Rao Vinnakota
CMSC 220
Kerri-Ann Norton
April 23, 2018

Project Proposal

For my final project, I'll be using multiple sequence alignment, and classification algorithms to classify breast cancer vs normal cell lines. As mentioned, the process will take several steps:

- Gather the data on the various cell types from NCBI (National Center for Biotechnology Information)

- I'll use multiple sequence alignment to align the various sequences

- Calculate the similarity scores of the sequence, and then cluster the sequences into two groups. This step will require implementing a classification algorithm

- Determine rate of false positives and negatives

- Use hypothesis testing to compare the two similarity scores and make a statistically significant comparison

Personally, I've always found cancer research interesting. It's been years since research started, and so much about cancer has been discovered, but there's still so much to know.