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?
Above the chromosome image you will see the number of genes and base pairs on that particular chromosome.

A) Chromosome 13

2 & 3. STATE THE NUMBER OF GENES **AND** BASE PAIRS ON THE CHROMOSOME YOU CHOSE.
Scan the chromosome map.

A) Contains approximately 800 genes

B) Contains over 110 million base pairs

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

A) RB1

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.

A) Rb is found in all cells of the body, where under normal conditions it acts as a brake on the cell division cycle by preventing certain regulatory proteins from triggering DNA replication.

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.

A) 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.

A) The encoded protein functions as a chloride channel, making it unique among members of this protein family, and controls ion and water secretion and absorption in epithelial tissues.

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.

A) Mutations in this gene cause 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.

A) 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.

A) Gorilla gorilla gorilla

11. WHAT IS THE COMMON NAME FOR THIS SPECIES?

A) lowland gorilla

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

A) It does not surprise me that this species also has 100% similarity because there is scientific evidence that humans evolved form apelike ancestors.

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?

A) Nomascus leucogenys

b. WHAT IS THE COMMON NAME?

A) **White-cheeked Gibbon**

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.

A) **0**

14. WHAT IS A GAP IN SEQUENCE ALIGNMENTS? (This is something you'll have to search for on the Internet.)

A) **A gap in sequence alignments is the absence of a region in one sequence that is present in another.**

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

A) **Mus musculus TAR DNA binding protein (Tardbp), transcript variant 1, mRNA**

16. NM_013444

A) **Homo sapiens ubiquilin 2 (UBQLN2), mRNA**

17. NM_001010850

A) **Homo sapiens fusion (involved in t(12;16) in malignant liposarcoma) (FUS), transcript variant 2, mRNA**

18. KJ174530

A) **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…)

A) **amyotrophic lateral sclerosis (ALS)**
B) **Lou Gehrig's disease**

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.)

A) **GenBank ® is the NIH genetic sequence database, an annotated collection of all publicly available DNA sequences**

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.)

A) **cDNA (short for copy DNA; also called complementary DNA) is synthetic DNA that has been transcribed from a specific mRNA through a reaction using the enzyme reverse transcriptase. cDNA only contains coding regions.**

Enter the following cDNA sequence on "tblastx":
ACATTTGCTTCTGACACAATTGTGTTCACTAGCAACCTCAAACAGACACCATGGTGCATCTGACTC
CTGAGGAGAAGTCTGCCGTTACTGCCCTGTGGGGCAAGGTGAACGTGGATGAAGTTGGTGGTGAG
GCCCTGGGCAG
22. WHAT IS THE SEQUENCE MATCH?

A) **beta-globin [Homo sapiens] Sequence ID: AAA88069.1**

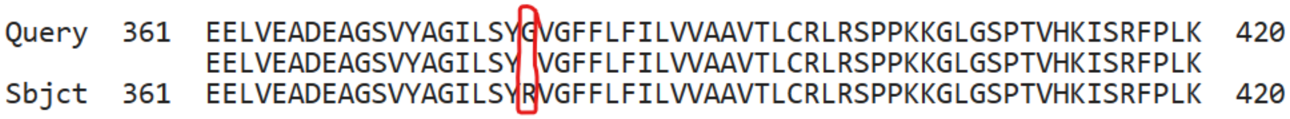
Assignment: While still on BLAST, choose "Protein BLAST". Check the box which says "Align two or more sequences".
Copy and paste the following sequence into the "Enter Query Sequence" box:
Person 1/Sequence 1:
MGAPACALALCVAVAIVAGASSESLGTEQRVVGRAAEVPGPEPGQQEQLVFGSGDAVELSCPP
PGGGPMGPTVWVKDGTGLVPSERVLVGPQRLQVLNASHEDSGAYSCRQRLTQRVLCHFSVRV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

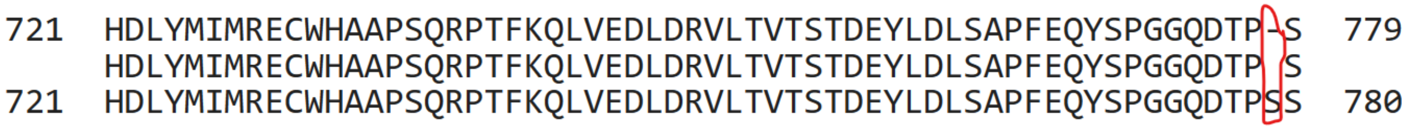
Click on the "BLAST" box.
After you receive results, choose the "Description" box for more information about the alignment.
23. DO YOU SEE ANY DIFFERENCES BETWEEN THE TWO AMINO ACID SEQUENCES? (Look for a space between same amino acid comparison for both sequences.)

A) **yes**

24. IF YOU SAW DIFFERENCES, WHAT WERE THEY?



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



```
721  HDLYMIMRECWHAAPSQRPTFKQLVEDLDRVLTVTSTDEYLDLSAPFEQYSPGGQDTP-S  779
     HDLYMIMRECWHAAPSQRPTFKQLVEDLDRVLTVTSTDEYLDLSAPFEQYSPGGQDTP S
721  HDLYMIMRECWHAAPSQRPTFKQLVEDLDRVLTVTSTDEYLDLSAPFEQYSPGGQDTPSS  780
```

25. ARE THERE ANY GAPS IN THE SEQUENCE ALIGNMENT?

A) yes

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.

A) **fibroblast growth factor receptor 3**

27. WHAT IS THE FUNCTION OF THIS PROTEIN? Click on "Description" and then "Gene" under "Related Information" to the right to get this information.

A) **This gene encodes a member of the fibroblast growth factor receptor (FGFR) family, with its amino acid sequence being highly conserved between members and among divergent species. FGFR family members differ from one another in their ligand affinities and tissue distribution. A full-length representative protein would consist of an extracellular region, composed of three immunoglobulin-like domains, a single hydrophobic membrane-spanning segment and a cytoplasmic tyrosine kinase domain. The extracellular portion of the protein interacts with fibroblast growth factors, setting in motion a cascade of downstream signals, ultimately influencing mitogenesis and differentiation. This particular family member binds acidic and basic fibroblast growth hormone and plays a role in bone development and maintenance.**

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

A) **Mutations in this gene lead to craniosynostosis and multiple types of skeletal dysplasia.**

29. WHAT IS THE CONNECTION AMONG THE FOLLOWING: NIH, NLM, NCBI, and HHS?

A) **The National Center for Biotechnology Information (NCBI) is part of the United States National Library of Medicine (NLM), a branch of the National Institutes of Health (NIH). It is funded by the U.S. Department of Health and Human Services (HHS).**

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

A) One thing that I learned from this assignment was that the GenBank is comprised of international data from Europe, Japan, and the U.S.