

Metabolomic Machine-Learning Predictor for Diagnosis of Early-Stage Non-Small-Cell Lung Cancer

Qiang Sun^{1,2}, Yaran Liu², Raojun Lio¹, Lin Zhongjie¹, Ping Wang², Mengzhe Guo³, Shanshan Xie⁴, and Ziyi Zhu¹

1. Department of Thoracic Surgery, Sir Run Run Shaw Hospital, School of Medicine, Zhejiang University, Hangzhou, Zhejiang, China.

2. Center for RNA Medicine, the Fourth Affiliated Hospital of School of Medicine, and International School of Medicine, International Institutes of Medicine, Zhejiang University, Yiwu, Zhejiang, China.

3. College of Pharmacy, Xuzhou Medical University, Xuzhou, Jiangsu, China.

4. Children's Hospital, Zhejiang University School of Medicine, National Clinical Research Center for Child Health, Hangzhou, Zhejiang, China

Objective

Non-small-cell lung cancer (NSCLC) has high cancer-related mortality. LDCT has limitations for screening, and conventional protein biomarkers lack sufficient sensitivity for early-stage NSCLC. This study aimed to characterize plasma metabolic alterations in stage IA NSCLC and build a machine-learning metabolite model for non-invasive early diagnosis.

Conclusions

Stage IA NSCLC shows obvious plasma metabolic dysregulation. The SMM model achieves superior non-invasive early-diagnostic performance compared with routine protein biomarkers. Combining SMM with clinical markers further increases diagnostic sensitivity. This metabolomics-machine-learning strategy has good translational prospects for NSCLC early screening. Further development of quantitative assays and simplified detection kits will support its clinical application.

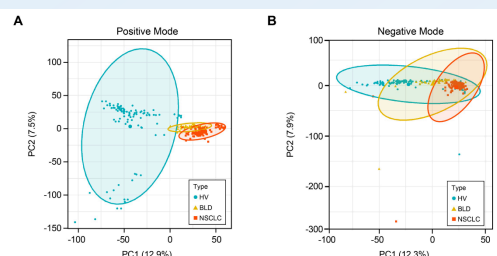


Figure 1. Significant differences in serum metabolic profiles of NSCLC patients.

Methods

- **Study cohorts:** 545 multi-center subjects: 363 pathologically confirmed stage IA NSCLC patients and 182 non-NSCLC controls. Patients receiving prior anti-cancer therapy were excluded.
- **Metabolomic profiling:** Plasma untargeted LC-MS metabolomics was performed. Differential metabolites were screened, followed by KEGG pathway enrichment analysis.
- **Model construction and validation:** LASSO regression for feature selection and random forest algorithm were used to construct the SMM diagnostic model. Model performance was validated in independent cohorts and compared with conventional tumor biomarkers and other machine-learning algorithms.

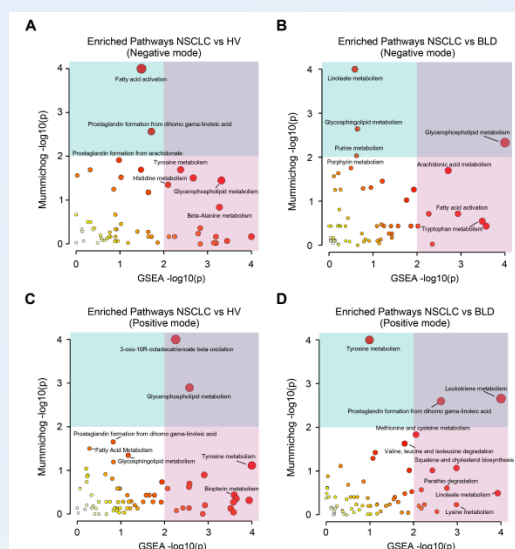


Figure 3. Identification of significantly dysregulated metabolic pathways in NSCLC patients.

Results

- NSCLC patients exhibited distinct plasma metabolic reprogramming. 45 differential metabolites were detected. Glutathione and cysteine-methionine metabolic pathways were significantly disturbed, indicating abnormal oxidative stress and perturbed methyl pool.
- The 9-metabolite SMM model achieved AUROC = 0.815 (95%CI: 0.784-0.846, sensitivity 0.832, specificity 0.910) in external validation. Citrate, kynurenine and arginine contributed most to prediction.
- SMM outperformed conventional tumor-marker panel. Combined use of SMM and routine biomarkers improved sensitivity to 0.874. Random-forest-based SMM showed the best performance among tested algorithms.

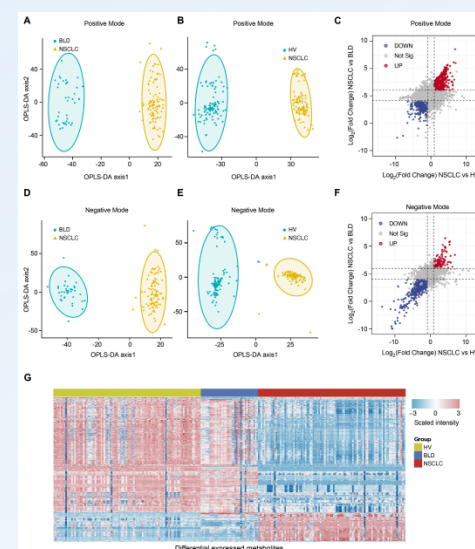


Figure 2. Identification of abnormally expressed serum metabolites in NSCLC patients.

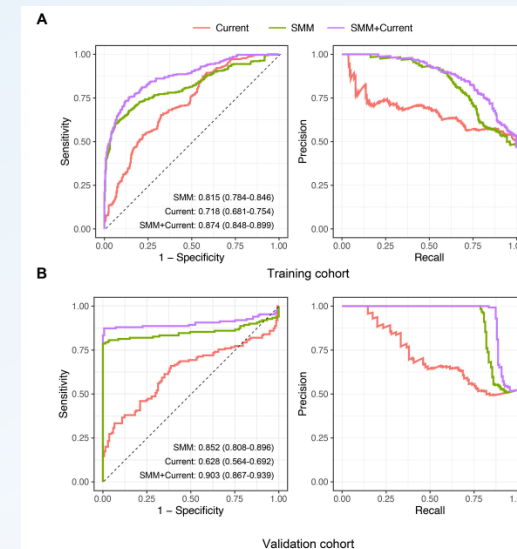


Figure 4. Construction of an early diagnostic model for NSCLC based on differential serum metabolites.