- 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.

Name: Tyler Rector
Date: 4/22/2024

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.

I choose Chromosome 3

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

The number of genes was 1900 and the number of base pairs is 200 million.

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

MLH1 which is linked with colon cancer.

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.**

The MLH1 creates protein products that correct for mistakes during DNA replication. If they are mutated then damage can accumulate in the DNA and cause colon cancer. (Source)

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 ggacccagc

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.

The gene encodes an ATP-binding cassette transporter protein which acts as a chloride channel and controls water secretion and absorption in epithelial tissue.

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.

When mutated these genes can 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.

It is found on 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(Pongo abelii)

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

No because this species shares a recent, on evolutionary time scales, common ancestor and shares many similarities with homo sapiens.

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.

Nomascus leucogenys CF transmembrane conductance regulator (CFTR), mRNA

a. WHAT IS THE GENUS AND SPECIES WITH THIS NUCLEOTIDE SEQUENCE?

Nomascus leucogenys
b. WHAT IS THE COMMON NAME?

northern white-cheeked gibbon

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.

There are 0/120 or 0% gaps between the two sequences.

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

According to the paper "Global and local sequence alignment with a bounded number of gaps" a gap in alignment is the absence or presence of a region that is not in the other compared sequence. (Source)

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 is just beginning…)

Lou Gehrig disease or ALS(amyotrophic lateral sclerosis)

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 the NIH's open access genetic sequence and protein translation database.

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 "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.)

cDNA is copy DNA which is synthetic DNA transcribed from mRNA through the enzyme reverse transcriptase.

Enter the following cDNA sequence on "blastx":
ACATTTGCTTCTGACACAATTGTGTTCACTAGCAACCTCAAACAGACACCATGGTGCATCTGACTCC
TGAGGAGAAGTCTGCCGTTACTGCCCTGTGGGGCAAGGTGAACGTGGATGAAGTTGGTGGTGAG
GCCCTGGGCAG
22. WHAT IS THE SEQUENCE MATCH?

The sequence's closest match is to beta-globin which is a component of the larger protein hemoglobin in 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:

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

WHAAPSQRPTFKQLVEDLDRVLTVTSTDEYLDLSAPFEQYSPGGQDTPSSSSSGDDSVFAHDLLP
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.)

There was one difference between the two amino acids.

24. IF YOU SAW DIFFERENCES, WHAT WERE THEY?

"AGILSYGVGFFLFIL"
"AGILSYRVGFFLFIL"
There was a single difference between the two amino acids with a change from a "G" to "R" on 2nd Amino acid.

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.

There is 1/806 gaps in the sequence.

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 or fibroblast growth factor receptor 3 [Homo sapiens (human)]

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 protein binds to fibroblast growth hormones to develop and maintain bones.

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 Skeletal Dysplasia are caused by mutations in this gene.

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
HHS is the Department of Health and Human Services
The Department of Health and Human Services manages the NIH as an agency. The NIH manages the NLM and NCBI as agency initiatives.

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

Positive: I got to learn how to use these tools.

Negative: The tools are extremely slow and delayed the assignment significantly.