Identification and Analysis of Key Candidate Genes and Pathways in Lung Adenocarcinoma by Integrated Bioinformatics

Gupta, Prateek (School: American Heritage School)

Lung adenocarcinoma (LUAD) accounts for forty percent of all lung cancer cases in the United States. Treatment currently focuses on surgery and chemotherapy but survival rates remain lower than fifteen percent. The purpose of this investigation is to use an integrated bioinformatics approach to identify key mutated genes in LUAD that are differentially expressed, environmentally responsive, functionally important, and affect survival rates in patients. First, 23,483 genes from nine cancer databases and published studies were identified that are either differentially expressed, mutated, or environmentally responsive. From these, there were 107 significant genes common to all 9 sources used for further analysis. The Database for Annotation, Visualization and Integrated Discovery showed that many of the 107 genes were significantly involved in common processes and pathways, such as DNA replication, DNA repair, ATP binding, and pathways in cancer. Among the 107 genes, the Comparative Toxicogenomics Database found 38,532 gene-environment interactions. A protein-protein interaction network was mapped with the Search Tool for the Retrieval of Interacting Genes/Proteins. This revealed 1116 protein-to-protein interactions among the 107 genes, 25 of which had more than 30 interactions each. Finally, these 25 genes were analyzed with OncoLnc, which resulted in three genes, Checkpoint Kinase 1, Cyclin E1 and Exonuclease 1 (EXO1), that had a significant correlation between increased differential expression in LUAD and worsened patient survival. Among these, EXO1 has been the least studied in LUAD. This study can help focus research on genes likely significant in LUAD and use those genes as targets in gene therapies.