Integrative ML-Guided Multi-Omics and Structure-Based Drug Design to Unveil Diagnostic Biomarkers and Host-Directed Therapeutics Against Dengue Virus

Muthuraja Arun Pravin[1], Ruba Devi Senthil[1], Thiruselvam Premkumar[2] and Sanjeev Kumar Singh[1]*

[1] Computer Aided Drug Design and Molecular Modelling Lab, Department of Bioinformatics, Science Block, Alagappa University, Karaikudi-630004, Tamil Nadu, India.

(Email: pravinarun550@gmail.com, rubadevis2003@gmail.com,

**Corresponding author: skysanjeev@gmail.com)*

[2] School of Bio Sciences and Technology, Vellore Institute of Technology, Vellore-632014, Tamil Nadu, India

(Email: premkumart.2402@gmail.com)

Abstract

Dengue virus (DENV) has become an escalating global health challenge. Despite affecting millions annually, the complexity of its pathogenesis driven by interactions between viral serotypes and host immune responses remains inadequately understood. With no specific antiviral treatment available, understanding the host transcriptional landscape during infection is crucial. Therefore, our study employs high-throughput transcriptomics and systems biology approaches to elucidate key molecular signatures and immune mechanisms underlying DHF progression, with the goal of contributing to more effective diagnostic and therapeutic strategies. As a starting point, the gene expression datasets (GSE279208 & GSE215835) were retrieved from the Gene Expression Omnibus (GEO) database. Following the identification of DEGs between healthy controls, dengue patients, and dengue hemorrhagic patients, a total of 87 common DEGs were identified. Functional enriched analysis revealed these genes were enriched in defense response to virus, and diseases like Asthma, African trypanosomiasis and Coronavirus disease – COVID-19. Subsequently, a PPI network was constructed using STRING Database, and 13 hub genes were identified using the cytoHubba plugin. To further refine gene prioritization, ML-based feature selection and model construction was performed using LASSO and SVM-RFE. The model was statistically validated via ROC analysis, identifying a robust 6-gene signature. Additionally, single-cell transcriptomic profiling of PBMCs was integrated to map identified biomarkers to specific immune cell subsets, reinforcing their mechanistic relevance in DHF progression. Building on these findings, the biomarker was subjected to structure-based drug discovery approaches, including protein modeling, molecular docking and molecular dynamics analysis, to evaluate the druggability

of key host targets and identify potential antiviral compounds. This study provides a comprehensive framework that identifies potential biomarkers for the early diagnosis and prognosis of DHF, while demonstrating a robust integrative pipeline combining multi-omics, systems biology, CADD, and machine-learning approaches.

Keywords: RNA-seq, DEGs, PPI network, LASSO, SVM-RFE, single-cell transcriptomics, molecular docking, molecular dynamics