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 17**

2 & 3. STATE THE NUMBER OF GENES **AND** BASE PAIRS ON THE CHROMOSOME YOU CHOSE.
   - **1600 Genes and 80 million base pairs**

4. LIST ONE GENE THAT IS LOCATED ON THIS CHROMOSOME.
   - **p53**

5. STATE THE FUNCTION OF THE GENE YOU LISTED IN #4.
   - **A tumor repressor protein. Its function is to regulate cell division and stop cell from proliferating too fast or in a way that is uncontrollable. (It is important that you state the physiological function of the gene product you select, not the disease listed for the gene.)**

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 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.
   - **The encoded protein functions as a chloride channel by inscribing a member of the ATP-binding cassette (ABC). It makes it unique among the members of the protein family. It controls ion and water secretion as well as 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.
   - **The disease mutated from this gene is the cause of 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.
   - **located in 7q31.2**

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.

- **Pan Troglodytes CR CFTR, mRNA**

11. WHAT IS THE COMMON NAME FOR THIS SPECIES?

- **Chimpanzee**

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

- **Not quite because chimpanzees and human have similarities to one another so for pan troglodytes to have 100% similarity in identity is not as surprising as if a lion or fruit fly would do so.**

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?

- **Genus: Robust Capuchin Monkeys**
- **Species: 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, but only 1/119 which equals to 0%.**

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

- **It is the maximum sequence of consecutive insertions or deletions of letters. It can be described as the absence of a region that is present in another sequence.**

(Barton C., Flouri T., Iliopoulos S.C., Pissis P.S., Global and local sequence alignment with a bounded number of gaps, *Theoretical Computer Science*, Volume 582, 2015, pg.1-16, https://doi.org/10.1016/j.tcs.2015.03.016.)

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, describe 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 ubiquitin 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 is just beginning…)

- **ALS, Amyotrophic Lateral Sclerosis. Its common name is "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 comprehensive database that contains nucleotide sequences for 300,000 or more organisms named at the genus level. It is obtained primarily through submissions of individual laboratories and batch submissions of large-scale sequencing.**

(Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., & Sayers, E. W. (2010). GenBank. *Nucleic acids research*, 38, D46–D51. https://doi.org/10.1093/nar/gkp1024)

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) search for similarities between amino acid sequences.
Assignment: Access the BLAST site at http://blast.ncbi.nlm.nih.gov/Blast.cgi
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. It a DNA that has been transcribed from a specific mRNA using the enzyme reverse transcriptase. It is often use by scientists as a toolfor gene cloning.**

(Shchelochkov, A.O., (2024). National Human Genome Research Institute, *CDNA (Copy DNA)*. https://www.genome.gov)

Enter the following cDNA sequence on "blastx":
ACATTTGCTTCTGACACAATTGTGTTCACTAGCAACCTCAAACAGACACCATGGTGCATCTGACTCC
TGAGGAGAAGTCTGCCGTTACTGCCCTGTGGGGCAAGGTGAACGTGGATGAAGTTGGTGGTGAG
GCCCTGGGCAG

22. WHAT IS THE SEQUENCE MATCH?

- **Beta-globin**

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:

MGAPACALALCVAVAIVAGASSESLGTEQRVVGRAAEVPGPEPGQQEQLVFGSGDAVELSCPPP
GGGPMGPTVWVKDGTGLVPSERVLVGPQRLQVLNASHEDSGAYSCRQRLTQRVLCHFSVRVTD
APSSGDDEDGEDEAEDTGVDTGAPYWTRPERMDKKLLAVPAANTVRFRCPAAGNPTPSISWLK
NGREFRGEHRIGGIKLRHQQWSLVMESVVPSDRGNYTCVVENKFGSIRQTYTLDVLERSPHRPI
LQAGLPANQTAVLGSDVEFHCKVYSDAQPHIQWLKHVEVNGSKVGPDGTPYVTVLKTAGANT
TDKELEVLSLHNVTFEDAGEYTCLAGNSIGFSHHSAWLVVLPAEEELVEADEAGSVYAGILSYG
VGFFLFILVVAAVTLCRLRSPPKKGLGSPTVHKISRFPLKRQVSLESNASMSSNTPLVRIARLSSGE
GPTLANVSELELPADPKWELSRARLTLGKPLGEGCFGQVVMAEAIGIDKDRAAKPVTVAVKML
KDDATDKDLSDLVSEMEMMKMIGKHKNIINLLGACTQGGPLYVLVEYAAKGNLREFLRARRPP
GLDYSFDTCKPPEEQLTFKDLVSCAYQVARGMEYLASQKCIHRDLAARNVLVTEDNVMKIADF
GLARDVHNLDYYKKTTNGRLPVKWMAPEALFDRVYTHQSDVWSFGVLLWEIFTLGGSPYPGIP
VEELFKLLKEGHRMDKPANCTHDLYMIMRECWHAAPSQRPTFKQLVEDLDRVLTVTSTDEYLD
LSAPFEQYSPGGQDTPSSSSGDDSVFAHDLLPPAPPSSGGSRT

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

MGAPACALALCVAVAIVAGASSESLGTEQRVVGRAAEVPGPEPGQQEQLVFGSGDAVELSCPPP
GGGPMGPTVWVKDGTGLVPSERVLVGPQRLQVLNASHEDSGAYSCRQRLTQRVLCHFSVRVTD
APSSGDDEDGEDEAEDTGVDTGAPYWTRPERMDKKLLAVPAANTVRFRCPAAGNPTPSISWLK
NGREFRGEHRIGGIKLRHQQWSLVMESVVPSDRGNYTCVVENKFGSIRQTYTLDVLERSPHRPI
LQAGLPANQTAVLGSDVEFHCKVYSDAQPHIQWLKHVEVNGSKVGPDGTPYVTVLKTAGANT
TDKELEVLSLHNVTFEDAGEYTCLAGNSIGFSHHSAWLVVLPAEEELVEADEAGSVYAGILSYR
VGFFLFILVVAAVTLCRLRSPPKKGLGSPTVHKISRFPLKRQVSLESNASMSSNTPLVRIARLSSGE
GPTLANVSELELPADPKWELSRARLTLGKPLGEGCFGQVVMAEAIGIDKDRAAKPVTVAVKML
KDDATDKDLSDLVSEMEMMKMIGKHKNIINLL
GACTQGGPLYVLVEYAAKGNLREFLRARRPPGLDYSFDTCKPPEEQLTFKDLVSCAYQVARGME
YLASQKCIHRDLAARNVLVTEDNVMKIADFGLARDVHNLDYYKKTTNGRLPVKWMAPEALF
DRVYTHQSDVWSFGVLLWEIFTLGGSPYPGIPVEELFKLLKEGHRMDKPANCTHDLYMIMREC
WHAAPSQRPTFKQLVEDLDRVLTVTSTDEYLDLSAPFEQYSPGGQDTPSSSSGDDSVFAHDLLP
PAPPSSGGSRT

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**

24. IF YOU SAW DIFFERENCES, WHAT WERE THEY?

- **It was between the Query 361 and Subject 361, wherein they both were continuously matching from the 1st sequence letter to the 19th sequence letter (EELVEADEAGSVYAGILSY…) then a space appeared between the columns of G and R. It was the 20th sequence letter.**

25. ARE THERE ANY GAPS IN THE SEQUENCE ALIGNMENT?

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.

- **Yes, the gaps between the sequence alignment is 1%.**

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.

- **It encodes for the production of the FGFR3 protein that converts cartilage to bone.**

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

- **The function of this protein is to enable ATP binding, enable fibroblast growth factor binding, and enable fibroblast growth facter receptor activity.**

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

- **The most common form of mutation caused by fibroblast growth factor receptie 3 is “*achondroplasia*”, which is a common form of bony dysplasia.**

29. WHAT IS THE CONNECTION AMONG THE FOLLOWING: NIH, NLM, NCBI, and HHS? (What do the abbreviations stand for? Who oversees what?)

- **NLM (National Library of Medicine), NCBI (National Center for Biotechnology Information), NIH (National Institutes of Health), and HHS (Health and Human Services). The NLM and NCBI distribute the results of biomedical research that is funded and carried out by the NIH.**

30. WHAT WAS ONE POSITIVE THING AND ONE NEGATIVE THING YOU ENCOUNTERED WHILE DOING THIS ASSIGNMENT?

- **The negative thing that I have encountere doing this assignment is having to identity the sequence gaps and protein on the last 6 questions. It took me longer than the rest. The positive thing that I have encountered as I did the assignment is knowing what to look for and actually applying what I have been taught throughout the course semester.**