Old Dominion University/BIOL294-Genetics/Rinehart-Kim/Module 9/Genome Assignment/FA23

- You are to address the statements/questions that are in red. You do not need to use complete sentences in your replies. You may either answer directly on this worksheet, or put just your answers on another word document.
- The following is an excellent reference: Genetics Home Reference: http://ghr.nlm.nih.gov
- **As of now (11/19), 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. WHAT CHROMOSOME DID YOU CHOOSE?
Chromosome 19

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.
Over 1700 genes and over 60 million base pairs

Scan the chromosome map.
4. LIST ONE GENE THAT IS LOCATED ON THIS CHROMOSOME.
GNG7

5. STATE THE 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 physiological function of the gene product you select, not the disease listed for the gene.
GNG7 has physiological functions that are speculated to occur before or within fear responses and locomotory behavior and that are predicted to include the ability to start G-protein beta-subunit binding activity and to help regulate adenylate cyclase activity. There is also evidence that shows its involvement in the receptor guanylyl cyclase signaling and G protein-coupled receptor signaling pathways.

Introduction to Nucleotide 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 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 **SECOND** SEQUENCE DESCRIPTION MATCH FOR YOUR QUERY SEQUENCE? For this answer, you should give the description listed. Do not give the actual sequence.
Homo sapiens CFTR (CFTR) gene, partial cds

7. WHAT DOES THE ENCODED PROTEIN DO IN THE BODY? To find this information, click on the sequence description and you will get a nucleotide match. To the right of that nucleotide match you will find "Related Information" and under that you will find "Gene". Click on "Gene" to find out more information about this protein.
Being a part of the ATP-binding cassette (ABC) transporter superfamily, this unique encoded protein regulates ion and water secretion and absorption in epithelial tissues in its role as a chloride channel.

8. 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

9. ON WHAT CHROMOSOME IS THE GENE LOCATED? You should be able to get this information by looking at the "Genomic context" which is just below the description of the gene.
Chromosome 7

10. Return to the original nucleotide sequence alignment descriptions. CHOOSE A SPECIES (STATE THE SCIENTIFIC NAME) OTHER THAN *HOMO SAPIENS* THAT ALSO HAS A 100% IDENTITY (Per. Ident) FOR THIS SEQUENCE? There will be about 3 species that you could name, so choose one of them.
Pongo abelii

11. WHAT IS THE COMMON NAME FOR THIS SPECIES?
Sumatran orangutan [1]

12. DOES IT SURPRISE YOU THAT THIS SPECIES ALSO HAS A 100% SIMILARITY IN IDENTITY? WHY OR WHY NOT?
This does not surprise me as orangutans are primates like us. Primate (animal) genetics are very similar to ours, hence why we are believed to have evolved from them.

13. Return to the original nucleotide sequence alignment **DESCRIPTION**. Find the first match that has less than 100% similarity identity. Click on the description to answer this question.
a. WHAT IS THE GENUS AND SPECIES WITH THIS NUCLEOTIDE SEQUENCE?
Sapajus, *Sapajus apella*

b. WHAT IS THE COMMON NAME?
Black-capped or Tufted capuchin [2]

c. HOW MANY GAPS OCCUR BETWEEN THE TWO SEQUENCES (THE ONE YOU ORIGINALLY SUBMITTED AND ONE THAT HAS LESS THAN 100% QUERY COVER)? The gap information is just above the sequence alignment.
1/119

14. WHAT IS A GAP IN SEQUENCE ALIGNMENTS? (This is something you'll have to search for on the Internet.)
Gaps are created by insertions and deletions that occur when a gene mutates and are then shown as empty spaces when comparing two similar sequences. [3]

You can also do BLAST searches using an accession number that has been assigned to a particular sequence when was entered into the database. Go back to the Blast home page (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 (#15-18). (Again, give the description of the gene or gene product, not the nucleotide sequence.)
15. NM_145556
Mus musculus TAR DNA binding protein (Tardbp), transcript variant 1, mRNA

16. NM_013444
Homo sapiens ubiquilin 2 (UBQLN2), mRNA

17. NM_001010850
Homo sapiens fusion (involved in t(12;16) in malignant liposarcoma) (FUS), transcript variant 2, mRNA

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

19. Search Google to answer the following: WHAT DISEASE IS ASSOCIATED WITH MUTATIONS OF THE GENES REFERENCED IN #15-#18? WHAT IS A “COMMON NAME” OF THE DISEASE? (The name of a person; Hint, hint…Baseball season just finished…)
ALS – amyotrophic lateral sclerosis (all name ALS within each gene’s description) also known as Lou Gehrig's disease [4]

20. BLAST is possible because of the submission of DNA sequences to GenBank. WHAT IS GENBANK? (You can do an Internet search to find this information.)
GenBank is database of all known genetic sequences. This database was constructed to have a storage method and public access point for the increasingly large number of sequences found in nature.[5]

Introduction to Protein BLAST

Assignment Goal: To use the Internet-based site BLAST, Basic Local Alignment Search Tool (NCBI), to 1) translate cDNA and 2) to search for similarities between amino acid sequences.
Assignment: Access the BLAST site at http://blast.ncbi.nlm.nih.gov/Blast.cgi
Click on “tblastx”. (This is a smaller button next to “Nucleotide BLAST”.
21. First, answer this question: WHAT IS cDNA? (Again, you can do an Internet search to find this information.)
copy DNA (previous information learned, did not need to search)

Enter the following cDNA sequence on “blastx”:
ACATTTGCTTCTGACACAATTGTGTTCACTAGCAACCTCAAACAGACACCATGGTGCATCTGACTC
CTGAGGAGAAGTCTGCCGTTACTGCCCTGTGGGGCAAGGTGAACGTGGATGAAGTTGGTGGTGAG
GCCCTGGGCAG
22. WHAT IS THE SEQUENCE MATCH?
beta-globin [Homo sapiens]

Assignment: While still on BLAST, choose “Protein BLAST”. Check the box which says “Align two or more sequences”.
Copy and paste the following sequence into the “Enter Query Sequence” box:
Person 1/Sequence 1:

MGAPACALALCVAVAIVAGASSESLGTEQRVVGRAAEVPGPEPGQQEQLVFGSGDAVELSCPP
PGGGPMGPTVWVKDGTGLVPSERVLVGPQRLQVLNASHEDSGAYSCRQRLTQRVLCHFSVRV
TDAPSSGDDEDGEDEAEDTGVDTGAPYWTRPERMDKKLLAVPAANTVRFRCPAAGNPTPSIS
WLKNGREFRGEHRIGGIKLRHQQWSLVMESVVPSDRGNYTCVVENKFGSIRQTYTLDVLERSP

HRPILQAGLPANQTAVLGSDVEFHCKVYSDAQPHIQWLKHVEVNGSKVGPDGTPYVTVLKTA
GANTTDKELEVLSLHNVTFEDAGEYTCLAGNSIGFSHHSAWLVVLPAEEELVEADEAGSVYAG
ILSYGVGFFLFILVVAAVTLCRLRSPPKKGLGSPTVHKISRFPLKRQVSLESNASMSSNTPLVRIA
RLSSGEGPTLANVSELELPADPKWELSRARLTLGKPLGEGCFGQVVMAEAIGIDKDRAAKPVT
VAVKMLKDDATDKDLSDLVSEMEMMKMIGKHKNIINLLGACTQGGPLYVLVEYAAKGNLRE
FLRARRPPGLDYSFDTCKPPEEQLTFKDLVSCAYQVARGMEYLASQKCIHRDLAARNVLVTED
NVMKIADFGLARDVHNLDYYKKTTNGRLPVKWMAPEALFDRVYTHQSDVWSFGVLLWEIFT
LGGSPYPGIPVEELFKLLKEGHRMDKPANCTHDLYMIMRECWHAAPSQRPTFKQLVEDLDRVL
TVTSTDEYLDLSAPFEQYSPGGQDTPSSSSGDDSVFAHDLLPPAPPSSGGSRT

Copy and paste the following sequence into the “Enter Subject Sequence” box:
Person 2/Sequence 2:

MGAPACALALCVAVAIVAGASSESLGTEQRVVGRAAEVPGPEPGQQEQLVFGSGDAVELSCPP
PGGGPMGPTVWVKDGTGLVPSERVLVGPQRLQVLNASHEDSGAYSCRQRLTQRVLCHFSVRV
TDAPSSGDDEDGEDEAEDTGVDTGAPYWTRPERMDKKLLAVPAANTVRFRCPAAGNPTPSIS
WLKNGREFRGEHRIGGIKLRHQQWSLVMESVVPSDRGNYTCVVENKFGSIRQTYTLDVLERSP
HRPILQAGLPANQTAVLGSDVEFHCKVYSDAQPHIQWLKHVEVNGSKVGPDGTPYVTVLKTA
GANTTDKELEVLSLHNVTFEDAGEYTCLAGNSIGFSHHSAWLVVLPAEEELVEADEAGSVYAG
ILSYRVGFFLFILVVAAVTLCRLRSPPKKGLGSPTVHKISRFPLKRQVSLESNASMSSNTPLVRIA
RLSSGEGPTLANVSELELPADPKWELSRARLTLGKPLGEGCFGQVVMAEAIGIDKDRAAKPVT
VAVKMLKDDATDKDLSDLVSEMEMMKMIGKHKNIINLL
GACTQGGPLYVLVEYAAKGNLREFLRARRPPGLDYSFDTCKPPEEQLTFKDLVSCAYQVARG
MEYLASQKCIHRDLAARNVLVTEDNVMKIADFGLARDVHNLDYYKKTTNGRLPVKWMAPEA
LFDRVYTHQSDVWSFGVLLWEIFTLGGSPYPGIPVEELFKLLKEGHRMDKPANCTHDLYMIMR
ECWHAAPSQRPTFKQLVEDLDRVLTVTSTDEYLDLSAPFEQYSPGGQDTPSSSSSGDDSVFAHD
LLPPAPPSSGGSRT

Click on the “BLAST” box.
After you receive results, choose the “Description” box for more information about the alignment.

23. DO YOU SEE ANY DIFFERENCES BETWEEN THE TWO AMINO ACID SEQUENCES? (Look for a space between same amino acid comparison for both sequences.)
There is one space between the two amino acid sequences.

24. IF YOU SAW DIFFERENCES, WHAT WERE THEY?
There was an R in the place of a G in the Subject sequence.

```
Query  361  EELVEADEAGSVYAGILSYGVGFFLFILVVAAVTLCRLRSPPKKGLGSPTVHKISRFPLK  420
            EELVEADEAGSVYAGILSY VGFFLFILVVAAVTLCRLRSPPKKGLGSPTVHKISRFPLK
Sbjct  361  EELVEADEAGSVYAGILSYRVGFFLFILVVAAVTLCRLRSPPKKGLGSPTVHKISRFPLK  420
```

25. ARE THERE ANY GAPS IN THE SEQUENCE ALIGNMENT?
There is one gap (1/806).

Return to the BLAST home page (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Return to the Protein BLAST search to identify the polypeptide which you have been analyzing. (You may use either sequence.). Copy and paste the sequence into the “Enter Query Sequence box. Be sure that the box “Align two or more sequences” is **NOT** checked. Choose “Quick BLASTP” under the Program Selection. Click the “Blast” button at the bottom.

26. WHAT GENE ENCODES FOR THE POLYPEPTIDE YOU WERE ANALYZING?
Click on “Description”, then choose “Gene” under “Related Information” to answer the following two questions.
FGFR3

27. WHAT IS THE FUNCTION OF THIS PROTEIN? Click on “Description” and then “Gene” under “Related Information” to the right to get this information.
As a member of the fibroblast growth factor receptor (FGFR) family, this protein has an effect on mitogenesis and differentiation due to its extracellular interactions with fibroblast growth factors. This is not an immediate influence as it sets in motion downstream signals that have an ultimate effect on these systems. The functions specific to this protein include the ability to bind acidic and basic fibroblast growth hormone and effect the development and maintenance of bone.

28. WHAT HUMAN DISEASE IS CAUSED BY A MUTATION IN THIS GENE? Again, this information can be gained by clicking on “Description” and then “Gene” under “Related Information” to the right to get this information.
Craniosynostosis and different types of skeletal dysplasia.

Finally, but not required, if you click on the “AlphaFold Structure” under “Related Information” you will be given a structural diagram of this protein.

29. WHAT IS THE CONNECTION AMONG THE FOLLOWING: NIH, NLM, NCBI, and HHS?
All are focused on bettering human health through research and access to information. They make up databases that contain information related to biomedical sciences and biotechnology.

30. REFLECT ON ONE THING THAT YOU LEARNED FROM DOING THIS ASSIGNMENT.

I believe that I now have a better understanding of the complexity of a chromosome. The depth at which you can examine the genes on a chromosome is difficult to describe, so seeing it in this context gave me a better understanding of the amount of information a chromosome can hold. Being able to have a refresher on programs like this is incredibly beneficial and aids in furthering my understanding of genetics, especially regarding amino acid sequencing.

Works Cited

1. Schoch CL, et al. NCBI Taxonomy: a comprehensive update on curation, resources and tools. Database (Oxford). baaa062. PubMed: 32761142 PMC: PMC7408187 (2020).

2. Schoch CL, et al. NCBI Taxonomy: a comprehensive update on curation, resources and tools. Database (Oxford). baaa062. PubMed: 32761142 PMC: PMC7408187 (2020).

3. Higgins DG, Blackshields G, Wallace IM. Mind the gaps: progress in progressive alignment. *Proc Natl Acad Sci U S A*. **26;102(30)**,10411-2 (2005). doi: 10.1073/pnas.0504801102. EpPMID: 16027352; PMCID: PMC1180805.

4. National Institutes of Neurological Disorders and Stroke, Amyotrophic Lateral Sclerosis (ALS), https://www.ninds.nih.gov/health-information/disorders/amyotrophic-lateral-sclerosis-als (2023).

5. National Library of Medicine, GenBank Overview, https://www.ncbi.nlm.nih.gov/genbank/ (2023).