Identifying Genetic Biomarkers for Essential Tremor Using Bioinformatics and Machine Learning

Djedjos, Nicholas (School: Mississippi School for Mathematics and Science)

Nearly seven million individuals in the U.S. have Essential Tremor (ET), making it one of the most common neurological disorders. Current ET research aligns it with a Purkinje cell disorder in the cerebellum, the motor control center of the brain. ET is associated with life-threatening neurological diseases such as Parkinson's and dementia, yet still remains understudied. This study uses the raw RNA-seq data from 55 post-mortem cerebellum samples to understand the genetic background of ET. The genetic data were used to develop machine learning models for prognosis and further identification of ET genetic biomarkers. Differential Gene Expression (DGE) identified 86 differentially expressed transcribed gene transcripts ($p < 0.001$, FDR < 0.25). The gene transcripts were then input into Gene Set Enrichment Analysis (GSEA) where five pathways were identified as dysregulated after comparisons with the Hallmark and KEGG gene sets: Fatty Acid Metabolism, Cholesterol Metabolism, Ribosome, Axonal Guidance, and Parkinson's Disease. The gene transcripts were also input in Random Forest and Logistic Regression models for further analyses. After filtering the 86 genes to 32 with Random Forest optimization, the classification model predicted ET and control accurately 85% of the time. Logistic Regression was utilized to analyze the 32 genes individually, and 8 genes had a higher accuracy than 80%: SFTPA2, NLRP14, PLCD1, SCRG1, ANKZF1, INPPFD, EVA1C, and BTN3A1Identifying the aforementioned biomarkers both advanced and corroborated with existing scientific literature and could be used to diagnose ET. The addition of machine learning models with higher statistical power and a larger dataset would strengthen the genetic findings.

Awards Won: