BC367, Fall 2019 Problem set #3 Due: Friday, September 20

(Don't forget to answer each question on a separate page. Please type narrative questions)

"Protein homology searching – a case study"

The situation:

An adult male presents with fatigue and shortness of breath. He has had a relatively severe cough with chest pain. The cough is productive and often contains blood. He is also experiencing weight loss and minor swelling of the face and neck.

Genetic tests reveal a chromosomal abnormality and a DNA sequence that, base for base, has not previously been reported. The sequence translates into the following linear polypeptide sequence.

PPGTRVRAMAIYKQSQHMTEVLRHFPHHERCSDSDGLAPPQHLIRVEGNLRVEYL
DDRNTFRHSVVVPYEPPEVGSDCTTIHYNYMCNSSCMGGMNRRPILTIITLEDSSG

About BLAST:

The Basic Local Alignment Search Tool (BLAST) finds regions of local similarity between a queried protein or nucleic acid sequence and sequences in published databases. The program calculates the statistical significance of matches. BLAST can be used to infer functional and evolutionary relationships between sequences as well as help identify members of gene families.

Procedure:

1) Go to the BLAST home page.
2) Select "Protein BLAST" button.
3) Copy and paste the above polypeptide sequence into the search field (labeled "Enter accession number, gi, or FASTA sequence"). Ensure that the search set is the non-redundant protein sequences database. Enter 'human' in the organism field, ("human taxid:9606" should pop up). Change no other parameters
4) Click "BLAST" on the bottom left.
5) When the search is complete (it may take several minutes), the first thing you'll see is a list of descriptions of matching sequences and some corresponding information.

Questions (a few sentences to a small paragraph each … TYPED on separate pages):

1) Look at the first few aligned sequences shown in the "Descriptions" tab. What is the most likely protein associated with your mutant sequence? What other information do you see?

2) Click on the "Alignments" tab. For each alighnement, the 'Query' sequence is the one you entered – the 'Sbjct' sequence is the one from the database to which your sequence is being compared – the sequence between the two shows any identity or homology. Scan the alignment and identify the differences. What are the mutations and how do they compare at the amino acid level? Are these mutations likely to result in a dramatic functional difference among the query sequence and the subject sequences? Explain.

3) Click on the Sequence ID links of the first few alignments. The complete database entries for each sequence should appear. What kind of information do you see? Other than sequences from unpublished results, you will be able to click on a PubMed number that takes you to the reference for the affiliated publications. What ***kind*** of protein does our sequence likely come from? What is the significance of identification with respect to the patient in this study (you should use other sources of information for this question)?

4) Try searching in Google using what you know about this patient's symptoms and mutated protein and predict a diagnosis. Within PubMed or Scopus, search using the name of the mutated protein and your predicted diagnosis. Scan the titles of the journal articles and read some of the corresponding abstracts. Based on what you saw in the Sequence ID pages and these titles, what is a plausible diagnosis for our patient? Add the search term 'therapy' and click 'search.' By inspecting the titles and abstracts of the journal articles that you see, what is the latest in current clinical research with this disease? What does the future hold?

5) Repeat the same BLAST search, but with 'mouse' (taxid:10088) as the organism. Identify sequence differences as before. Would the mouse be an appropriate model organism to study potential therapies for our patient's disease? Explain.