Genome Assignment   
Kristin May

1. WHICH CHROMOSOME DID YOU CHOOSE?

Chromosome 3

1. STATE THE NUMBER OF GENES

1900 Genes

1. BASE PAIRS ON THE CHROMOSOME YOU CHOSE

200 million base pairs

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

MLH1 Gene

1. STATE THE NORMAL FUNCTION OF THE GENE YOU LISTED IN #4.

Provides instructions for making a protein that plays an essential role in repairing DNA

1. WHAT IS THE TOP SEQUENCE DESCRIPTION MATCH FOR YOUR QUERY SEQUENCE?

Homo sapiens CFTR promoter region (LOC111674463) on chromosome 7

1. IS THIS A SEQUENCE FOR A PROTEIN OR ANOTHER PART OF THE GENE? IF IT IS “ANOTHER PART OF THE GENE”, EXPLAIN ITS PURPOSE.

provides instructions for making a protein called the CF transmembrane conductance regulator (CFTR).

1. WHAT DOES THE ENCODED PROTEIN ASSOCIATED WITH THE ABOVE SEQUENCE DO IN THE BODY? This protein functions as a channel across the membrane of cells that produce mucus, sweat, saliva, tears, and digestive enzymes. The channel transports negatively charged particles called chloride ions into and out of cells. The transport of chloride ions helps control the movement of water in tissues, which is necessary for the production of thin, freely flowing mucus
2. A MUTATED FORM OF THIS GENE IS RESPONSIBLE FOR A WELL-STUDIED DISEASE. WHAT IS THAT DISEASE? Cystic Fibrosis results from mutation of the cystic fibrosis transmembrane conductance regulator (CFTR) gene.
3. ON WHAT CHROMOSOME IS THE GENE LOCATED?   
   Chromosome 7
4. 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.  
   Gorillas
5. WHAT IS THE COMMON NAME FOR THIS SPECIES?

Gorilla

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

No because they’re primates and we’re closely related to them

1. DESCRIBE THE FIRST MATCH THAT HAS LESS THAN 100% QUERY COVER BUT IS NOT PREDICTED OR HOMO SAPIENS. STATE THE SCIENTIFIC AND COMMON NAMES.

- Saimiri boliviensis boliviensis, black capped squirrel monkey

1. HOW MANY GAPS OCCUR BETWEEN THE TWO SEQUENCES (THE ONE YOU SUBMITTED AND THE FIRST ONE THAT HAS LESS THAN 100% QUERY COVER)?   
   1/119
2. WHAT IS A GAP IN SEQUENCE ALIGNMENTS?   
   one or more amino acid residues have been deleted from the sequence, or we could also say that there is an insertion in the second sequence.
3. NM\_145556: TAR DNA binding protein (Tardbp)
4. NM\_013444: ubiquilin 2 (UBQLN2)
5. NM\_001010850: FUS RNA binding protein (FUS)
6. KJ174530: superoxide dismutase-1 (SOD-1) gene
7. WHAT DISEASE IS ASSOCIATED WITH MUTATIONS OF THE GENES REFERENCED IN #17-#20? WHAT IS A “COMMON NAME” OF THE DISEASE?

Amyotrophic lateral sclerosis, common name is Lou Gehrig’s Disease

1. WHAT IS GENBANK?   
   The GenBank sequence database is an open access, annotated collection of all publicly available nucleotide sequences and their protein translations.
2. WHAT IS cDNA? How can we obtain cDNA in the lab?  
   -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.   
   - ProtoScript® II Reverse Transcriptase is a recombinant M-MuLV reverse transcriptase with reduced RNase H activity and increased thermostability. It can be used to synthesize first strand cDNA at higher temperatures than the wild-type M-MuLV.
3. USING THE SAME PROGRAM YOU USED IN THE INTRODUCTION TO BLAST ABOVE, WHAT IS THE SEQUENCE MATCH?  
   - partial HBB gene for hemoglobin beta chain, exon 1
4. WHAT IS AN OPEN READING FRAME?

An open reading frame is a portion of a DNA molecule that, when translated into amino acids, contains no stop codons.

1. 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”?

Methionine

1. WHICH 5’ TO 3’ FRAME IS MOST LIKELY TO BE AN OPEN READING FRAME? WHY DID YOU CHOOSE THAT FRAME?

-frame 3 because it is the longest after the M which indicates the open reading frame

1. DO YOU SEE ANY DIFFERENCES BETWEEN THE TWO AMINO ACID SEQUENCES?

yes

1. IF YOU SAW DIFFERENCES, WHAT WERE THEY?

-361 there’s an R G swap   
-781 there’s an S on one and an absence on the other

1. WHAT IS THE FUNCTION OF THIS PROTEIN?

fibroblast growth factor receptor 3 isoform 1 precursor

1. WHAT HUMAN DISEASE IS CAUSED BY A MUTATION IN THIS GENE?

Achondroplasia.

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

-I learned a lot about the data bases available and how easy it is now to look up these types of things. I learned what cDNA was too, and more about mutations and how directly they can impact the diseases we get from our chromosomes.

References other than the ones provided

<https://medlineplus.gov/genetics/gene/mlh1/#:~:text=The%20MLH1%20gene%20provides%20instructions,in%20preparation%20for%20cell%20division>.

<https://medlineplus.gov/genetics/gene/cftr/>

<https://pubmed.ncbi.nlm.nih.gov/22201896/>

<https://medlineplus.gov/genetics/gene/tardbp/>

<https://medlineplus.gov/genetics/condition/amyotrophic-lateral-sclerosis/#causes>

<https://medlineplus.gov/genetics/gene/fgfr3/#:~:text=Achondroplasia&text=Two%20mutations%20in%20the%20FGFR3,change%20in%20the%20FGFR3%20protein>.