Non-Invasive microRNA-based Diagnostics for the Detection of Multiple Cancers

Zhou, Grace (School: Laramie High School)

Detecting cancer in blood enables diagnostic before symptoms develop. I developed a diagnostic model based on an Ovarian Cancer Dataset for improved performance from a previous 4-microRNA model based on a Lung-Cancer Dataset by Zhang and Hu (2022), striving to achieve at least 70% sensitivity for 12 other cancer types while maintaining 99% specificity. A Combination Dataset (3792) and a Lung Cancer Dataset (3744) published by Zhang & Hu (2022) from the Gene Expression Omnibus was used. From the Combination Dataset, an Ovarian Cancer data set was split into training and testing datasets, and the training dataset was matched by non-cancer controls in the Lung Cancer Dataset where clinical characteristics were available. A model was developed using limma for differential analysis, 10-fold cross-validation to select the optimal number of miRNAs used, and linear combination of the expression level to derive a diagnostic score. Results were graphed with the area under curve of the Receiving Operational Characteristic curve, where sensitivity and specificity showed model performance. The model achieved AUC of 0.994 in the Ovarian Cancer testing set with an enhanced sensitivity of 80%. Validation from the 12 other cancers showed above 70% sensitivity for 8 of the 12 cancers and four underperformed. A diagnostic model for cancer detection can be developed from detecting microRNAs in the blood with high sensitivity and specificity. Basing the model on a different cancer may lead to higher sensitivity for some cancers and it may be significant to develop a model on the cancers that underperformed.