Old Dominion University/BIOL294-Genetics/Rinehart-Kim/Module 9/Genome Assignment/SP24

- 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 just your answers on another word document.
- The following is an excellent reference: Genetics Home Reference: http://ghr.nlm.nih.gov
- As of now (3/31), 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? 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.

## 400 genes and 40 million base pairs

Scan the chromosome map.

4. LIST ONE GENE THAT IS LOCATED ON THIS CHROMOSOME. SOD1

5. STATE THE FUNCTION OF THE GENE YOU LISTED IN #4. This is possible by clicking on the gene you stated in #4. You must state the physiological function of the gene product you select, not the disease listed for the gene. Codes for "an antioxidant enzyme protecting the cell from reactive oxygen species toxicity" <u>https://pubmed.ncbi.nlm.nih.gov/25492944/</u>

### 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 <u>http://blast.ncbi.nlm.nih.gov/Blast.cgi</u> 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. Encodes a protein that 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. **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

12. DOES IT SURPRISE YOU THAT THIS SPECIES ALSO HAS A 100% SIMILARITY IN IDENTITY? WHY OR WHY NOT? No, humans are closest to apes according to our evolutionary lineage.

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 apella

b. WHAT IS THE COMMON NAME? tufted capuchin

c. ARE THERE ANY GAPS 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. Yes, (1/119, or 0%)

14. WHAT IS A GAP IN SEQUENCE ALIGNMENTS? (This is something you'll have to search for on the Internet.) Gaps in sequence alignments happen when accounting for insertions or deletions when comparing two sequences of DNA. For example, a gap is placed on the sequence where a deletion occurred and a gap is placed on the other sequence where an insertion occurred on the main sequence to make up for the extra space. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1180805/

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 (<u>www.ncbi.nlm.nih.gov/BLAST.cgi</u>) 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 FUS RNA binding protein (FUS), transcript variant 2, non-coding RNA

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 is just beginning...) Amyotrophic lateral sclerosis, or 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.) It is a database of all the DNA sequences that are available to the public. <u>https://www.ncbi.nlm.nih.gov/genbank/</u>

# **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 <u>http://blast.ncbi.nlm.nih.gov/Blast.cgi</u>

Click on "blastx". (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.) It is short for "copy DNA" and it is synthetic DNA that is used in gene cloning or other genetic experiments. <u>https://www.genome.gov/genetics-glossary/Copy-DNA</u>

Enter the following cDNA sequence on "blastx":

ACATTTGCTTCTGACACAATTGTGTTCACTAGCAACCTCAAACAGACACCATGGTGCATCTGACTC CTGAGGAGAAGTCTGCCGTTACTGCCCTGTGGGGGCAAGGTGAACGTGGATGAAGTTGGTGGTGAG 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:

MGÅPACALALCVAVAIVAGASSESLGTEQRVVGRAAEVPGPEPGQQEQLVFGSGDAVELSCPP PGGGPMGPTVWVKDGTGLVPSERVLVGPQRLQVLNASHEDSGAYSCRQRLTQRVLCHFSVRV TDAPSSGDDEDGEDEAEDTGVDTGAPYWTRPERMDKKLLAVPAANTVRFRCPAAGNPTPSIS WLKNGREFRGEHRIGGIKLRHQQWSLVMESVVPSDRGNYTCVVENKFGSIRQTYTLDVLERSP HRPILQAGLPANQTAVLGSDVEFHCKVYSDAQPHIQWLKHVEVNGSKVGPDGTPYVTVLKTA GANTTDKELEVLSLHNVTFEDAGEYTCLAGNSIGFSHHSAWLVVLPAEEELVEADEAGSVYAG ILSYRVGFFLFILVVAAVTLCRLRSPPKKGLGSPTVHKISRFPLKRQVSLESNASMSSNTPLVRIA RLSSGEGPTLANVSELELPADPKWELSRARLTLGKPLGEGCFGQVVMAEAIGIDKDRAAKPVT VAVKMLKDDATDKDLSDLVSEMEMMKMIGKHKNIINLLGACTQGGPLYVLVEYAAKGNLRE FLRARRPPGLDYSFDTCKPPEEQLTFKDLVSCAYQVARGMEYLASQKCIHRDLAARNVLVTED NVMKIADFGLARDVHNLDYYKKTTNGRLPVKWMAPEALFDRVYTHQSDVWSFGVLLWEIFT LGGSPYPGIPVEELFKLLKEGHRMDKPANCTHDLYMIMRECWHAAPSQRPTFKQLVEDLDRVL TVTSTDEYLDLSAPFEQYSPGGQDTPSSSSSGDDSVFAHDLLPPAPPSSGGSRT

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.) yes, near the end of the sequence

24. IF YOU SAW DIFFERENCES, WHAT WERE THEY? There was an extra S insertion near the end of the second sequence

25. ARE THERE ANY GAPS IN THE SEQUENCE ALIGNMENT? Yes, the first sequence was missing an S near the end

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? **fibroblast growth** 

### factor receptor 3 isoform 1 precursor [Homo sapiens]

Click on "Description", then choose "Gene" under "Related Information" to answer the following two questions. 27. WHAT IS THE FUNCTION OF THIS PROTEIN? Click on "Description" and then "Gene" under "Related Information" to the right to get this information. "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 **Achondroplasia as well as several other disorders** 

29. WHAT IS THE CONNECTION AMONG THE FOLLOWING: NIH, NLM, NCBI, and HHS? (What do the abbreviations stand for? Who oversees what?) NIH is the National Institute of Health, NLM is the National Library of Medicine, NCBI is the National Center for Biotechnology Information, and HHS is the Department of Health and Human Services. The NIH is run by the HHS, and the NLM and NCBI are part of the NIH.

30. WHAT WAS ONE POSITIVE THING AND ONE NEGATIVE THING YOU ENCOUNTERED WHILE DOING THIS ASSIGNMENT? I liked finding out what species was the most genetically similar to us. I did not like how slow it was to use the BLAST system and how it was slightly confusing to navigate the website.