[BIOINFO 2025]

Machine learning on RNA-seq enables early detection of Alzheimer's Disease

Hyun Woo Park[1], Ki Wook Lee[1], Jimin Seo[1], Na Young Kwon[1], Hye Jung Min[1], Heeje Cho[1], Balachandran Manavalan[1], and Young-Jun Jeon[1,*]

[1] *Department of Integrative Biotechnology, Sungkyunkwan University, Suwon, South Korea*
** Corresponding author: jeon2020@skku.edu*

Alzheimer's disease (AD) is a progressive neurodegenerative disorder where early and objective diagnosis is hindered by the lack of reliable molecular biomarkers. Existing clinical tools, such as the Mini-Mental State Examination (MMSE) or image-based approaches, largely depend on subjective interpretation and typically detect disease only after significant neurological decline, thereby delaying timely intervention. To address this gap, we employed RNA-sequencing (RNA-seq) of brain-tissue transcriptomes to capture genome-wide expression changes associated with AD and to develop predictive models that integrate molecular data with computational approaches. Using a rigorously designed machine-learning framework, we trained classifiers to distinguish AD from cognitively unimpaired (CU) individuals and to identify transcriptomic signatures characteristic of AD pathology. Data preprocessing, normalization, and quality control ensured robust input for model training. The resulting classifiers consistently revealed discriminative gene sets enriched in neuroinflammation and synaptic signaling—two biological domains central to AD progression. These transcriptomic signatures were not incidental but functionally meaningful, aligning with established pathological mechanisms while highlighting additional expression shifts that may underlie early disease stages. Classification performance was evaluated comprehensively using AUROC, Matthews correlation coefficient, accuracy, sensitivity, and specificity, with results demonstrating strong predictive capacity. Moreover, targeted feature selection and systematic hyperparameter optimization substantially improved generalization across datasets. Collectively, our findings demonstrate that RNA-seq–derived transcriptomic signatures can serve as biologically relevant markers that both discriminate AD from CU and provide mechanistic insight. This integrative approach lays the groundwork for predictive tools that unite diagnostic precision with molecular interpretation, ultimately advancing the prospects for earlier detection and informed therapeutic decision-making.