[BIOINFO 2025]

AI-Driven Transcriptome-based Immune Phenotyping for Predicting Immunotherapy Response in Lung Cancer

Ki Wook Lee[1], Hyun Woo Park[1], Balachandran Manavalan[1], Sehhoon Park[2] and Young-Jun Jeon[1]*

[1] *Department of Integrative Biotechnology, Sungkyunkwan University*
[2] *Department of Medicine, Samsung Medical Center*
**Corresponding author: jeon2020@skku.edu*

Immune checkpoint inhibitors (ICIs) benefit only a subset of patients with advanced non-small-cell lung cancer (NSCLC), and reliable pretreatment predictors remain limited. To address this clinical unmet need, we developed a transcriptome-based AI platform leveraging immune phenotyping as an alternative to image-based TIL quantification, which is labor-intensive, observer-dependent, and sampling-constrained .

Using 449 TCGA-LUAD profiles annotated by LUNIT-Scope, we divided samples into training (80%) and test (20%) sets. Rather than restricting analyses to DEGs, we identified AI-based important genes (AIGs) using six tree-based classifiers across the whole transcriptome, then trained 14 machine learning (ML) algorithms to classify immune-infiltrated (IF) vs non-IF tumors, followed by ensemble refinement. The final model achieved an AUC of 0.907 in training, 0.810 in test, and 0.848 in external validation, supporting transcriptome-derived immune phenotyping as a practical alternative to image-based approaches. Notably, our model outperformed LUNIT-Scope in predicting ICI response based-on immune phenotype within tumor proportion score (TPS) subgroups. Among the AIGs, we discovered two previously unreported genes strongly associated with ICI response, underscoring the novelty and biological relevance of our platform.

We further developed a random survival forest (RSF) model using selected AIGs to predict progression-free survival (PFS) in ICI-treated patients. The AIG-based PFS model demonstrated strong predictive performance (R = 0.94), and its risk score showed consistently high predictive ability for ICI response in training (AUC = 0.964) and across two external validation sets (AUC = 0.887 and 0.849).

Our study highlights that transcriptome-derived ML framework not only demonstrates high accuracy and adaptability but also offers significant potential to improve patient stratification, reduce

ineffective treatments and advance precision immunotherapy in lung cancer.