1). Chromosome Y

2 & 3) Contains over 200 genes and over 50 million base pairs

4.) SRY Testes-determining factor

5) Binds to other DNA and in doing so distorts it, altering the properties causing the formation of testes

6) *Homo sapiens* CFTR promoter region

7) Part of the gene associated with gene details

8) "The CFTR gene promoter is clearly important for maintaining levels of CFTR gene expression, but apparently it does not contain any tissue-specific elements.

McCarthy, V.A. & Harris, A. The CFTR gene-and regulation of its expression. *Pediatric Pulmonology* 40, 1-8 (2005).

9) Cystic Fibrosis

10) Chromosome 7

11) *Pongo pygmaeous abelii*

12) Sumatran Orangutan

13) No, many of the great apes share similarity between genomic codes and attributes

14) *Pan troglodyles* – Common Chimpanzee

15) 1/119

16) "A gap in one of the sequences means that one or more amino acids residues have been deleted from the sequence, or we could also say there is an insertion in the second sequence."

Alignment scoring, Gaps & Similarities. *Sequence Alignment: Scores, Gaps and Gap Penalties* Available at: https://proteinstructures.com/sequence/sequence-alignment/#:~:text=A%20gap%20in%20one%20of,insertion%20in%20the%20second%20sequence. (Accessed: 6th December 2022)

17) TAR DNA Binding Protein – important segment for protein stability

18) Ubiquilinz, RNA – Deals with health of stem cell derived motor neurons

19) Induces mitochondrial dysfunction and cytosolic aggregates

20) mutation in Downs Syndrome – Superoxide dismutase

21) Trisomy 21 aka Downs Syndrome

22) The NIH genetic sequence database, on annotated collection of all publicly available DNA sequences.

23) Synthetic DNA that has been transcripted from a specific mRNA through a reaction using enzyme reverse transcriptase.

Shchelochkov, O. A. CDNA (copy DNA). *Genome.gov* (2022). Available at: https://www.genome.gov/genetics-glossary/Copy-DNA. (Accessed: 6th December 2022)

24) *Homo sapiens* partial HBB gene for hemoglobin betachain, exon 1, isolate 04593664

25) Spans of DNA sequence between the start and stop codons.

26) M acts as the beginning of the one of the open reading frames

27) Frame 3 3'5' – least # of conflicting reds that is not highlighted

28 & 29) G → R 361 v. 361 - → S 381 v. 381

30) Fibroblast growth factor – development at tissue

31) Pfeifer syndrome

Sargar, K. M., Singh, A. K. & Kao, S. C. Imaging of skeletal disorders caused by fibroblast growth factor receptor gene mutations. *RadioGraphics* **37,** 1813–1830 (2017).

32) I learned that there was a very intriguing website to locate a genomic code's origin.

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

- You are to address the statements/questions that are in red. You do not need to use complete sentences in your replies.
- The following is an excellent reference: https://medlineplus.gov/genetics/
- **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. WHICH CHROMOSOME DID YOU CHOOSE?

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.

Scan the chromosome map.

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

5. STATE THE NORMAL 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 NORMAL physiological function of the gene product you select.

Introduction to 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 TOP SEQUENCE DESCRIPTION MATCH FOR YOUR QUERY SEQUENCE? For this answer, you should give the description listed. Do not choose a Predicted sequence.
7. IS THIS A SEQUENCE FOR A PROTEIN OR ANOTHER PART OF THE GENE? IF IT IS "ANOTHER PART OF THE GENE", EXPLAIN ITS PURPOSE.
8.. WHAT DOES THE ENCODED PROTEIN ASSOCIATED WITH THE ABOVE SEQUENCE DO IN THE BODY? Search the PubMed site at www.ncbi.nlm.nih.gov/entrez/query.fcgi?DB=pubmed to answer this question. Under "Article types," choose "Review". CITE THE PAPER YOU USED TO DETERMINE THE PURPOSE OF THE ENCODED PROTEIN.
Click on the top match to find the following.
9. A MUTATED FORM OF THIS GENE IS RESPONSIBLE FOR A WELL-STUDIED DISEASE. WHAT IS THAT DISEASE? You should be able to get this information from the description of the gene. You may need to "probe" the gene description.
10. ON WHAT CHROMOSOME IS THE GENE LOCATED? You should be able to get this information by clicking on the description of the gene.
11. Return to the original nucleotide sequence alignment descriptions. WHAT SPECIES (STATE THE SCIENTIFIC NAME) OTHER THAN *HOMO SAPIENS* ALSO HAS A 100% IDENTITY (Ident) FOR THIS SEQUENCE? USE THE TOP SEQUENCE LISTED, BUT DO NOT USE THE PREDICTED SEQUENCES.
12. WHAT IS THE COMMON NAME FOR THIS SPECIES?
13. DOES IT SURPRISE YOU THAT THIS SPECIES ALSO HAS A 100% SIMILARITY IN IDENTITY? WHY OR WHY NOT?
14. DESCRIBE THE FIRST MATCH THAT HAS LESS THAN 100% QUERY COVER BUT IS NOT PREDICTED OR *HOMO SAPIENS*. STATE THE SCIENTIFIC AND COMMON NAMES.
15. Click on the description to answer this question. HOW MANY GAPS OCCUR BETWEEN THE TWO SEQUENCES (THE ONE YOU SUBMITTED AND THE FIRST ONE THAT HAS LESS THAN 100% QUERY COVER)?
16. WHAT IS A GAP IN SEQUENCE ALIGNMENTS? (This is something you'll have to search for elsewhere.)
You can also do BLAST searches using an accession number that has been assigned to a particular sequence when it has been 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 (#16-19). Give the description of the gene and gene product. You do not need to state the organism source.
17. NM_145556
18. NM_013444
19. NM_001010850
20. KJ174530
21. Search Google to answer the following: WHAT DISEASE IS ASSOCIATED WITH MUTATIONS OF THE GENES REFERENCED IN #17-#20? WHAT IS A "COMMON NAME" OF THE DISEASE? (The name of a person; Hint, hint…We just finished the World Series…)
BLAST is possible because of the submission of DNA sequences to GenBank.
22. WHAT IS GENBANK? (You can do an Internet search to find this information.)

Introduction to Swiss-Prot to Study Protein Sequences

Assignment Goal: To use the Internet-based site ExPASy (Expert Protein Analysis System) to translate cDNA, and the Internet-based database UniProt KB/Swiss-Prot to access a complete polypeptide.
23. WHAT IS cDNA? How can we obtain cDNA in the lab?

Assignment: Access the BLAST site at www.ncbi.nlm.nih.gov/BLAST.cgi
Click on "Nucleotide Blast"
Enter the following sequence:
ACATTTGCTTCTGACACAATTGTGTTCACTAGCAACCTCAAACAGACACCATGGTGCATCTGACTC
CTGAGGAGAAGTCTGCCGTTACTGCCCTGTGGGGCAAGGTGAACGTGGATGAAGTTGGTGGTGAG
GCCCTGGGCAG
24. USING THE SAME PROGRAM YOU USED IN THE INTRODUCTION TO BLAST ABOVE, WHAT IS THE SEQUENCE MATCH?
Now access the Expasy translate tool at https://web.expasy.org/translate/.
Enter the above DNA sequence.
Click "Translate Sequence".
25. WHAT IS AN OPEN READING FRAME?
26. ALL OF THE PROPOSED OPEN READING FRAMES (HIGHLIGHTED IN RED) START WITH THE AMINO ACID "M". FROM WHAT YOU KNOW ABOUT POLYPEPTIDES, WHAT IS "M"?
27. WHICH 5' TO 3' FRAME IS MOST LIKELY TO BE AN OPEN READING FRAME? WHY DID YOU CHOOSE THAT FRAME?

Amino Acid Sequence Comparisons

Assignment Goal: To use the Internet-based site Expasy SIM program to align two amino acid sequences. Knowing the sources of these sequences will allow one to determine the mutation and potential cause of a human disease.

Assignment: Access the Expasy site at https://web.expasy.org/sim/. Copy and paste each of the following sequences into the "Sequence" text boxes as User-entered sequence.

Person 1/Sequence 1:
MGAPACALALCVAVAIVAGASSESLGTEQRVVGRAAEVPGPEPGQQEQLVFGSGDAVELSCPPPGGGP
MGPTVWVKDGTGLVPSERVLVGPQRLQVLNASHEDSGAYSCRQRLTQRVLCHFSVRVTDAPSSGDDE
DGEDEAEDTGVDTGAPYWTRPERMDKKLLAVPAANTVRFRCPAAGNPTPSISWLKNGREFRGEHRIG
GIKLRHQQWSLVMESVVPSDRGNYTCVVENKFGSIRQTYTLDVLERSPHRPILQAGLPANQTAVLGSD
VEFHCKVYSDAQPHIQWLKHVEVNGSKVGPDGTPYVTVLKTAGANTTDKELEVLSLHNVTFEDAGEY
TCLAGNSIGFSHHSAWLVVLPAEEELVEADEAGSVYAGILSYGVGFFLFILVVAAVTLCRLRSPPKKGL
GSPTVHKISRFPLKRQVSLESNASMSSNTPLVRIARLSSGEGPTLANVSELELPADPKWELSRARLTLGK
PLGEGCFGQVVMAEAIGIDKDRAAKPVTVAVKMLKDDATDKDLSDLVSEMEMMKMIGKHKNIINLL
GACTQGGPLYVLVEYAAKGNLREFLRARRPPGLDYSFDTCKPPEEQLTFKDLVSCAYQVARGMEYLA
SQKCIHRDLAARNVLVTEDNVMKIADFGLARDVHNLDYYKKTTNGRLPVKWMAPEALFDRVYTHQS
DVWSFGVLLWEIFTLGGSPYPGIPVEELFKLLKEGHRMDKPANCTHDLYMIMRECWHAAPSQRPTFK
QLVEDLDRVLTVTSTDEYLDLSAPFEQYSPGGQDTPSSSSGDDSVFAHDLLPPAPPSSGGSRT

Person 2/Sequence 2:
MGAPACALALCVAVAIVAGASSESLGTEQRVVGRAAEVPGPEPGQQEQLVFGSGDAVELSCPPPGGGP
MGPTVWVKDGTGLVPSERVLVGPQRLQVLNASHEDSGAYSCRQRLTQRVLCHFSVRVTDAPSSGDDE
DGEDEAEDTGVDTGAPYWTRPERMDKKLLAVPAANTVRFRCPAAGNPTPSISWLKNGREFRGEHRIG
GIKLRHQQWSLVMESVVPSDRGNYTCVVENKFGSIRQTYTLDVLERSPHRPILQAGLPANQTAVLGSD
VEFHCKVYSDAQPHIQWLKHVEVNGSKVGPDGTPYVTVLKTAGANTTDKELEVLSLHNVTFEDAGEY
TCLAGNSIGFSHHSAWLVVLPAEEELVEADEAGSVYAGILSYRVGFFLFILVVAAVTLCRLRSPPKKGL
GSPTVHKISRFPLKRQVSLESNASMSSNTPLVRIARLSSGEGPTLANVSELELPADPKWELSRARLTLGK
PLGEGCFGQVVMAEAIGIDKDRAAKPVTVAVKMLKDDATDKDLSDLVSEMEMMKMIGKHKNIINLL
GACTQGGPLYVLVEYAAKGNLREFLRARRPPGLDYSFDTCKPPEEQLTFKDLVSCAYQVARGMEYLA
SQKCIHRDLAARNVLVTEDNVMKIADFGLARDVHNLDYYKKTTNGRLPVKWMAPEALFDRVYTHQS
DVWSFGVLLWEIFTLGGSPYPGIPVEELFKLLKEGHRMDKPANCTHDLYMIMRECWHAAPSQRPTFK
QLVEDLDRVLTVTSTDEYLDLSAPFEQYSPGGQDTPSSSSGDDSVFAHDLLPPAPPSSGGSRT

Submit the sequences for comparison.

28. DO YOU SEE ANY DIFFERENCES BETWEEN THE TWO AMINO ACID SEQUENCES? (Look for the absence of an asterisk, which indicates the same amino acid in both sequences.)
29. IF YOU SAW DIFFERENCES, WHAT WERE THEY?

Return to the BLAST home page (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Run a **PROTEIN** BLAST search to identify the polypeptide which you have been analyzing. (You may use either sequence.)

30. WHAT IS THE FUNCTION OF THIS PROTEIN?
31. WHAT HUMAN DISEASE IS CAUSED BY A MUTATION IN THIS GENE?

32. REFLECT ON ONE THING THAT YOU LEARNED FROM DOING THIS ASSIGNMENT. Please be honest. If you didn't learn anything, admit it…