Transcriptomic profiling and biomarker discovery in pre-eclampsia: An integrated approach leveraging WGCNA and LASSO with ROC validation

Tamil Barathi Palanisamy and **Mohanapriya Arumugam***

Department of Biotechnology, School of Biosciences and Technology, Vellore Institute of Technology, Vellore-632014.

*Corresponding author: mohanapriyaa@vit.ac.in

Abstract

Pre-eclampsia (PE) is still one of the leading causes of maternal and fetal morbidity, affecting 2-8% of pregnancies worldwide. Despite great efforts in research, the precise molecular mechanisms that underlie this complex disorder have not been found. We applied RNA sequencing (RNA-Seq) and advanced bioinformatics approaches to study the pathophysiology of PE. The weighted gene co-expression network analysis (WGCNA) method was used to construct a co-expression network of 239 differentially expressed genes (DEGs) between healthy and PE, which led to the identification of seven specific modules. Two modules, turquoise and yellow, showed strong co-relationships with PE. Further, functional enrichment pointed towards various important biological pathways such as NAD metabolism, HIF-1 signaling and glycolysis/gluconeogenesis, PI3K AKT-signalling pathway and JAK-STAT pathway. Further candidate genes were identified through clustering and analysis of protein interaction network and Least Absolute Shrinkage and Selection Operator (LASSO) regression analyses identified five crucial predictor genes such as GAPDH, LEP, PKM, TRIM24, and NDRG1 which are highly essential in PE. The prognostic potential of identified biomarkers was validated through Receivers Operating Characteristics (ROC) curve analysis, achieving an AUC of 0.981, demonstrating high discriminatory power between healthy and PE groups. Furthermore, drug-gene interaction analysis using DGIDB database revealed interaction for only three biomarkers, such as, GAPDH, PKM and LEP. These integrated systems biology approaches have identified key biomarkers and potential therapeutic targets for PE, providing a strong basis for future research into its molecular mechanism and clinical management.

Keywords: preeclampsia, weighted gene co-expression network analysis, Lasso regression, receiver operating characteristics analysis, drug-gene interaction.