Christopher Torres

Professor Rinehart-Kim

BIO 294

December 8th, 2022

Genome Assignment

Chromosome Maps

1. I decided to choose chromosome 5.
2. There are approximately 1700 genes on chromosome 5
3. There are approximately 180 million base pairs on chromosome 5. 95% have been discovered
4. One gene that is located on this chromosome is the gene for asthma.
5. Normally, this gene codes for key molecules that play a role in the inflammatory response, like tryptase, growth factors, cytokines, etc.

Introduction to BLAST

6. The top sequence description match for the query sequence that was not predicted was: Homo sapiens CFTR promoter region (LOC111674463) on chromosome 7
7. This sequence is another part of the gene. More specifically, it represents the promoter region of the cystic fibrosis transmembrane conductance regulator (CFTR) gene. Its primary purpose is to initiate transcription of this gene with the help of various types of transcription factors, enhancers, etc. This gene can also be expressed in various cell types.
8. The encoded protein associated with the sequence given takes effect in epithelial cells that secrete types of mucus that is associated with saliva, mucus, etc. It does this by regulating epithelial cell ion transport in/out of the cell.
    a. Paper used: Bergeron C, Cantin AM. New Therapies to Correct the Cystic Fibrosis Basic Defect. Int J Mol Sci. 2021 Jun 8;22(12):6193. doi: 10.3390/ijms22126193. PMID: 34201249; PMCID: PMC8227161.
9. A mutated form of the CFTR gene could lead to the well-studied disease of Cystic Fibrosis.
10. This gene is located on chromosome 7.
11. The other species that has a 100% identity other than Homo sapiens is Pongo abelii.
12. The common name for this species is the Sumatran orangutan.

13. It does not surprise me that this animal also has a 100% identity for this gene because orangutans like monkeys are genetically close to humans, so it would make sense that they would also have a 100% match for this gene.
14. Scientific Name: Saimiri boliviensis boliviensis
    a. Common Name: Bolivian squirrel monkey
15. Between these 2 matches, there is only 1 gap that is found in the Bolivian squirrel monkey.
16. A gap in sequence alignments means that there are either one/more residues of amino acids that have been deleted from the sequence, or that there was an insertion of an amino acid residue found in the second sequence.
    a. Source: https://proteinstructures.com/sequence/sequence-alignment/#:~:text=A%20gap%20in%20one%20of,insertion%20in%20the%20second%20sequence.
17. This gene is otherwise known as the TARDBP gene. The gene product of this gene is the TAR DNA binding protein.
18. This gene is otherwise known as the UBQLN2 gene. The gene product of this gene is the protein ubiquilin, which is a ubiquitin like protein that tags other proteins for disposal.
19. This gene is otherwise known as the FUS gene. The gene product is the FUS RNA binding protein (FUS) that is a protein that takes part of the heterogeneous nuclear ribonucleoprotein (hnRNP) complex. This complex assists with pre-mRNA splicing and the export of fully processed mRNA to the cytoplasm.
20. This gene is otherwise known as the SOD-1 gene. The gene product of this gene is a protein called Superoxide dismutase 1, this protein binds to copper/zinc ions and destroys free superoxide radicals in the body.
21. The disease, Amyotrophic Lateral Sclerosis, is the disease that is associated with mutations of the genes represented in #17-20. The common name for this disease is ALS.
22. GenBank is the NIH genetic sequence database that is a collection of all DNA sequences that have been sequenced and available to the public.

Introduction to Swiss-Prot to Study Protein Sequences

23. cDNA is the DNA that is synthesized from a single-stranded RNA template from the enzyme called reverse transcriptase. We can get cDNA in the lab from the process of reverse transcription.
    a. Source - https://www.genome.gov/genetics-glossary/Copy-DNA
24. Using the same program as before, the top searches have a 100% sequence match. More specifically, it seems to be part of the HBB gene for hemoglobin for Homo sapiens.
25. The open reading frame is the portion of a DNA molecule, between the start and stop codons, that when translated to amino acids, have no stop codons. Only readable nucleotides.
    a. Source - https://www.genome.gov/genetics-glossary/Open-Reading-Frame

26. From what I know about amino acids, M for amino acids, represents the amino acid Methionine. Also represents the start codon AUG.
27. I believe that frame 3 is most likely to be an open reading frame because it begins with Methionine, a start codon, and does not seem to end because there is no stop codon near the 3’ end.

Amino Acid Sequence Comparisons

28. Yes, there is a slight difference between the two sequences.
29. The differences are the G(1) and R(2) on base/nucleotide number 380, and the – blank - (1) and S(2) on base/nucleotide number 783.
30. The function of this protein is to assist in the differentiation process for cells to become fibroblasts. It also plays a role in bone development and maintenance.
31. One human disease that can be caused by a mutation of the FGFR3 gene is Achondroplasia. Another condition caused by this mutation is a condition known as Muenke Syndrome.
32. After completing this assignment, I learned that there are various resources that people can use to conduct research of their own, whether it is for professional genetic research or just to better understand the genetics of certain genes/conditions.