1. WHAT CHROMOSOME DID YOU CHOOSE?

Chromosome 11

2 & 3. STATE THE NUMBER OF GENES AND BASE PAIRS ON THE CHROMOSOME YOU CHOSE.

Approx. 2000 genes, over 130 million base pairs

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

GRK2 (G protein-coupled receptor kinase 2)

5. STATE THE FUNCTION OF THE GENE YOU LISTED IN #4.

"This gene encodes a member of the G protein-coupled receptor kinase family of proteins. The encoded protein phosphorylates the beta-adrenergic receptor as well as a wide range of other substrates including non-GPCR cell surface receptors, and cytoskeletal, mitochondrial, and transcription factor proteins. Data from rodent models supports a role for this gene in embryonic development, heart function and metabolism. Elevated expression of this gene has been observed in human patients with heart failure and Alzheimer's disease."

6. WHAT IS THE **SECOND** SEQUENCE DESCRIPTION MATCH FOR YOUR QUERY SEQUENCE?

Homo sapiens CFTR (CFTR) gene, partial cds

7. WHAT DOES THE ENCODED PROTEIN DO IN THE BODY?

"This gene encodes a member of the ATP-binding cassette (ABC) transporter superfamily. The encoded protein functions as a chloride channel, making it unique among members of this protein family, and controls ion and water secretion and absorption in epithelial tissues. Channel activation is mediated by cycles of regulatory domain phosphorylation, ATP-binding by the nucleotide-binding domains, and ATP hydrolysis. Mutations in this gene cause cystic fibrosis, the most common lethal genetic disorder in populations of Northern European descent. The most frequently occurring mutation in cystic fibrosis, DeltaF508, results in impaired folding and trafficking of the encoded protein. Multiple pseudogenes have been identified in the human genome."

8. FOR WHAT DISEASE IS A MUTATED FORM OF THIS GENE RESPONSIBLE?

Cystic fibrosis

9. ON WHAT CHROMOSOME IS THE GENE LOCATED?

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?

Gorilla gorilla gorilla

11. WHAT IS THE COMMON NAME FOR THIS SPECIES?

Western lowland gorilla

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

Not really surprised considering we share a recent common ancestor.

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?

Saimiri boliviensis boliviensis

b. WHAT IS THE COMMON NAME?

Bolivian squirrel monkey

c. ARE THERE ANY GAPS BETWEEN THE TWO SEQUENCES (THE ONE YOU ORIGINALLY SUBMITTED AND ONE THAT HAS LESS THAN 100% QUERY COVER)?

1/119 gaps

14. WHAT IS A GAP IN SEQUENCE ALIGNMENTS?

A sequence of consecutive insertions or deletions

FOR EACH, STATE WHAT THE GENE IS (#15-18).

5. NM_145556

Tardbp TAR DNA binding protein [Mus musculus (house mouse)]

"Enables RNA polymerase II cis-regulatory region sequence-specific DNA binding activity and pre-mRNA intronic binding activity. Involved in positive regulation of protein import into nucleus; regulation of circadian rhythm; and regulation of protein stability. Acts upstream of or within RNA splicing. Located in nucleus. Is expressed in several structures, including branchial arch; central nervous system; early conceptus; genitourinary system; and heart. Used to study Grn-related frontotemporal lobar degeneration with Tdp43 inclusions; amyotrophic lateral sclerosis type 10; and frontotemporal dementia. Human ortholog(s) of this gene implicated in Parkinson's disease; amyotrophic lateral sclerosis; amyotrophic lateral sclerosis type 10; and motor neuron disease. Orthologous to human TARDBP (TAR DNA binding protein)."

16. NM_013444

UBQLN2 ubiquilin 2 [Homo sapiens (human)]

"This gene encodes an ubiquitin-like protein (ubiquilin) that shares high degree of similarity with related products in yeast, rat and frog. Ubiquilins contain a N-terminal ubiquitin-like domain and a C-terminal ubiquitin-associated domain. They physically associate with both proteasomes and ubiquitin ligases; and thus, are thought to functionally link the ubiquitination machinery to the proteasome to affect in vivo protein degradation. This ubiquilin has also been shown to bind the ATPase domain of the Hsp70-like Stch protein."

17. NM_001010850

FUS FUS RNA binding protein [Homo sapiens (human)]

"This gene encodes a multifunctional protein component of the heterogeneous nuclear ribonucleoprotein (hnRNP) complex. The hnRNP complex is involved in pre-mRNA splicing and the export of fully processed mRNA to the cytoplasm. This protein belongs to the FET family of RNA-binding proteins which have been implicated in cellular processes that include regulation of gene expression, maintenance of genomic integrity and mRNA/microRNA processing. Alternative splicing results in multiple transcript variants. Defects in this gene result in amyotrophic lateral sclerosis type 6."

18. KJ174530

SOD1 superoxide dismutase 1 [Homo sapiens (human)]

"The protein encoded by this gene binds copper and zinc ions and is one of two isozymes responsible for destroying free superoxide radicals in the body. The encoded isozyme is a soluble cytoplasmic protein, acting as a homodimer to convert naturally-occuring but

harmful superoxide radicals to molecular oxygen and hydrogen peroxide. The other isozyme is a mitochondrial protein. In addition, this protein contains an antimicrobial peptide that displays antibacterial, antifungal, and anti-MRSA activity against E. coli, E. faecalis, S. aureus, S. aureus MRSA LPV+, S. agalactiae, and yeast C. krusei. Mutations in this gene have been implicated as causes of familial amyotrophic lateral sclerosis. Rare transcript variants have been reported for this gene."

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?

Lou Gehrig's disease

20. BLAST is possible because of the submission of DNA sequences to GenBank. WHAT IS GENBANK?

Genbank is the NIH (National Institue of health) genetic sequence database, a database containing all publically known sequences of DNA.

21. First, answer this question: WHAT IS cDNA?

cDNA is synthetic DNA transcribed from an mRNA sequence using reverse transcriptase enzymes

22. WHAT IS THE SEQUENCE MATCH?

hemoglobin beta [Homo sapiens] (60%)

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

yes.

24. IF YOU SAW DIFFERENCES, WHAT WERE THEY?

Compared to the first sequence, the second sequence has an R instead of a G roughly in the middle between AA 361 and 420, and an insertion of S at position 778 in the second sequence.

25. ARE THERE ANY GAPS IN THE SEQUENCE ALIGNMENT?

yes, there is one gap

26. WHAT GENE ENCODES FOR THE POLYPEPTIDE YOU WERE ANALYZING?

fibroblast growth factor receptor 3 isoform 1 precursor [Homo sapiens]

27. WHAT IS THE FUNCTION OF THIS PROTEIN?

"This gene encodes a member of the fibroblast growth factor receptor (FGFR) family, with its amino acid sequence being highly conserved between members and among divergent species. FGFR family members differ from one another in their ligand affinities and tissue distribution. A full-length representative protein would consist of an extracellular region, composed of three immunoglobulin-like domains, a single hydrophobic membrane-spanning segment and a cytoplasmic tyrosine kinase domain. The extracellular portion of the protein interacts with fibroblast growth factors, setting in motion a cascade of downstream signals, ultimately influencing mitogenesis and differentiation. This particular family member binds acidic and basic fibroblast growth hormone and plays a role in bone development and maintenance. Mutations in this gene lead to craniosynostosis and multiple types of skeletal dysplasia."

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

craniosynostosis

29. WHAT IS THE CONNECTION AMONG THE FOLLOWING: NIH, NLM, NCBI, and HHS?

The NIH (National Institute of Health) is the primary agency in the US gov. related to biomedical and public health research. The NLM (National Library of Medicine) is a library which contains information and research and makes it available to scientists, medical professions, and the public. THe NCBI (National Center for Biotechnology Information) provides access to genomic information, creates databases, and develops software for analyzing data. The HHS (Department of Health and Human services) oversees the NIH, FDA, and CDC.

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

One positive thing I encountered was useful online tools and skills used by real researchers.

One negative I encountered was that the red font with all capitals made it feel like you were yelling at us :(