To Identify a Multi-Gene Biomarker Based on Most Frequently Occurring Genetic Mutations Across Cancers Through Bioinformatics and Verify Its Efficacy to Detect a Wide Range of Cancers Using Statistical Analysis

Saha, Sahishnu (School: James Clemens High School)

Most cancer biomarkers are specific to a cancer type(s). By focusing on the most frequently occurring gene mutations across different types of cancers, a multi-gene biomarker was identified through bioinformatics and its efficacy validated using cancer data through statistical analysis to detect a broad range of 15 cancers. This project used TCGA data of adults between ages 18–89 years covering 33 cancers. Identified 50 genes with the most frequent mutations in 15 types of cancers and selected the common 25 genes as a biomarker using bioinformatics. Used RSEM to calculate the gene expression levels of the identified 25 genes for each of the 15 cancers using data of cancerous and healthy cells to validate that their values differed between cancer patients and healthy individuals. Used Multiple Variable Binary Logistic Regression, for statistical analysis, to determine the accuracy of the 25 gene biomarker in detecting each of the 15 cancers using cancer patient data. Gene expression levels of 25 genes differed between normal and cancerous cells and varied depending on the type of cancer. Cancers with greater gene overlap with the 25 gene biomarker had higher Deviance-R-squared values, obtained through Multiple Variable Binary Logistic Regression, indicating the biomarker's higher ability to detect those cancers. Stomach cancer had the highest Deviance-R-squared of 92% and 10 other cancers had Deviance-R-squared >60%. With Odds Ratio in 95% Confidence Interval >1 and p-value <0.05, it can be inferred that the 25 gene biomarker is potentially an effective detector of at least 11 types of cancers.

Awards Won: