Number Ninjas

Bofan Chen

Anton Hung

Kevin Rouse

Key Issues Solved -
- Found which genes were different between the groups
- Determined which group was healthy as well as what the ailment was using these genes
- Developed a diagnostic tool for the disease
- Explored ethical and equitable aspects of this diagnostic tool

Analytics-
- Used Wilcoxon ranked sum test with Bonferroni correction between the groups to determine which genes to analyze and confirmed these genes with random forest and support vector machine
- Compared up and down-regulation of our genes to the literature and determined the disease
- Justified how our disease is related to the gene pathways obtained from KEGG

Results-
- The disease of interest was colorectal cancer, and we used the literature trends to classify the two groups as sick/not sick
- Determined that the metabolic pathways from KEGG were likely due to the behavior of cancer cells versus normal cells
- Developed a diagnostic tool using the KNN machine learning algorithm

Relevance-
- Given the "perfectness" of the data, we would be hesitant to use this diagnostic tool in a clinical setting
- Although we determined that the disease was colorectal cancer, it is possible that different cancers express these genes similarly and the disease is a different form of cancer
- Given patient symptoms along with medical results, this dataset could be better analyzed