Investigating Biofilm-Related Biomarkers to Predict Therapy Response in Colorectal Cancer Tumors: Year 2

Manjith, Meghna (School: Wiregrass Ranch High School)

Colorectal cancer (CRC) is the second leading cause of cancer-related deaths. Often covered in biofilms, right sided tumors are deadly and often identified in advanced stages, highlighting the need to identify biofilm-related biomarkers related to CRC. Mucins, a type of biomarker, are upregulated in the presence of biofilms, proposing their use as a novel biomarker usable for predicting therapy response. For this project, a novel connection between mucin biomarkers and therapy responses of four categories of treatments was explored by (i) cleaning relevant CRC datasets (ii) identifying the relationship between mucin levels/ mutation rates and CRC subtypes and (iii) developing a novel model to predict therapy response from mucin biomarkers and patient demographics using a Random Forest Machine Learning Models with specialized hyperparameters and 10-fold cross validation. Datasets were first cleaned and preprocessed using PyJanitor, then 20% of the data was augmented and reconfirmed for cleanliness. Evaluation of the efficacy of specific mucin biomarkers were tested and resulted in an average accuracy of 0.81 when both MUC5B and MUC2 were used (AUC-ROC>0.75). Using the novel biomarkers, a Random Forest Classifier Machine Learning Model was developed, able to predict the therapy response (survival in months) of patients at 0.86 accuracy (AUC-ROC>0.85), suggesting the identification of novel biomarkers that can predict therapy responses for CRC patients. This model will continue to be trained with robust datasets to increase accuracy and efficacy. Future work includes experimentation with different optimization and regularization techniques to prevent any potential over and underfitting of data and delving into deep learning machine learning models

Awards Won: