CIRCA - CircularRNA for Cancer Active Immunotherapy: A Machine Learning Model To Predict Liver Cancer and Top Genes for Cancer Vaccine

Koushik, Aditya (School: La Cueva High School)

Circular RNAs (circRNAs) are closed long non-coding RNAs with excellent prognostic and diagnostic biomarker properties for many diseases including cancer. By using liver tissues of Hepatocellular Carcinoma (HCC) patient dataset, this study designed and tested a robust machine learning pipeline to predict circRNA targeted hub genes and their immunogenicity for immunotherapy for cancer. First, publicly available circRNA microarray datasets were analyzed in Python for the top twelve strongest deregulated circRNAs in tumor tissue compared to adjacent healthy tissue from HCC patients. Next, multiple classification models were trained and tested on the circRNA data and evaluated on multiple metrics. microRNA (miRNA) and gene targets (mRNA) of the deregulated circRNAs were predicted and top circRNA targeted hub genes were found from gene interaction network analysis in Cytoscape. Finally, an immunogenicity predictor in Python was built with a T-cell epitope prediction framework. This study found: 1) hsa_circ_0005284 is significantly positively correlated (upregulated) with tumor occurrence and hsa_circRNA_089372 is significantly downregulated in tumor occurrence, 2) the Logistic Regression and Naive Bayes classification models most accurately predicted tumor occurrence based on circRNA data, 3) the TMED10 and RAB1A hub genes were predicted to be the most immunogenic hub genes in Python analysis. In conclusion, this project shows a possible therapeutic option involving direct targeting of hsa_circ_0005284 and hsa_circRNA_089372, and shows that the RAB1A and TMED10 peptides can be promising candidates for a liver cancer vaccine because of their high immunogenicity.

Awards Won: