PGSS Biology Core 2022 Homework #1
Due Thursday, June 30

The sequence of DNA in Figure 1A is from an exon near the center of a specific wild type (normally functioning) human gene. This gene is transcribed, or "expressed", by many human tissues including the intestinal epithelium.

1. Take the DNA sequence and transcribe it into mRNA.

2. Take the mRNA sequence and translate it into protein (use the single letter abbreviations for the amino acids) using the genetic code.

3. Used by biologists around the world, BLAST is a freely available computational tool provided by the National Center for Biotechnology Information (NCBI) for analyzing DNA and protein sequences. It does this by comparing your sequence (the query), to the vast database of DNA and protein sequences maintained by the NCBI.

(https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins)

Take the amino acid sequence and identify the protein using BLAST. See Figure 2 for guidance. After the search is complete you will need to scroll down to find the part of the results that shows you alignments between your sequence and sequences from the database.

4. Take the **mutant** DNA sequence in Figure 1B and transcribe it into mRNA.

5. Take the mRNA sequence and translate it into protein using the genetic code.

6. Predict what will happen to the protein. Then, go to protein schematic in Figure 3 to see where this mutation is located in the gene and the corresponding amino acid location in the protein. Based on that information about this specific protein, predict the consequence to the mutant protein (e.g. what functions will it retain, which will it lose?)

7. PubMed is a database of biomedical literature maintained by the NIH. Go to PubMed, search using the name of the protein as you identified it using BLAST including the terms “mutations” and “signaling” in the search and find out what effect mutations in this gene has in real life. In addition, discover something about the normal function of the protein. Write a brief synopsis (a few sentences) of your findings.

https://www.ncbi.nlm.nih.gov/pubmed

8. a. Examine the plot that shows mutation frequency in Figure 3 and you will notice that mutations in our gene of interest that are associated with human disease can be either somatic or germline. Diseased human cells with mutations in our gene of interest (you should know what kind of cells these are from your answer to question 7) always seem to be homozygous mutant; in the majority of patients, both disease alleles will be the result of somatic mutation, while in other patients, one allele will result from somatic mutation, and the other from a germline mutation. In light of that information, describe where you think the mutations came from in those two classes of patients.

b. The frequency of mutation varies across the gene. Speculate about the basis for the observed pattern and what that tells you about the function of the different parts of the protein.

Figure 1A

DNA sequence (wild type)

The top strand is the coding strand and the bottom strand is the template strand.

5′ CAAGAGGCTGATAGCGCCAATACACTTCAAATCGCTGAGATCAAAGAAAAAATCGGGACACGAAGTGCTGAGGATCCCGTC 3′
3′ GTTCTCCGACTATCGCGGTTATGTGAAGTTTAGCGACTCTAGTTTCTTTTTTAGCCCTGTGCTTCACGACTCCTAGGGCAG 5′

Figure 1B

DNA sequence (mutant found in a patient population)

The top strand is the coding strand and the bottom strand is the template strand.

5′ CAAGAGGCTGATAGCGCCAATACACTTCAAATCGCTGAGATCAAATAAAAAATCGGGACACGAAGTGCTGAGGATCCCGTC 3′
3′ GTTCTCCGACTATCGCGGTTATGTGAAGTTTAGCGACTCTAGTTTATTTTTTAGCCCTGTGCTTCACGACTCCTAGGGCAG 5′

Figure 2

1. Enter the single letter amino acid sequence here

2. Do not adjust the parameters

3. BLAST!

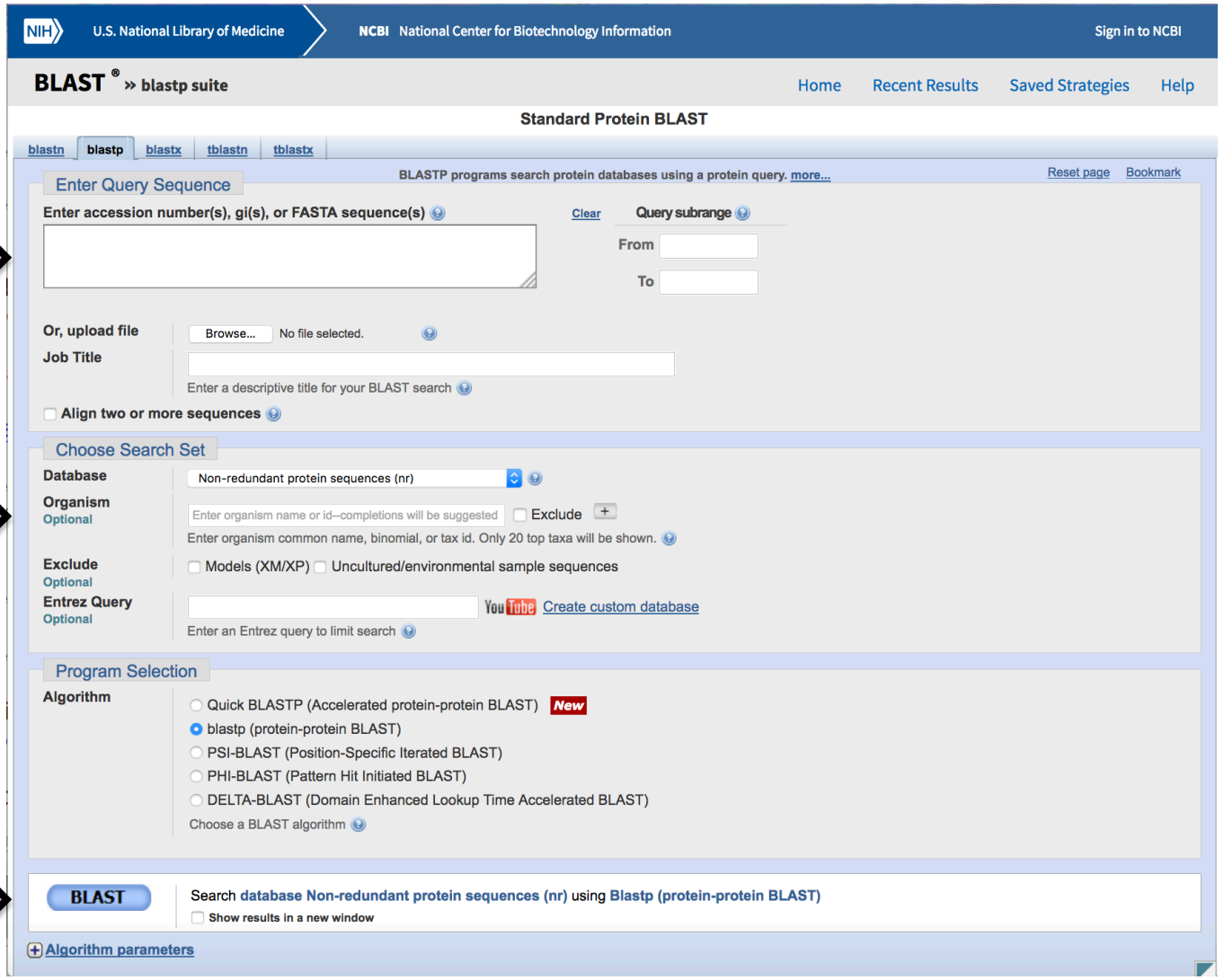
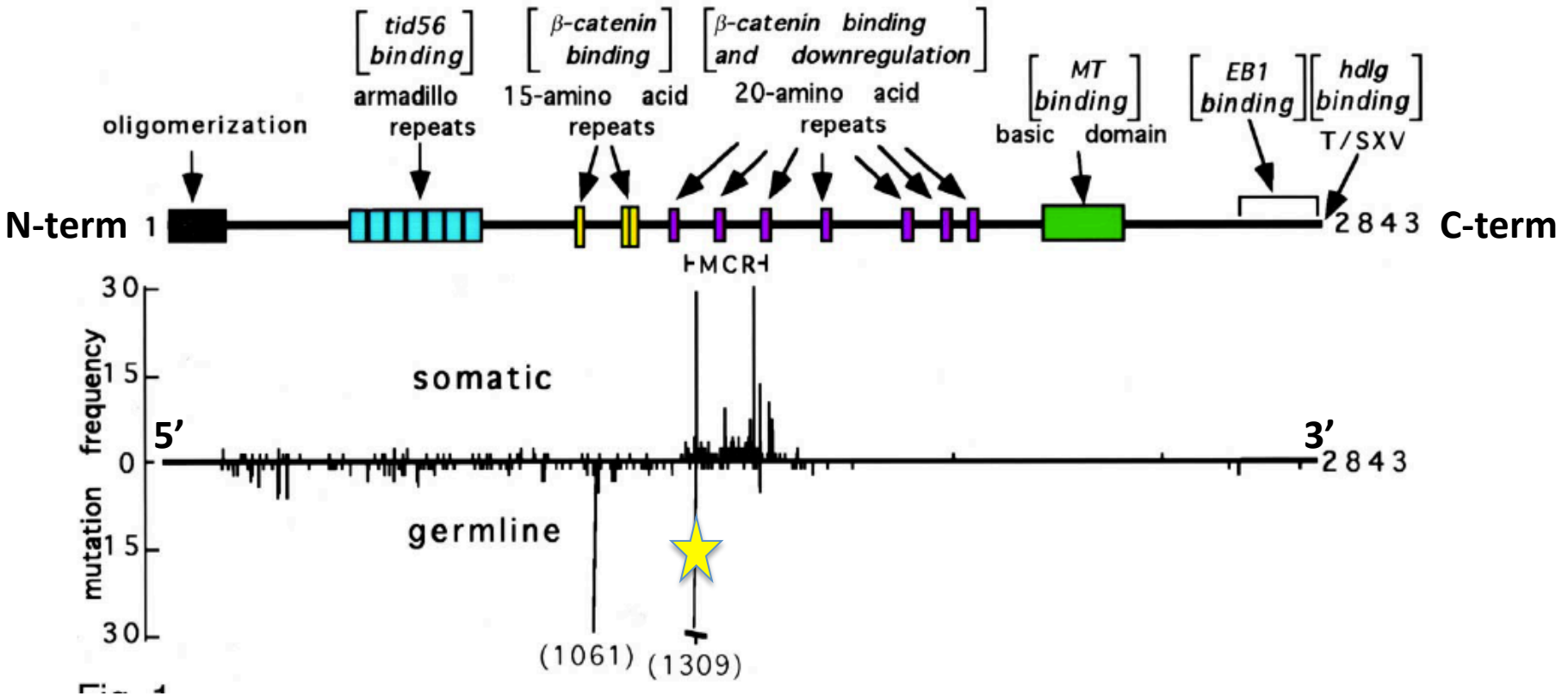


Figure 3



- Each colored box in the protein cartoon represents a different functional region of the protein
- MCR = mutational cluster region
- the yellow star indicates the position of the mutation described in questions 4 and 5
- mutation frequency refers to how often a mutation at a specific nucleotide has been identified in a cancer patient population. Higher numbers indicate that the mutation is more frequent.