Old Dominion University/BIOL294-Genetics/Rinehart-Kim/Module 9/Genome 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: https://medlineplus.gov/genetics/
* **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.

**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. WHICH CHROMOSOME DID YOU CHOOSE? I chose chromosome 21

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.

Chromosome 21 contains over 400 genes, and 40 million base pairs.

Scan the chromosome map.

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

One gene on chromosome 21 is the APS1 gene which encodes for Autoimmune polyglandular syndrome.

5. STATE THE NORMAL FUNCTION OF THE GENE YOU LISTED IN #4. This is possible by clicking on the gene you stated in #4. It is important that you state the NORMAL physiological function of the gene product you select.

The normal physiological function of the gene APS1 found on chromosome 21 is autoimmune regulator. A protein product of this gene is a transcription factor, which is used in gene expression.

**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 cloned a piece of gene in the laboratory in which you work:

aattggaagc aaatgacatc acagcaggtc agagaaaaag ggttgagcgg caggcaccca gagtagtagg tctttggcat taggagcttg agcccagacg gccctagcag ggaccccagc

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.

6. WHAT IS THE TOP SEQUENCE DESCRIPTION MATCH FOR YOUR QUERY SEQUENCE? For this answer, you should give the description listed. Do not choose a Predicted sequence.

The top sequence description match for the query sequence was Homo Sapiens CFTR promoter region (LOC111674463) on chromosome 7.

7. IS THIS A SEQUENCE FOR A PROTEIN OR ANOTHER PART OF THE GENE? IF IT IS “ANOTHER PART OF THE GENE”, EXPLAIN ITS PURPOSE.

This sequence is for a promoter region of a gene. This region’s purpose is to be the binding site of RNA polymerase, which initiates transcription. The protein associated with this region is the cystic fibrosis transmembrane conductance regulator.

8.. WHAT DOES THE ENCODED PROTEIN ASSOCIATED WITH THE ABOVE SEQUENCE DO IN THE BODY? Search the PubMed site at [www.ncbi.nlm.nih.gov/entrez/query.fcgi?DB=pubmed](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?DB=pubmed) to answer this question. Under “Article types,” choose “Review”. CITE THE PAPER YOU USED TO DETERMINE THE PURPOSE OF THE ENCODED PROTEIN.

The cyctic fibrosis transmembrane conductance regulator (CFTR) functions as a regulator of different ion channels and participates in intracellular transport processes.

Source used: Vankeerberghen, Anne et al. “The cystic fibrosis transmembrane conductance regulator: an intriguing protein with pleiotropic functions.” *Journal of cystic fibrosis: official journal of the European Cystic Fibrosis Society* vol. 1,1 (2002): 13-29. doi:10.1016/s1569-1993(01)00003-0

Click on the top match to find the following.

9. A MUTATED FORM OF THIS GENE IS RESPONSIBLE FOR A WELL-STUDIED DISEASE. WHAT IS THAT DISEASE? You should be able to get this information from the description of the gene. You may need to “probe” the gene description

The disease caused by a mutated form of this gene is cystic fibrosis (CF)

10. ON WHAT CHROMOSOME IS THE GENE LOCATED? You should be able to get this information by clicking on the description of the gene.

The CFTR gene is located on chromosome 7

11. Return to the original nucleotide sequence alignment descriptions. WHAT SPECIES (STATE THE SCIENTIFIC NAME) OTHER THAN *HOMO SAPIENS* ALSO HAS A 100% IDENTITY (Ident) FOR THIS SEQUENCE? USE THE TOP SEQUENCE LISTED, BUT DO NOT USE THE PREDICTED SEQUENCES.

*Pongo abelii*.

12. WHAT IS THE COMMON NAME FOR THIS SPECIES?

Sumatran orangutan

13. DOES IT SURPRISE YOU THAT THIS SPECIES ALSO HAS A 100% SIMILARITY IN IDENTITY? WHY OR WHY NOT?

No, it does not. Considering the genetic proximity of humans and primates, this match was expected.

14. DESCRIBE THE FIRST MATCH THAT HAS LESS THAN 100% QUERY COVER BUT IS NOT PREDICTED OR *HOMO SAPIENS*. STATE THE SCIENTIFIC AND COMMON NAMES.

The first match that is less than 100 percent is *saimiri boliviensis boliviensis*

15. Click on the description to answer this question. HOW MANY GAPS OCCUR BETWEEN THE TWO SEQUENCES (THE ONE YOU SUBMITTED AND THE FIRST ONE THAT HAS LESS THAN 100% QUERY COVER)?

*Homo sapiens* CFTR promoter region: 0/120 (0%) gaps

*saimiri boliviensis boliviensis* clone: 1/119 (0%) gaps

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

A gap in sequence alignments occur when one or more amino acid residues are deleted or inserted into a sequence. This allows for better alignment for matching sequences.

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 ([www.ncbi.nlm.nih.gov/BLAST](http://www.ncbi.nlm.nih.gov/BLAST).cgi ) and again choose “Nucleotide Blast”. Look up the following sequences using the given accession numbers. (Under “Program Selection” near the bottom of the page, choose “somewhat similar sequence (blastn)”. (Again, click on the “BLAST” button at the bottom of the page after you have entered the accession number.)

FOR EACH, STATE WHAT THE GENE IS (#16-19). Give the description of the gene and gene product. You do not need to state the organism source.

17. NM\_145556: Mus musculus TAR DNA binding protein (Tardbp), transcript variant 1, mRNA

18. NM\_013444: Homo sapiens ubiquilin 2 (UBQLN2), mRNA

19. NM\_001010850: Homo sapiens fusion (involved in t(12;16) in malignant liposarcoma) (FUS), transcript variant 2, mRNA

20. KJ174530: Homo sapiens superoxide dismutase-1 (SOD-1) gene, exon 1 and partial cds

21. Search Google to answer the following: WHAT DISEASE IS ASSOCIATED WITH MUTATIONS OF THE GENES REFERENCED IN #17-#20? WHAT IS A “COMMON NAME” OF THE DISEASE? (The name of a person; Hint, hint…We just finished the World Series…)

BLAST is possible because of the submission of DNA sequences to GenBank.

The disease associated with genes referenced in #17-20 is Amyotrophic lateral sclerosis, commonly known as ALS.

Source: <https://medlineplus.gov/genetics/gene/sod1/#:~:text=At%20least%20200%20mutations%20in,an%20inability%20to%20control%20movement>.

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

A genbank is a publicly available database that contains nucleotides sequences from a variety of different organisms.

**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.

23. WHAT IS cDNA? How can we obtain cDNA in the lab?

**Assignment**: Access the BLAST site at [www.ncbi.nlm.nih.gov/BLAST](http://www.ncbi.nlm.nih.gov/BLAST).cgi

Click on “Nucleotide Blast”

Enter the following sequence: ACATTTGCTTCTGACACAATTGTGTTCACTAGCAACCTCAAACAGACACCATGGTGCATCTGACTCCTGAGGAGAAGTCTGCCGTTACTGCCCTGTGGGGCAAGGTGAACGTGGATGAAGTTGGTGGTGAG

GCCCTGGGCAG

24. 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, exon 1, isolate 04593664

Now access the Expasy translate tool at https://web.expasy.org/translate/.

Enter the above DNA sequence.

Click “Translate Sequence”.

25. WHAT IS AN OPEN READING FRAME?

An open reading frame is a section of DNA that does not contain stop codons.

26. ALL OF THE PROPOSED OPEN READING FRAMES (HIGHLIGHTED IN RED) START WITH THE AMINO ACID “M”. FROM WHAT YOU KNOW ABOUT POLYPEPTIDES, WHAT IS “M”?

“M” is Methionine

27. WHICH 5’ TO 3’ FRAME IS MOST LIKELY TO BE AN OPEN READING FRAME? WHY DID YOU CHOOSE THAT FRAME?

5’ to 3’ Frame 3 is most likely to be an open reading frame because it has a Methionine in the beginning, and the highlighted portion extends longer than any other frame.

**Amino Acid Sequence Comparisons**

**Assignment Goal**: To use the Internet-based site Expasy SIM program 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 as User-entered sequence.

Person 1/Sequence 1:

MGAPACALALCVAVAIVAGASSESLGTEQRVVGRAAEVPGPEPGQQEQLVFGSGDAVELSCPPPGGGPMGPTVWVKDGTGLVPSERVLVGPQRLQVLNASHEDSGAYSCRQRLTQRVLCHFSVRVTDAPSSGDDEDGEDEAEDTGVDTGAPYWTRPERMDKKLLAVPAANTVRFRCPAAGNPTPSISWLKNGREFRGEHRIGGIKLRHQQWSLVMESVVPSDRGNYTCVVENKFGSIRQTYTLDVLERSPHRPILQAGLPANQTAVLGSDVEFHCKVYSDAQPHIQWLKHVEVNGSKVGPDGTPYVTVLKTAGANTTDKELEVLSLHNVTFEDAGEYTCLAGNSIGFSHHSAWLVVLPAEEELVEADEAGSVYAGILSYGVGFFLFILVVAAVTLCRLRSPPKKGLGSPTVHKISRFPLKRQVSLESNASMSSNTPLVRIARLSSGEGPTLANVSELELPADPKWELSRARLTLGKPLGEGCFGQVVMAEAIGIDKDRAAKPVTVAVKMLKDDATDKDLSDLVSEMEMMKMIGKHKNIINLLGACTQGGPLYVLVEYAAKGNLREFLRARRPPGLDYSFDTCKPPEEQLTFKDLVSCAYQVARGMEYLASQKCIHRDLAARNVLVTEDNVMKIADFGLARDVHNLDYYKKTTNGRLPVKWMAPEALFDRVYTHQSDVWSFGVLLWEIFTLGGSPYPGIPVEELFKLLKEGHRMDKPANCTHDLYMIMRECWHAAPSQRPTFKQLVEDLDRVLTVTSTDEYLDLSAPFEQYSPGGQDTPSSSSGDDSVFAHDLLPPAPPSSGGSRT

Person 2/Sequence 2:

MGAPACALALCVAVAIVAGASSESLGTEQRVVGRAAEVPGPEPGQQEQLVFGSGDAVELSCPPPGGGPMGPTVWVKDGTGLVPSERVLVGPQRLQVLNASHEDSGAYSCRQRLTQRVLCHFSVRVTDAPSSGDDEDGEDEAEDTGVDTGAPYWTRPERMDKKLLAVPAANTVRFRCPAAGNPTPSISWLKNGREFRGEHRIGGIKLRHQQWSLVMESVVPSDRGNYTCVVENKFGSIRQTYTLDVLERSPHRPILQAGLPANQTAVLGSDVEFHCKVYSDAQPHIQWLKHVEVNGSKVGPDGTPYVTVLKTAGANTTDKELEVLSLHNVTFEDAGEYTCLAGNSIGFSHHSAWLVVLPAEEELVEADEAGSVYAGILSYRVGFFLFILVVAAVTLCRLRSPPKKGLGSPTVHKISRFPLKRQVSLESNASMSSNTPLVRIARLSSGEGPTLANVSELELPADPKWELSRARLTLGKPLGEGCFGQVVMAEAIGIDKDRAAKPVTVAVKMLKDDATDKDLSDLVSEMEMMKMIGKHKNIINLL

GACTQGGPLYVLVEYAAKGNLREFLRARRPPGLDYSFDTCKPPEEQLTFKDLVSCAYQVARGMEYLASQKCIHRDLAARNVLVTEDNVMKIADFGLARDVHNLDYYKKTTNGRLPVKWMAPEALFDRVYTHQSDVWSFGVLLWEIFTLGGSPYPGIPVEELFKLLKEGHRMDKPANCTHDLYMIMRECWHAAPSQRPTFKQLVEDLDRVLTVTSTDEYLDLSAPFEQYSPGGQDTPSSSSSGDDSVFAHDLLPPAPPSSGGSRT

Submit the sequences for comparison.

28. 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, I did see differences in between the two amino acids. For instance, there is an absence of an asterisk on the polypeptide 361, which differs by the amino acids G and R.

29. IF YOU SAW DIFFERENCES, WHAT WERE THEY?

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.)

30. WHAT IS THE FUNCTION OF THIS PROTEIN?

The identification of the polypeptide from the second sequence is a fibroblast growth factor receptor 3 isoform 1 precursor in Homo sapiens. Its function is to regulate growth of bone by limiting ossification.

31. WHAT HUMAN DISEASE IS CAUSED BY A MUTATION IN THIS GENE?

A mutation of the FGFR3 gene causes Muenke syndrome, which causes cariosynostosis.

Source: https://medlineplus.gov/genetics/gene/fgfr3/#:~:text=A%20single%20mutation%20in%20the,foot%20abnormalities%2C%20and%20developmental%20delay.

32. REFLECT ON ONE THING THAT YOU LEARNED FROM DOING THIS ASSIGNMENT. Please be honest. If you didn’t learn anything, admit it…

This assignment showed me that there are many available tools online that can help build my knowledge and experience when it comes to sequencing of genes. I am eager to continue exploring these different websites and learning more about ways to use them. Thank you!