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 18

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.

* Number of Genes: Contains over 600 genes
* Number of Base Pairs: Contains over 70 millions base pairs.

Scan the chromosome map.

4. LIST ONE GENE THAT IS LOCATED ON THIS CHROMOSOME.

* Gene DPC4

5. STATE THE **NORMAL** FUNCTION OF THE GENE YOU LISTED IN #4.

* DPC4 regulates development, signaling pathways, and eventually slows cellular growth.

6. STATE THE POSSIBLE DISEASE(S) RELATED TO THIS GENE. (This should be possible by clicking on the gene you stated in #4.)

* Pancreatic Cancer

**GenBank**

7. WHAT IS GENBANK? (You can do an Internet search to find this information.)

* a comprehensive public database of nucleotide sequences and supporting bibliographical and biological annotation.
* a database for genetic sequences, it is run by the National Institute of Health.

**Introduction to BLAS**T

**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 CFTR on chromosome 7

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.

* The actual region is a promoter region which will not be transcribed. However, the CFTR protein, or Cystic fibrosis transmembrane conductance regulator protein, acts as an ion channel transport protein.
* https://pubmed.ncbi.nlm.nih.gov/30340955/

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?

* Papio Anubis (Olive Baboon)

13. HOW MANY GAPS OCCUR BETWEEN THE TWO SEQUENCES (THE ONE YOU SUBMITTED AND THE FIRST ONE THAT HAS < 100% QUERY COVER)?

* One

14. WHAT IS A GAP IN SEQUENCE ALIGNMENTS? (This is something you’ll have to search for elsewhere.)

* a gap in sequence alignment is what is not included, or gets left behind, when one attempts to align DNA or protein sequences. They also report that when you have a gap you must “use padding or null characters to match homologous residues”

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.)

* SARS-Cov-2, which is also known as severe acute respiratory syndrome coronavirus 2

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?

* What is different about the other sequences is the ID number, the lengths , along with the reference sequences.

**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: ACATTTGCTTCTGACACAATTGTGTTCACTAGCAACCTCAAACAGACACCATGGTGCATCTGACTCCTGAGGAGAAGTCTGCCGTTACTGCCCTGTGGGGCAAGGTGAACGTGGATGAAGTTGGTGGTGAGGCCCTGGGCAG

18. USING THE SAME PROGRAM YOU USED IN THE INTRODUCTION TO BLAST ABOVE, WHAT IS THE SEQUENCE MATCH?

* HBB gene/ 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?

* 6 frames in total, which most of them go from 5’ to 3’

20. BASED UPON THE LENGTH OF THE POLYPEPTIDE, WHICH FRAME(S) IS (ARE) MOST LIKELY THE CORRECT ONE?

* Frame 3

**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:

MGAPACALALCVAVAIVAGASSESLGTEQRVVGRAAEVPGPEPGQQEQLVFGSGDAVELSCPPPGGGPMGPTVWVKDGTGLVPSERVLVGPQRLQVLNASHEDSGAYSCRQRLTQRVLCHFSVRVTDAPSSGDDEDGEDEAEDTGVDTGAPYWTRPERMDKKLLAVPAANTVRFRCPAAGNPTPSISWLKNGREFRGEHRIGGIKLRHQQWSLVMESVVPSDRGNYTCVVENKFGSIRQTYTLDVLERSPHRPILQAGLPANQTAVLGSDVEFHCKVYSDAQPHIQWLKHVEVNGSKVGPDGTPYVTVLKTAGANTTDKELEVLSLHNVTFEDAGEYTCLAGNSIGFSHHSAWLVVLPAEEELVEADEAGSVYAGILSYGVGFFLFILVVAAVTLCRLRSPPKKGLGSPTVHKISRFPLKRQVSLESNASMSSNTPLVRIARLSSGEGPTLANVSELELPADPKWELSRARLTLGKPLGEGCFGQVVMAEAIGIDKDRAAKPVTVAVKMLKDDATDKDLSDLVSEMEMMKMIGKHKNIINLLGACTQGGPLYVLVEYAAKGNLREFLRARRPPGLDYSFDTCKPPEEQLTFKDLVSCAYQVARGMEYLASQKCIHRDLAARNVLVTEDNVMKIADFGLARDVHNLDYYKKTTNGRLPVKWMAPEALFDRVYTHQSDVWSFGVLLWEIFTLGGSPYPGIPVEELFKLLKEGHRMDKPANCTHDLYMIMRECWHAAPSQRPTFKQLVEDLDRVLTVTSTDEYLDLSAPFEQYSPGGQDTPSSSSGDDSVFAHDLLPPAPPSSGGSRT

Person 2/Sequence 2:

MGAPACALALCVAVAIVAGASSESLGTEQRVVGRAAEVPGPEPGQQEQLVFGSGDAVELSCPPPGGGPMGPTVWVKDGTGLVPSERVLVGPQRLQVLNASHEDSGAYSCRQRLTQRVLCHFSVRVTDAPSSGDDEDGEDEAEDTGVDTGAPYWTRPERMDKKLLAVPAANTVRFRCPAAGNPTPSISWLKNGREFRGEHRIGGIKLRHQQWSLVMESVVPSDRGNYTCVVENKFGSIRQTYTLDVLERSPHRPILQAGLPANQTAVLGSDVEFHCKVYSDAQPHIQWLKHVEVNGSKVGPDGTPYVTVLKTAGANTTDKELEVLSLHNVTFEDAGEYTCLAGNSIGFSHHSAWLVVLPAEEELVEADEAGSVYAGILSYRVGFFLFILVVAAVTLCRLRSPPKKGLGSPTVHKISRFPLKRQVSLESNASMSSNTPLVRIARLSSGEGPTLANVSELELPADPKWELSRARLTLGKPLGEGCFGQVVMAEAIGIDKDRAAKPVTVAVKMLKDDATDKDLSDLVSEMEMMKMIGKHKNIINLL

GACTQGGPLYVLVEYAAKGNLREFLRARRPPGLDYSFDTCKPPEEQLTFKDLVSCAYQVARGMEYLASQKCIHRDLAARNVLVTEDNVMKIADFGLARDVHNLDYYKKTTNGRLPVKWMAPEALFDRVYTHQSDVWSFGVLLWEIFTLGGSPYPGIPVEELFKLLKEGHRMDKPANCTHDLYMIMRECWHAAPSQRPTFKQLVEDLDRVLTVTSTDEYLDLSAPFEQYSPGGQDTPSSSSSGDDSVFAHDLLPPAPPSSGGSRT

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, there are a total of two differences

22. WHY WOULD DIFFERENCES BE IMPORTANT AND POSSIBLY CLINICALLY USEFUL?

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.)

* It can be clinically useful because it can help with evolutionary trends and relationship with/ against other species.

23. WHAT IS THE PROTEIN THAT YOU WERE ANALYZING?

Click on “Description”, then choose “Gene”.

* FGFR3

24-25. REFLECT ON ONE THING YOU ALREADY KNEW, AND ONE THING THAT YOU LEARNED FROM DOING THIS ASSIGNMENT.

* I used BLAST in General Biology lab, so I was aware of this program when I was completing this assignment. One thing I learned from doing this assignment is all the genetic research and information that is open to the public. I learned that the mutation in FGFR3 gene can lead to the misshapen head of some people/children.