Novel Prenatal ASD Diagnosis Tool

Maliar, Emma (School: William A. Shine Great Neck South High School)

Autism spectrum disorder (ASD) is a heterogeneous category of impairments characterized by repetitive behaviors and impaired social communication which is currently diagnosed behaviorally. It is estimated that one in 54 children in the United States has ASD. This disorder is highly heritable, yet has varied genetic causes, posing an obstacle to molecular diagnosis. A public dataset of epigenetic changes in gene expression in 214 maternal blood samples from North Californian mothers with at least one child already diagnosed with ASD was used. There were three diagnosis categories: ASD, non-typical development (non-TD), and typical development (TD). Diagnosis was made based on the child's score on standard scales used to measure development and ASD (ADOS and MSEL). Using R, a box-and-whisker graph was used to plot the value distributions, and volcano plots were used to display differential gene expression. A random forest classification algorithm written using scikit-learn in Python identified genes associated with ASD. An accuracy of 83.87% was achieved using the random forest classification algorithm. The maternal blood transcriptome of ASD and TD children was compared to find 1,254 genes with significant changes in gene expression measured using mean decrease in impurity (MDI) (P-value < 0.05), 603 of which were upregulated and 651 of which were downregulated. Of these 1,254 genes, 433 were also listed in the Simons Foundation Autism Research Initiative (SFARI) dataset of genes with strong evidence of correlation with ASD including ones related to ion signaling and synaptic plasticity. Significant upregulation of genes associated with immunity and epigenetic changes supports theories that these factors can cause ASD.