Final Project Proposal

Literature Review:

Williams syndrome occurs when people are missing some genes and it results in certain types of physical traits such as bubbly, extroverted personalities, heart defects, intellectual disability etc. In a study [1] it was seen that **WBSCR17** and other genes near it were important in dog evolution. Variations in this same region have been shown to correlate with **Williams syndrome** in humans [2]. In 2017 [3], researchers found that threes genes (one of them was **WBSCR17**) were correlated with the behavioral traits of dogs and wolves. For these three genes some variants were found mostly in friendly dogs and wolves and others were found more often in less friendly animals. This information hints towards the possibility that there might be some similarity between the gene sequences of humans with **Williams Syndrome** , dogs and friendlier wolves.

Our Plan:

In our project, we plan to compare this specific region of DNA (which contains the gene WBSCR17) in dogs and wolves (both friendly and unfriendly) with the functionally equivalent region of DNA (which contains gene WBSCR17) in humans with and without Williams syndrome.

Our Goal:

We want to find out if there is similarity between the DNA of domesticated and friendly animals and people with Williams Syndrome. Also, we want to see if there is similarity between the DNA of unfriendly animals and people without Williams Syndrome. Instead of comparing the DNA, we plan to convert them to mRNA and then compare the RNA sequences based on well-established algorithms (ex: Edit Distance, LCS, Sequence Alignment, KMP, Z-value etc.). For now, we plan to write the code ourselves instead of using tools. We also plan to predict the secondary structure of the mRNA based on MFE and then compare the secondary structures as another means of comparison.

Data source:

For our project, we have access to the original paper's data which can be found on NCBI's website in the SRA section. We were pointed to it by the paper's author over email. An example can be found here [https://www.ncbi.nlm.nih.gov/]. We don't yet fully understand the format of this data or how to connect the samples to the individuals in the study. As a backup, we plan on using reference genomes for chromosome 6 in both Canis lupus and Canis lupus familiaris, found on NCBI's website under the nucleotide section. Finding genomic sequences of humans with Williams syndrome has proven to be more difficult. In the event that we cannot find any,

we can still examine the location associated with this condition (Chr band 7q11.23) in healthy human genomes and search for the same type of mutation (microdeletion) in the same functional area of Canis lupus familiaris genomes as compared to Canis lupus genomes.

Further Work:

If there is still time, we want to generate Structural Variant tables for animals (dogs and wolves) and humans (with and without William Syndrome) based on our data (similar to the Structural Variant table shown in 3). Then we plan to compare our results based on different parameters to find out if there is any significant relation.

Timeline:

Week	**Update**
Week 7-8	Collecting and analyzing data; Figuring out which reads to use for our project
Week 9-10	Converting the DNA sequences to mRNA sequences, comparing them using various techniques and analyzing the results
Week 11-12	Predicting the Secondary Structure of RNA sequences using MFE
Week 13-15	Comparing the folded structures and analyzing their results
Further work	Generating Structural variant tables and comparing them to figure out if there is any significance

Reference

[1] Pollinger, J. P. (n.d.). Genome-wide SNP and haplotype analyses reveal a rich history underlying dog domestication. *Nature, 464(7290), 898*.

[2] Rogers, N. (n.d.). Rare Human Syndrome May Explain Why Dogs are So Friendly. *Inside Science*.

[3] Shuldiner, E. K. (n.d.). Structural variants in genes associated with human Williams-Beuren syndrome underlie stereotypical hypersociability in domestic dogs. *Science Advances, 3(7), e1700398*.