Identification of Lung Cancer Biomarkers Through Coding Sequences Targeted by miRNAs

Yang, Kyle (School: Carmel High School)

Lung Cancer as of the current day is one of the deadliest cancers around the globe due to its high incidence and mortality rates. Identifying factors aiding researchers in developing treatments for Lung Cancer Patients is the purpose of my project. I selected 55 highly expressed lung cancer candidate genes by analyzing the TCGA lung cancer dataset. Three prime untranslated region sequences, or three prime UTR, coding sequences, miRNAs target regions, and superfamilies were all studied using a variety of bioinformatics tools and databases. Nine genes were highly conserved between Humans, Mice, and Groundhogs, according to the results of my sequence alignment results provided by MUSCLE V5. I then checked the Tarbase Database to see if any miRNAs target these nine genes. I found that only ITM2B, CD74, and SFTPB had miRNA target information available in the database. To check if the genes were targeted by miRNAs in their coding regions, I used Diana-MicroT to identify specific Coding Sequences (CDS) targeted by miRNAs. There were four CDS targeting miRNAs, hsa-miR-33a-3p, hsa-miR-409-3p, hsa-miR-545-5p, and hsa-miR-2110. Analyzing the data from the UCSC Genome Browser, I conclude that the targeting sites by hsa-miR-2110 for CD74 are not as specific as ITM2B targeting sites by other miRNAs. Furthermore, the two CDS targeting miRNA, hsa-miR-33a-3p, and hsa-miR-409-3p have similar amino acid binding patterns. Through Uniprot ID-Mapping and NCBI CD-Search, I was able to obtain many common superfamilies between these genes targeted by miRNAs. My research provides clues for identified candidate biomarkers through their CDS regions targeted by miRNAs.

Awards Won: