in the same pathway. For each of the four genes, indicate if it belongs in the DAD1 pathway and explain why you believe this to be true.

3. You are searching for a biomarker for ovarian cancer that occurs in approximately 2 per 2,500 people. You have determined that a new protein that you have discovered has a sensitivity of 100% and a specificity of 98%. In a random population of 10,000 patients, how many positives and how many false positives will your biomarker detect?

4. You have an interesting protein related to heart disease in mice. Careful analysis of the mRNA sequence shows the mutation to be a 3 amino acid deletion in the middle of the protein sequence. You decide to confirm this by peptide fragmentation analysis in a mass spectrometer. However, after database searching on your data (using the mouse reference proteome), you only find peptides upstream or downstream of your region of interest. Why did your approach fail?

5. Go to <http://proteome.gs.washington.edu/classes/Genome372/final/sequence.txt>  
The file contains human genomic DNA sequence. Identify the following:

What chromosome is the DNA from: \_\_\_\_\_

Find all the unique genes in the sequence (partial or complete) and indicate whether they have higher expression in liver or kidney using expression data from Affy U133A and GNF1H Chips.

6. How does SELDI differ from MALDI?

For questions 7-9, choose from the following list of experimental methods:

- A) Antibody Pull-Down
- B) ICAT
- C) Top-Down Mass Spectrometry
- D) Bottom-Up (Shotgun) Mass Spectrometry
- E) Antibody Array
- F) Protein-interaction Array
- G) Targeted Proteomics
- H) Western Blot

7. A colleague of yours comes to you and asks you to measure the relative abundance of their favorite protein in 50 different samples grown under 10 different conditions. Which experimental method should you use and why?

8. A colleague is working on a protein that they know contains three phosphorylation sites on different regions of the protein. They would like to know if only one modification occurs at a time or if some combinations of the modifications are more prevalent than others. The amount of sample is not a limiting factor. Which experimental method should you use and why?

9. A colleague hands you a box of 100 antibodies that target 20 yeast proteins. Five antibodies were generated for each protein, and each antibody has a different epitope. Your colleague would like to know which antibody is the most specific for each of the 20 proteins. Which experimental method should you use and why?

10. How does a “Targeted Proteomics” differ from “Shotgun Proteomics using data dependent acquisition”?

11. You perform a next-generation sequencing experiment on human DNA from an anonymous donor that produces 120 million 36bp reads. Of these, 86.4 million can be mapped back to unique locations in the human genome using a special fast alignment algorithm for short read sequences (ELAND).

(a) What are two explanations for why the other 33.6 million cannot be mapped to unique locations?

(b) Assuming the mapping reads are randomly distributed, what is the average level (X) of sequence coverage of the genome?

(c) 10 reads from the sequencing run that map to this exon are available here:

[http://proteome.gs.washington.edu/classes/Genome372/final/next\\_gen\\_sequence.txt](http://proteome.gs.washington.edu/classes/Genome372/final/next_gen_sequence.txt)

Unfortunately, your informatics personnel forgot to tell you exactly where these sequences map in the genome. Identify the genomic coordinates to which these sequences map. Identify the gene(s) if any.

|

(d) Several of the reads from part (c) overlap one another, providing redundant coverage of some genomic bases. What is the maximum per nucleotide sequence coverage you observe? What is the average level of coverage of

(e) In another region of the genome with redundant coverage you find the following:

AAGGACAAATTTGCAAAGGCCAAAGAGTTGGGTGCC  
AAGGCCAAAGAGTTGGGTGCCACTGAATGCATCAAC  
AATTTGCAAAGGCCAAAGTGTTGGGTGCCACTGAAT

What is the explanation for the discrepancy in bases at the indicated sequence position?

12. You now use your next generation sequencer and sequence mapping and assembly algorithms to “finish” the genome of your anonymous donor individual from Question 11. The quality and depth of the finished sequence allows you to find reliably all of the nucleotide positions at which your sequence differs from the human reference sequence (i.e., SNPs). At approximately how many positions will your sequence differ from the reference? You may express the answer as a number, as a percentage, or as a rate (i.e., 1 difference per \_\_\_ bases).

13. You have a gene of interest and wish to determine if it is expressed in the dentate gyrus of the mouse brain.

(a) Describe two different experimental technologies you could use to determine if your gene was expressed in this tissue structure.

(b) For one of the strategies in part (a), list the main experimental steps involved starting with just the knowledge of the sequence of your gene and its exons, and ending with a determination of whether or not it was expressed in the dentate gyrus. (Hint: one process might involve cDNA. Note however that there are multiple possibilities for correct answers).

14. In living cells, the human genome is packaged into chromatin. Suppose you had a special set of atomic force tweezers and microscopes that would allow you to pluck all of the human genomic DNA out of a single skin cell nucleus without breaking the DNA.

(a) You perform the above experiment, and remove all the proteins. Then, you stretch out each chromosome DNA into a straight molecule and line the DNA molecules up end-to-end. Approximately how long is the resulting chain of DNA molecules?

(b) You perform the experiment again. However, this time you are careful not to disrupt the nucleosomes around which DNA is wound. You treat the chromatin fibers you plucked from the cell nucleus so that they relax into fibers with a 'beads-on-a-string' appearance under the electron microscope, of approximately 10nm in width. There is one fiber corresponding to each DNA molecule in part (a) (i.e., one fiber for each chromosome from the skin cell nucleus). You use your atomic force tweezers to stretch these 10nm wide chromatin molecules out, and line them up end-to-end. You then measure the length of the resulting chain of chromatin molecules. What is the approximate length of the chain? You may also express this as a fraction/multiple of the answer to part (a).

#### 15. True-or-false and brief questions

1. ( T / F ) Human DNA methylation occurs predominantly at cytosine residues followed by adenine residues.
2. ( T / F ) DNA methylation patterns can be inherited by daughter cells
3. ( T / F ) In the human genome, Small insertions and deletions (in/dels) are just as common as single nucleotide polymorphisms (SNPs).
4. ( T / F ) Some neurological diseases can be caused by deletions or duplications that involve entire genes.
5. DNA polymerase can copy DNA at a rate of \_\_\_ bases per minute.