Bri Purvis
Profesor Rinehart-Kim
Genetics- BIO 294
December 7, 2023

Genome Assignment

1. WHAT CHROMOSOME DID YOU CHOOSE?
    a. I choose chromosome 5.
2. STATE THE NUMBER OF GENES
    a. Chromosome 5 contains approximately 1700 genes.
3. STATE THE NUMBER OF BASE PAIRS ON THE CHROMOSOME YOU CHOSE
    a. Chromosome 5 contains approximately 180 million base pairs. Over 95% of these base pairs have already been determined.
4. LIST ONE GENE THAT IS LOCATED ON THIS CHROMOSOME
    a. One gene that is located on this chromosome is asthma.
5. STATE THE FUNCTION OF THE GENE YOU LISTED IN #4.
    a. Asthma affects more than 5% of people in the United States. It's a chronic inflammatory disorder that affects airways. Asthma symptoms include coughing, shortness of breath, and chest tightness, these can be made worse or triggered by things like viral respiratory infections, exercise, or smoking/ being around smoke. Physiological symptoms of asthma are narrowing of airways caused by fluid within cellular tissue space (edema) or an influx of inflammatory cells into the airway. There hasn't been a gene site for asthma specifically identified, it affects a number of chromosomes: 5, 6, 11, 12, and 14 have all been suspect to contribute to this heritable disease.
6. WHAT IS THE SECOND SEQUENCE DESCRIPTION MATCH FOR YOUR QUERY SEQUENCE?
    a. The second sequence description match that was produce from the query sequence search was Homo sapiens CFTR gene, partial cds.
7. WHAT DOES THE ENCODED PROTEIN DO IN THE BODY?
    a. This encoded protein is in charge of chloride channel in the body. It controls ion, water secretion, and absorption in epithelial tissue.
8. FOR WHAT DISEASE IS A MUTATED FORM OF THIS GENE RESPONSIBLE?
    a. Mutations of this gene can cause cystic fibrosis. The most common mutation of this gene results in impaired folding and trafficking of the encoded protein.
9. ON WHAT CHROMOSOME IS THE GENE LOCATED?
    a. This gene is located on Chromosome 7
10. CHOOSE A SPECIES (STATE THE SCIENTIFIC NAME) OTHER THAN HOMO SAPIENS THAT ALSO HAS A 100% IDENTITY (Per. Ident)FOR THIS SEQUENCE?
    a. Pongo abelii
11. WHAT IS THE COMMON NAME FOR THIS SPECIES?
    a. Sumatran orangutan
12. 12. DOES IT SURPRISE YOU THAT THIS SPECIES ALSO HAS A 100% SIMILARITY IN IDENTITY? WHY OR WHY NOT?
    a. This doesn't surprise me that a primate has 100% similarity in identity. We are thought to be descended from primates so it makes sense that we would still share genetic similarities.

13. (A)WHAT IS THE GENUS AND SPECIES WITH THIS NUCLEOTIDE SEQUENCE? (B) WHAT IS THE COMMON NAME? (C) HOW MANY GAPS OCCUR BETWEEN THE TWO SEQUENCES (THE ONE YOU ORIGINALLY SUBMITTED AND ONE THAT HAS LESS THAN 100% QUERY COVER)?
    a. Sapajus apella
    b. Tufted capuchin
    c. There is one gap between the original submission and the one that has less than 100% query coverage. Sapajus apella has a 98% query coverage.
14. WHAT IS A GAP IN SEQUENCE ALIGNMENTS?
    a. A gap in sequence alignments are the absence of a region that is not present in another sequence.
15. STATE WHAT THE GENE NM_145556 IS
    a. Mus musculus TAR DNA binding protein (Tardbp), transcript variant 1, mRNA
16. STATE WHAT THE GENE NM_013444 IS
    a. Homo sapiens ubiquilin 2 (UBQLN2), mRNA
17. STATE WHAT THE GENE NM_001010850 IS
    a. Homo sapiens fusion (involved in t(12;16) in malignant liposarcoma) (FUS), transcript variant 2, mRNA
18. STATE WHAT THE GENE KJ174530 IS
    a. Homo sapiens superoxide dismutase-1 (SOD-1) gene, exon 1 and partial cds
19. WHAT DISEASE IS ASSOCIATED WITH MUTATIONS OF THE GENES REFERENCED IN #15-#18? WHAT IS A "COMMON NAME" OF THE DISEASE?
    a. (15) Amyotrophic Lateral Sclerosis 7 (ALS7)
    b. (16) Dyspnea in Amyotrophic Lateral Sclerosis 15 (DALS-15), with or without frontotemporal dementia (FTD)
    c. (17) Amyotrophic Lateral Sclerosis 6 (ALS6), with or without frontotemporal dementia (FTD)
    d. (18) Craniosynostosis and multiple types of skeletal dysplasia
20. WHAT IS GENBANK?
    a. A GenBank is an open access database of all the annotated collection nucleotide sequences and their protein translations, it is publicly available.
21. WHAT IS cDNA?
    a. cDNA is also called complementary DNA, is synthetic DNA that has been transcribed from mRNA that has been reacted by using enzymes reverse transcriptase. cDNA only contains coding sequences unlike DNA which is composed of both coding and non-coding sequences.
22. WHAT IS THE SEQUENCE MATCH?
    a. Homo sapiens beta globin gene
23. DO YOU SEE ANY DIFFERENCES BETWEEN THE TWO AMINO ACID SEQUENCES?
    a. Yes, there are two difference between the two animal acid sequences;361 and 721
24. IF YOU SAW DIFFERENCES, WHAT WERE THEY?
    a. The difference of amino acid 361 is the first person has a G where person two has R. The difference of amino acid 721 is the first person has a dash line (-) where person to has an S.
25. ARE THERE ANY GAPS IN THE SEQUENCE ALIGNMENT?

a. Yes, there is one gap in the sequence alignment.

26. WHAT GENE ENCODES FOR THE POLYPEPTIDE YOU WERE ANALYZING?
    a. fibroblast growth factor receptor 3 (FGFR3)
27. WHAT IS THE FUNCTION OF THIS PROTEIN?
    a. The function of this protein is for protein coding
28. WHAT HUMAN DISEASE IS CAUSED BY A MUTATION IN THIS GENE?
    a. The human disease that is cause by this mutation is achondroplasia
29. WHAT IS THE CONNECTION AMONG THE FOLLOWING: NIH, NLM, NCBI, and HHS?
    a. National Institutes of Health (NIH) and The National Library of Medicine (NLM) are all under the U.S. Department of Health and Huamn Service (HHS). The HHS enhances the health and overall well being of Americans, by providing health and human service and making advances in sciences. The National Center for Biotechnology Information (NCBI) fits into this because it is under the NLM. All are approved and funded by the government.
30. REFLECT ON ONE THING THAT YOU LEARNED FROM DOING THIS ASSIGNMENT.
    a. The one thing that I learned from this assignment is there really are a lot of genes that go into making a person. I knew there was a lot but this assignment really put it in perspective, from the second and third question. The chromosome I choose had 1700 genes and 180 million base pairs, there is still 5% that are unknown.