CanceRNA: A Computational Method for Colorectal Cancer Diagnostic Biomarker Detection and Treatment Prediction

Hombal, Pratham (School: Floyd B. Buchanan High School)

Early detection and treatment of colorectal cancer is vital to prevent metastasis, but past models lack a cohesive pipeline from diagnosis to treatment to determine biomarkers and suggest treatment based on past genetic and demographic datasets. To address this issue, CanceRNA, a more accurate pipeline, was developed to identify stable ceRNA colorectal cancer diagnostic biomarkers and suggest efficacy of a Levamisole treatment against colorectal cancer. In the first part of the model, differential expression analysis was performed on clinical genetic data and an interactomic visualization of these ceRNA genes was constructed. Coexpressed diagnostic biomarkers were identified, revealing a six miRNA-lncRNA signature. In the second part, a novel random forest (RF) model was proposed for treatment effect prediction. Utilizing demographic data from clinical trials, the predicted risk reduction was computed through a linear regression model. For the random forest model, demographic hyperparameters (n_estimators, max_depth, min_samples_leaf, and random_state) were tuned for optimization. For training and validation, the set was divided into treatment and control groups with a 75/25 split, and a T-learner was used to determine absolute risk reduction. A c-for-benefit value for both the random forest and linear regression models was computed, revealing that the random forest model achieved an accuracy score of 0.65, while the logistic regression achieved a score of 0.54. The model proposed could be beneficial in cancer therapeutics, as it allows clinicians to accurately diagnose colorectal cancer in patients based on biomarkers determined with this method and suggest individualized treatment based on demographic factors.

Awards Won: