Differential Gene Expression Between Mucinous and Non-Mucinous Colorectal Adenocarcinomas

Bakshi, Aditya (School: Nashua High School South)

Colorectal cancer (CRC) has multiple subtypes, and it may take patients weeks to find out which subtype they have. The goal of this research is to use gene expression data to determine potential genetic biomarkers and create an efficient machine learning model that can distinguish between two common subtypes of CRC, mucinous and non-mucinous adenocarcinomas. It is hypothesized that significant differential gene expression is present between mucinous and nonmucinous adenocarcinomas, and staging may affect expression. 414 cases of non-mucinous adenocarcinomas and 62 cases of mucinous adenocarcinomas were obtained from the TCGA and analyzed using DESeq2, first for specific stages of CRC, and later across all stages. A Wald test was used to determine the significance of the differential gene expression. The results of this analysis were used to train and test a machine learning model that could distinguish between the subtypes. Two potential biomarkers were found. Gene LY6G6D was found to be downregulated in mucinous adenocarcinomas, and gene BEST4 was found to be upregulated mucinous adenocarcinomas. A lasso machine learning model was created that could differentiate between mucinous and non-mucinous adenocarcinomas with an accuracy score of 0.918. LY6G6D and BEST4 have both been found to be associated with CRC previously, and these findings suggest that expression of these genes may be associated with one specific subtype and stage of CRC more than the other. The lasso machine learning model proved to be effective, and with more samples and future testing, can be made more accurate.

Awards Won: