Old Dominion University/BIOL294-Genetics/Rinehart-Kim/Module 9/FA20Genome Assignment

- You are to address the statements/questions that are in red. You do not need to use complete sentences in your replies.
- The following is an excellent reference: Genetics Home Reference: http://ghr.nlm.nih.gov
- **All links are working and you should not have to pay to use them. However, remember that you might need to use different browsers. If you absolutely cannot get a link to work, please state that. Otherwise, an unanswered question will be considered as "incorrect".**
- If you use any source other than websites I have listed to answer your questions, you need to cite the source(s) that you used.
- **Each question is worth 2 points.**

Chromosome Maps

Assignment Goal: To use the Internet-based Genes and Disease site (NCBI) to view the assignment of genes to chromosomes.

Assignment: Access the Genes and Disease site at http://www.ncbi.nlm.nih.gov/books/NBK22183/
Under "Contents", select "Chromosome Map" (at the very bottom).
A karyotype will appear.
Click on a chromosome.
1. WHAT CHROMOSOME DID YOU CHOOSE? Chromosome X
Above the chromosome image you will see the number of genes and base pairs on that particular chromosome.
2 & 3. STATE THE NUMBER OF GENES **AND** BASE PAIRS ON THE CHROMOSOME YOU CHOSE.
Contains over 1400 genes and 150 million base pairs
Scan the chromosome map.
4. LIST ONE GENE THAT IS LOCATED ON THIS CHROMOSOME. HEMA
5. STATE THE **NORMAL** FUNCTION OF THE GENE YOU LISTED IN #4. Regulates blood clots
6. STATE THE POSSIBLE DISEASE(S) RELATED TO THIS GENE. (This should be possible by clicking on the gene you stated in #4.) Hemophilia

GenBank

7. WHAT IS GENBANK? (You can do an Internet search to find this information.) Genbank is a database that list publicly available nucleotide sequences for over 300,000 organisms.

Introduction to BLAST

Assignment Goal: To use the Internet-based site BLAST, Basic Local Alignment Search Tool (NCBI), to search for similarities between nucleotide sequences.

Assignment: Access the BLAST site at http://blast.ncbi.nlm.nih.gov/Blast.cgi
Click on "Nucleotide Blast"
Assume that you found this nucleotide sequence when you sequenced a piece of DNA in the laboratory in which you work:

```
tgtgtgtagg ggggaaggaa tttagctttc acatctctct tatgtttagt tctctgcatg  tgcagttaat
cctggaactc cggtgctaag gagagactgt tggcccttga aggagagctc  ctccctgtgg atgagagaga
aggactttac tctttggaat tatctttttg tgttgatgtt  atccaccttt tgttactcca
```

Enter the above sequence (you may copy and paste) into the "Enter Query Sequence" box at the top of the page.
Under "Program Selection" near the bottom of the page, choose "somewhat similar sequence (blastn)"
Click the "BLAST" button at the bottom of the page to run the search.
Give some time for the results of your search to show up.
You will be given significant matches for the sequence that you entered.
8. WHAT IS THE TOP SEQUENCE DESCRIPTION MATCH FOR YOUR QUERY SEQUENCE? **DO NOT CHOOSE THE PREDICTED SEQUENCE.** For this answer, you should give the description listed.
Homo Sapiens CF transmembrane conductance regulator (CTFR)

9. WHAT DOES THE ENCODED PROTEIN DO IN THE BODY? Search the PubMed site at https://pubmed.ncbi.nlm.nih.gov/ to answer this question. Do not give the disease associated with this gene, but the function of the gene itself. CITE THE PAPER YOU USED TO DETERMINE THE PURPOSE OF THE ENCODED PROTEIN. CTFR is a membrane protein connected to the ABC transporter family functioning as a chloride/anion channel in epithelial cells around the body (Meng X, et al. 2017).

Click on the top match to find the following.
10. FOR WHAT DISEASE IS A MUTATED FORM OF THIS GENE RESPONSIBLE? You should be able to get this information from the description of the gene. Cystic Fibrosis
11. ON WHAT CHROMOSOME IS THE GENE LOCATED? You should be able to get this information by clicking on the description of the gene. Chromosome 7
12. Scroll to the first described sequence that does not have 100% Query Cover. WHAT ORGANISM IS THE SOURCE OF THIS DNA? Olive Baboon
13. HOW MANY GAPS OCCUR BETWEEN THE TWO SEQUENCES (THE ONE YOU SUBMITTED AND THE FIRST ONE THAT HAS < 100% QUERY COVER)? There are no gaps.
14. WHAT IS A GAP IN SEQUENCE ALIGNMENTS? (This is something you'll have to search for elsewhere.) Gaps are the bits that get left behind when DNA or protein sequences are aligned. These are insertions and deletions within the genome.

BLAST has more sequences than 'just' human gene sequences. Also, you can also do BLAST searches using an accession number that has been assigned to a particular sequence when it has been entered into the database. Go back to the Blast home page (http://blast.ncbi.nlm.nih.gov/Blast.cgi) and again choose "Nucleotide Blast". Look up the following sequence using the given accession number. (Again, click on the "BLAST" button at the bottom of the page after you have entered the accession number.)
15. STATE WHAT THE FOLLOWING GENE IS: NC_045512.2
(Give the complete description of the gene or gene product.) Severe acute respiratory syndrome coronavirus
16. SCROLL DOWN THE LIST BELOW THIS SEQUENCE. THERE ARE MANY SEQUENCES THAT LOOK SIMILAR. CAN YOU DETECT WHAT IS DIFFERENT ABOUT THE OTHER SEQUENCES? The base pairs are different

Introduction to Swiss-Prot to Study Protein Sequences

Assignment Goal: To use the Internet-based site ExPASy (Expert Protein Analysis System) to translate cDNA, and the Internet-based database UniProt KB/Swiss-Prot to access a complete polypeptide.
17. WHAT IS cDNA? Complementary DNA
Assignment: Access the BLAST site at http://blast.ncbi.nlm.nih.gov/Blast.cgi
Click on "Nucleotide Blast"

Enter the following sequence:
ACATTTGCTTCTGACACAATTGTGTTCACTAGCAACCTCAAACAGACACCATGGTGCATCTG
ACTCCTGAGGAGAAGTCTGCCGTTACTGCCCTGTGGGGCAAGGTGAACGTGGATGAAGTTG
GTGGTGAGGCCCTGGGCAG

18. USING THE SAME PROGRAM YOU USED IN THE INTRODUCTION TO BLAST ABOVE, WHAT IS THE SEQUENCE MATCH? Homo sapiens partial HBB gene for hemoglobin beta chain
Now access the ExPASt translate tool at https://web.expasy.org/translate/
Enter the above DNA sequence.
Click "Translate Sequence".
19. HOW MANY 5' TO 3' FRAMES DID YOU OBTAIN? 3 frames
20. BASED UPON THE LENGTH OF THE POLYPEPTIDE, WHICH FRAME(S) IS (ARE) MOST LIKELY THE CORRECT ONE? Frame 2 is most likely the correct one.

Amino Acid Sequence Comparisons

Assignment Goal: To use the Internet-based site ExPASy program SIM to align two amino acid sequences. Knowing the sources of these sequences will allow one to determine the mutation and potential cause of a human disease.
Assignment: Access the ExPASy site at https://web.expasy.org/sim/
Copy and paste each of the following sequences into the “Sequence” text boxes in SIM.

Person 1/Sequence 1:
MGAPACALALCVAVAIVAGASSESLGTEQRVVGRAAEVPGPEPGQQEQLVFGSGDAVELSCPPPGGGP
MGPTVWVKDGTGLVPSERVLVGPQRLQVLNASHEDSGAYSCRQRLTQRVLCHFSVRVTDAPSSGDDE
DGEDEAEDTGVDTGAPYWTRPERMDKKLLAVPAANTVRFRCPAAGNPTPSISWLKNGREFRGEHRIG
GIKLRHQQWSLVMESVVPSDRGNYTCVVENKFGSIRQTYTLDVLERSPHRPILQAGLPANQTAVLGSD
VEFHCKVYSDAQPHIQWLKHVEVNGSKVGPDGTPYVTVLKTAGANTTDKELEVLSLHNVTFEDAGEY
TCLAGNSIGFSHHSAWLVVLPAEEELVEADEAGSVYAGILSYGVGFFLFILVVAAVTLCRLRSPPKKGL
GSPTVHKISRFPLKRQVSLESNASMSSNTPLVRIARLSSGEGPTLANVSELELPADPKWELSRARLTLGK
PLGEGCFGQVVMAEAIGIDKDRAAKPVTVAVKMLKDDATDKDLSDLVSEMEMMKMIGKHKNIINLL
GACTQGGPLYVLVEYAAKGNLREFLRARRPPGLDYSFDTCKPPEEQLTFKDLVSCAYQVARGMEYLA
SQKCIHRDLAARNVLVTEDNVMKIADFGLARDVHNLDYYKKTTNGRLPVKWMAPEALFDRVYTHQS
DVWSFGVLLWEIFTLGGSPYPGIPVEELFKLLKEGHRMDKPANCTHDLYMIMRECWHAAPSQRPTFK
QLVEDLDRVLTVTSTDEYLDLSAPFEQYSPGGQDTPSSSSGDDSVFAHDLLPPAPPSSGGSRT

Person 2/Sequence 2:
MGAPACALALCVAVAIVAGASSESLGTEQRVVGRAAEVPGPEPGQQEQLVFGSGDAVELSCPPPGGGP
MGPTVWVKDGTGLVPSERVLVGPQRLQVLNASHEDSGAYSCRQRLTQRVLCHFSVRVTDAPSSGDDE
DGEDEAEDTGVDTGAPYWTRPERMDKKLLAVPAANTVRFRCPAAGNPTPSISWLKNGREFRGEHRIG
GIKLRHQQWSLVMESVVPSDRGNYTCVVENKFGSIRQTYTLDVLERSPHRPILQAGLPANQTAVLGSD
VEFHCKVYSDAQPHIQWLKHVEVNGSKVGPDGTPYVTVLKTAGANTTDKELEVLSLHNVTFEDAGEY
TCLAGNSIGFSHHSAWLVVLPAEEELVEADEAGSVYAGILSYRVGFFLFILVVAAVTLCRLRSPPKKGL
GSPTVHKISRFPLKRQVSLESNASMSSNTPLVRIARLSSGEGPTLANVSELELPADPKWELSRARLTLGK
PLGEGCFGQVVMAEAIGIDKDRAAKPVTVAVKMLKDDATDKDLSDLVSEMEMMKMIGKHKNIINLL
GACTQGGPLYVLVEYAAKGNLREFLRARRPPGLDYSFDTCKPPEEQLTFKDLVSCAYQVARGMEYLA
SQKCIHRDLAARNVLVTEDNVMKIADFGLARDVHNLDYYKKTTNGRLPVKWMAPEALFDRVYTHQS
DVWSFGVLLWEIFTLGGSPYPGIPVEELFKLLKEGHRMDKPANCTHDLYMIMRECWHAAPSQRPTFK
QLVEDLDRVLTVTSTDEYLDLSAPFEQYSPGGQDTPSSSSSGDDSVFAHDLLPPAPPSSGGSRT
Click the “User entered sequence” button for each sequence entered.
Submit the sequences for comparison.
21. DO YOU SEE ANY DIFFERENCES BETWEEN THE TWO AMINO ACID SEQUENCES? (Look for the absence of an asterisk, which indicates the same amino acid in both sequences.) Yes

`UserSeq1      361 EELVEADEAGSVYAGILSYGVGFFLFILVVAAVTLCRLRSPPKKGLGSPTVHKISRFPLK` The G in sequence 1 is an R in sequence 2

22. WHY WOULD DIFFERENCES BE IMPORTANT AND POSSIBLY CLINICALLY USEFUL? Small changes in the amino acid sequence of a protein can cause devastating genetic diseases such as sickle cell anemia and Huntington's disease.
Return to the BLAST home page (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Run a **PROTEIN** BLAST search to identify the polypeptide which you have been analyzing. (You may use either sequence.)
23. WHAT IS THE PROTEIN THAT YOU WERE ANALYZING?
Click on "Description", then choose "Gene". Fibroblast Growth Factor Receptor 3 (Homo Sapiens(humans)).

24-25. REFLECT ON ONE THING YOU ALREADY KNEW, AND ONE THING THAT YOU LEARNED FROM DOING THIS ASSIGNMENT. I was able to learn how to successfully BLAST an amino sequence and determine which organism matches the sequence and characteristics of the gene. I have used BLAST and I already knew that the results are in chronological order based on percentage of closeness.

Works Cited

Meng X, Clews J, Kargas V, Wang X, Ford RC. The cystic fibrosis transmembrane conductance regulator (CFTR) and its stability. Cell Mol Life Sci. 2017 Jan;74(1):23-38. doi: 10.1007/s00018-016-2386-8. Epub 2016 Oct 12. PMID: 27734094; PMCID: PMC5209436.