Functional Genetic Biomarkers of Alzheimer's Disease and Gene Expression from Peripheral Blood

Sethi, Amish (School: Pine-Richland High School)
Ni, Andrew (School: Pine-Richland High School)

Detecting Alzheimer’s Disease (AD) at the earliest possible stage is key in advancing AD prevention and treatment but is challenged by confounders with the normal aging processes in addition to other neurodegenerative diseases. Recent genome-wide association studies (GWAS) have identified associated alleles, but it has been difficult to transition from non-coding genetic variants to underlying mechanisms of AD. We sought to reveal functional genetic variants and diagnostic biomarkers underlying AD using machine learning techniques. We first developed a Random Forest (RF) classifier using blood microarray gene expression data from 744 participants in Alzheimer’s Disease Neuroimaging Initiative cohort. After initial feature selection, 5-fold cross-validation of the 100-gene RF classifier achieved an accuracy of 98.4%. The high accuracy of the RF classifier supports the possibility of a powerful and minimally invasive tool for screening of AD. Then, unsupervised clustering was used to identify relationships among differentially expressed genes (DEGs) the RF selected. Results suggest downregulation of global sulfatase and oxidoreductase activities in AD through mutations in SUMF1 and SMOX respectively. Finally, we used Greedy Fast Causal Inference (GFCI) to find potential causes of AD within DEGs. In the causal graph, HLA-DPB1 emerges as the largest node. HLA-DPB1 is downregulated and indirectly causes AD, validated by its mechanisms in the immune system which lead to increased neuron death and the progression of neurodegenerative disorders through its role in T-cell receptors and antibody/antigen production. This study further advances understanding of molecular mechanisms underlying AD and provides potential gene targets for further experimentation.

Awards Won: