**BC367, Fall 2017          Problem set #3          Due:  Friday, September 22**

(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 when physically active. He is anemic and has a pale complexion.  He has discomfort on the left side of his abdomen from an enlarged spleen. He is also experiencing excessive sweating, weight loss, and inability to tolerate warm temperatures.

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.

VHSIPLTINKEDDESPGLYGFLNVIVHSATGFKQSSKALQRPVASDFEPQ
GLSEAARWNSKENLLAGPSDNDPNLFVALYDFVASGDNTLSITRGEKL
RVLGYNHNGEWCEAQTKNGQGWVPSNYITPVNSLEKHSWYHGPVSRN
AAEYLLSSGIN

### 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)      After a few moments, the first thing you'll see is a graphical representation of the search results – the red bars show highly conserved matches.  Rolling your mouse over the bars will identify the matched sequences.  Scroll down to the "Alignments" section.

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

1)      Look carefully at the first few aligned sequences.  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 is a composite.  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.

2)      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?  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)?

3)      In the "Related Information" section on the right column in each of your selected Sequence ID pages, click on 'Pubmed' to see a list of the other articles that have cited that from which this sequence came.  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 the likely diagnosis for our patient?  Copy and paste the complete name of the disease given in the abstract title into the search field on top of the PubMed page.  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?

4)      Now add the search terms 'imatinib' and 'resistance'.  What is imatinib?  What is the cause of imatinib resistance in this disease?  What strategies are researchers using to overcome this resistance in potential therapeutic regimens?

5)      Repeat the same BLAST search, but with 'mouse' 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.