Investigating the Underlying Molecular Patterns in Aggressive Brain Tumor ETMR to Develop a Robust Diagnostic Prediction Model

Dogra, Saanvi (School: Del Norte High School)

Rare, aggressive pediatric brain tumor ETMR is frequently misdiagnosed, reducing chances of survival. Early diagnosis is crucial due to its short average 12 month prognosis. While ETMR is characterized by amplification in the C19MC gene cluster, this isn't a unique molecular identifier. This project focuses on differentiating ETMR from other brain tumors through specific gene amplification patterns within C19MC and the development of a robust diagnostic model. It was hypothesized that excessively amplified miRNAs in C19MC would be responsible for oncogene functions. NCBI and Capper Dataset IDAT files were processed with R, developing CNV plots for each of 43 ETMR samples. Of the consistent significantly amplified probes in C19MC, ten with the greatest methylation differences were traced to their genes and functions. The 43 cases were then compiled with 43 non-ETMR cases, to prevent data imbalance, and model selection, K-Fold Cross Validation, and hyperparameter tuning were applied with Jupyter Notebook to develop a diagnostic Random Forest Classifier. This study demonstrates that, while the ten most amplified genes in C19MC and genes with high feature-importance from the model consist of oncogenes controlling ETMR indicators, many tumor suppressor genes were also present, potentially due to their deactivation from hypermethylation. These amplifications provide more specific hallmarks of ETMR and potential targets for therapeutic treatment. Additionally, the diagnostic model reached 0.961 AUC and other strong performance metrics, indicating overall improved reliability. Future data augmentation or asymmetric loss functions would further improve the model's accuracy, which, after significant testing and fine tuning, could be utilized in a clinical setting.