Integrated Machine Learning and Bioinformatics Analysis Reveal Hub Genes in Innate and Adaptive Immunity Linked to Autism Spectrum Disorder and Non-Typical Neurodevelopment from Prenatal Gene Expression Changes

Payam H. Kasani[1,2], Kee Hyun Cho[1,2 *], Cheol-Heui Yun[3,4,5 *]

[1] Department of Pediatrics, Kangwon National University Hospital, Chuncheon, Republic of Korea

[2] Department of Pediatrics, Kangwon National University School of Medicine, Chuncheon, Republic of Korea

[3] Department of Agricultural Biotechnology, and Research Institute of Agriculture and Life Sciences, Seoul National University, Seoul, Republic of Korea

[4] Center for Food and Bioconvergence, and Interdisciplinary Programs in Agricultural Genomics, Seoul National University, Seoul, Republic of Korea

[5] Institutes of Green Bio Science and Technology, Seoul National University, Pyeongchang, Republic of Korea

* Equal corresponding authors

Kee Hyun Cho; ckhfortune@gmail.com

Cheol-Heui Yun**;** cyun@snu.ac.kr

Abstract

Autism spectrum disorder (ASD) represents a diverse group of conditions characterized by challenges in communication and behavior, generally emerging before age 2. The prenatal period is crucial for ASD development, yet how it relates to the gene expression of expectant mothers, whose children are later diagnosed with ASD compare to typical development, remains unclear. The aim of this study is to identify the important immune-related genes that show differential expression in innate and adaptive immune responses. Gene expression data from the GSE148450 dataset were obtained from the Gene Expression Omnibus, followed by the identification of differentially expressed genes. Enrichment analysis was conducted to explore the signaling pathways using Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses. Furthermore, a protein-protein interaction (PPI) network was constructed using the cytoHubba plug-in of Cytoscape to identify hub genes. The effectiveness of these hub genes in detecting ASD was assessed using four widely-used machine learning models through the receiver operating characteristic curve (ROC). Pathway enrichment analysis revealed several significant pathways, with GO analysis showing that both innate and adaptive immune responses were prominently represented among the 33 upregulated genes compared to the 30 downregulated genes. Furthermore, PPI analysis revealed that in the adaptive immune response, KLRC2, HLA-DRB1, HLA-DQB1, HLA-DRB5, and CLEC6A, were identified as hub genes among the upregulated genes, while HLA-DQA1, KLA-B, HLA-F, HLA-DOA, and MR1 were recognized as hub genes for the downregulated genes in the comparison between typical development group

and ASD group. In the innate immune response, TLR7, RSAD2, C1QB, TREML4, and OAS3, were the hub genes in upregulated genes, and HLA-F, HLA-DOB, HLA-B, CLES5A, and MR1 were among the downregulated genes. Additionally, ROC curve analysis showed that these five hub genes have high diagnostic accuracy for ASD, highlighting their significant diagnostic value. Our study demonstrates a strong correlation between the identified hub genes and ASD development, suggesting that these genes may influence pathogenic mechanisms through immune-related signaling pathways.