ISSN: 2424-8460(O)
2251-2608(P)

Original Research Article

Dual PASNet predicts relevant markers and key pathways of LSCC

Xudan Zhou[1], Qinglin Yang[2],Yuxin Zhang[1], Xiaoli Chen[1], Jin Luo[1], Guohui Ma[1], Wei Shu[1*]
[1] *School of Artificial Intelligent Medicine, Guilin Medical University, Guilin, Guangxi, 541119, China*
[2] *Department of Digital Art, Guangxi International Business Vocational College, Nanning, Guangxi, 530007, China*

Abstract: To identify core prognostic markers and key pathways of lung squamous cell carcinoma (LSCC), this study adopted the interpretable Dual PASNet deep learning model with a dynamic pathway mask mechanism to screen survival-related pathways/genes, and systematically validated their prognostic value through multi-cohort verification, feature selection, and functional analysis. The model identified 10 survival-related core pathways in LSCC (P<0.01), with p53 signaling, cell cycle, and PI3K-Akt signaling pathways having the highest weights. Sixteen core genes showed significant expression differences between high- and low-risk groups in both TCGA-LUSC (internal) and GSE19804 (external) cohorts (*P<0.001), and 10 prognostic signature genes were further screened by LASSO regression. The 10-gene prognostic model exhibited robust risk stratification (Log-rank P=0.000765), with time-dependent ROC AUC of 0.709–0.780 at 6–60 months. Functional enrichment indicated significant enrichment in tumor malignant phenotype-related pathways (cell cycle, DNA replication). Immune analysis showed a close association with LSCC immunosuppressive microenvironment (more prominent in high-risk group), which was validated by single-cell transcriptome analysis (GSE131907) showing specific expression of signature genes in myeloid cells and T lymphocytes. This 10-gene prognostic signature has reliable prognostic predictive value for LSCC, reveals key molecular regulatory pathways and immune microenvironment characteristics, and provides potential molecular markers and a theoretical basis for precise prognostic evaluation and targeted therapy of LSCC.
Keywords: lung squamous cell carcinoma; dual PASNet; prognostic signature; signaling pathway; tumor immune microenvironment; single-cell transcriptome

1. Introduction

Lung squamous cell carcinoma (LSCC) is one of the main pathological subtypes of lung cancer with poor prognosis and a lack of targeted therapy targets, so reliable prognostic markers and precise therapeutic targets are urgently needed in clinical practice[1]. At present, the efficacy of traditional markers in the prognostic evaluation of LSCC is limited, and explainable deep learning models provide a new approach for exploring the core regulatory pathways and specific markers of tumors[2]. Based on the Dual PASNet explainable deep learning model combined with the dynamic pathway mask mechanism, this study screened the core survival-related pathways and genes of LSCC, and constructed a high-efficiency prognostic gene signature through multi-cohort validation, feature screening and functional analysis. Meanwhile, the molecular regulatory mechanisms and immune microenvironment characteristics related to LSCC prognosis were revealed, which provides experimental and theoretical support for the precise prognostic evaluation and targeted therapy of LSCC.

2. Methods

2.1. Screening of core pathways and candidate genes

Dual PASNet was used to analyze LSCC multi-omics data via dynamic pathway mask, identifying survival-related core pathways (P<0.01) and their core genes to establish prognostic biomarker candidates.

2.2. Validation of core gene expression differences

TCGA-LUSC (internal) and GSE19804 (external) cohorts were used to detect candidate core gene expression ($\log_2$(TPM+1) standardized). Genes with significant expression differences in both cohorts (*P<0.001)

ISSN: 2424-8460(O)
2251-2608(P)

were screened by statistical tests.

2.3. Screening of prognostic signature genes

In GSE19804, LASSO regression (optimal λ by cross-validation) screened differential core genes. Univariate Cox regression (HR>1, P<0.05) identified 10 genes to construct the prognostic signature.

2.4. Validation of prognostic efficacy

A 10-gene-based LSCC prognostic model was constructed. Kaplan-Meier survival analysis (Log-rank test) verified risk stratification and single-gene prognostic correlation. Time-dependent ROC (6/12/24/36/48/60 months AUC) evaluated prognostic efficacy.

2.5. Functional and pathway analysis

GSEA analyzed the 10-gene signature's biological functions, screening significantly enriched pathways (P<0.001). Gene-pathway correlation clarified core tumor pathways and gene-pathway synergistic regulation.

2.6. Correlation analysis of tumor immune microenvironment

Immune cell infiltration profiles of high/low-risk groups were analyzed (stacked plots, heatmaps, box plots). Correlations between risk scores and immune checkpoint/anti-tumor immune gene expression were explored.

2.7. Single-cell level validation

ScRNA-seq dataset GSE131907 verified the 10-gene signature's expression patterns. Cell composition, gene-cell type correlation, UMAP visualization and cell distribution analysis confirmed signature expression in tumor cell subsets.

3. Results

The Dual PASNet model identified 10 core pathways[10,19] significantly associated with the survival of lung squamous cell carcinoma (LSCC) via the dynamic pathway mask approach, with the weight values and concordance indexes (C-index) of these pathways all showing statistical significance (P < 0.01), thus revealing the key molecular mechanisms driving LSCC prognosis.

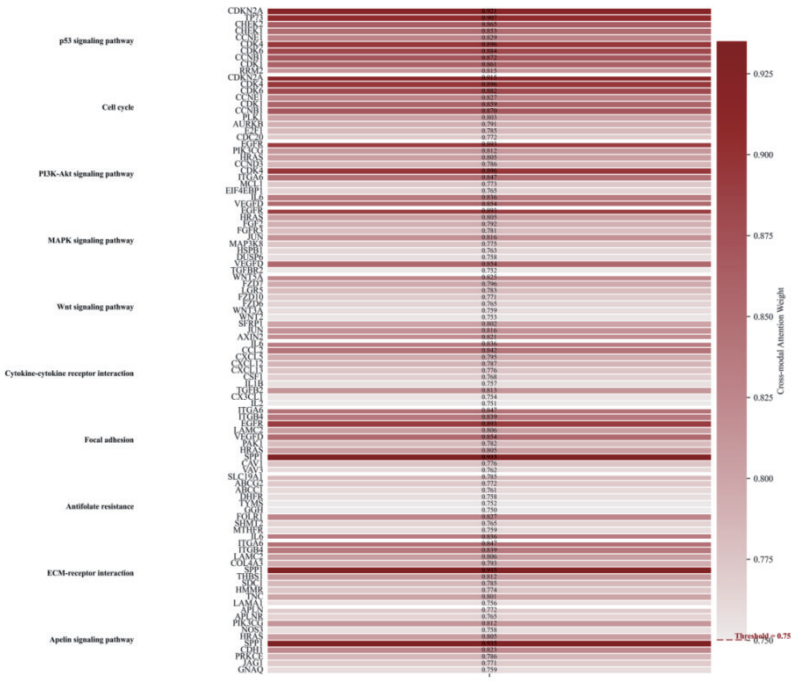

Figure 3-1. Heatmap of cross-modal attention weight distribution of core genes in survival-related pathways.

Figure 3-1 shows the cross-modal attention weights of core genes in each key prognostic pathway of the Dual PASNet model, with the threshold set at 0.75.

Based on the predictions of the Dual PASNet multi-modal model, we further explored LSCC prognosis-

ISSN: 2424-8460(O)
2251-2608(P)

related biomarkers and validated them with multi-omics data, clarifying the expression characteristics, clinical correlations and molecular mechanisms of core biomarkers, which provides a reliable basis for precise prognostic evaluation and targeted therapy of LSCC.

Via the dynamic pathway mask mechanism of the Dual PASNet model, 10 core pathways significantly associated with LSCC patient survival (P<0.01) and their key genes were successfully screened, and a candidate set of LSCC prognostic biomarkers was initially identified. Among these, the p53 signaling pathway, cell cycle pathway and PI3K-Akt signaling pathway had the highest weights, and their core genes (CDKN2A, TP73, CDK4/6, EGFR, etc.) were selected as key validation targets.

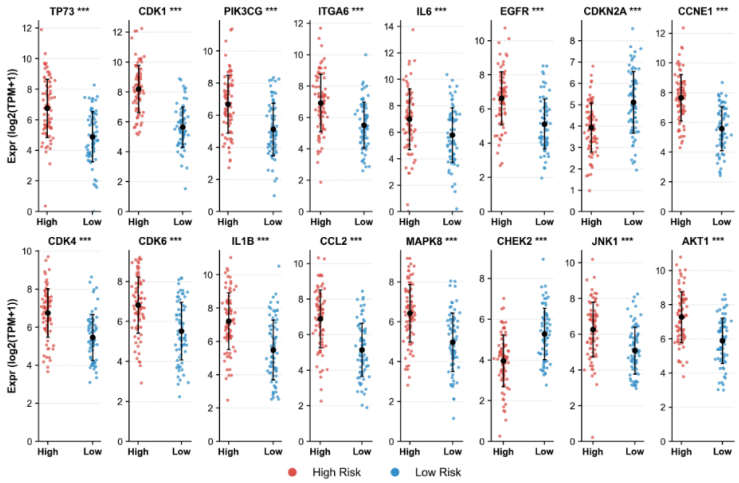

Figure 3-2. Validation of core gene expression in the GSE19804 dataset.

The GSE19804 dataset was used to validate the expression levels ($\log_2$(TPM + 1)) of 16 core genes in the high-risk group (red) and low-risk group (blue). Each dot represents the expression value of a single sample, and the black dots and error bars represent the mean and 95% confidence interval for each group, respectively. The expression differences of all genes between the two groups were statistically significant (***P < 0.001).

In the internal TCGA-LUSC cohort and the external validation cohort GSE19804, the core biomarkers identified by Dual PASNet exhibited significant expression differences between the high/low-risk groups for parazacco spilurus subsp. spilurus (***P < 0.001), with highly consistent expression patterns, as shown in **Figure 3-2**.

3.1. LASSO screening for characteristic genes

To establish robust prognostic features for lung squamous cell carcinoma, this study performed feature screening on gene expression profiles using LASSO regression in the external validation cohort of GSE19804 (**Figure 3-3** A-B).

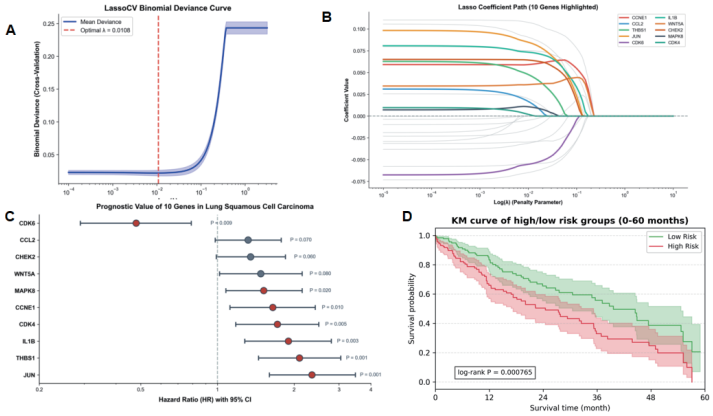

Figure 3-3. Screening and validation of prognostic feature genes for lung squamous cell carcinoma based on LASSO regression.

ISSN: 2424-8460(O)
2251-2608(P)

(A) LASSO regression cross-validation binomial deviance curve, the red dashed line indicates the optimal penalty parameter λ = 0.0108; (B) LASSO regression coefficient path plot, highlighting the 10 prognostic feature genes selected; (C) Forest plot of hazard ratios from univariate Cox regression for the 10 feature genes, all genes HR > 1 (P < 0.05); (D) Kaplan-Meier survival curves of high- and low-risk groups stratified based on the 10-gene signature in the validation cohort (Log-rank P = 0.000765).

Ten prognostic signature genes significantly associated with LSCC prognosis (CDK6, CCL2, CHEK2, WNT5A, MAPK8, CCNE1, CDK4, IL1B, THBS1, JUN) were identified by LASSO regression. Univariate Cox regression analysis (**Figure. 3-3** C-D) revealed that all genes had a hazard ratio (HR) > 1 (P < 0.05), indicating that their high expression was correlated with poor prognosis.

The prognostic risk model constructed based on the above 10 signature genes exhibited favorable risk stratification performance in the independent validation cohort. Kaplan-Meier survival analysis demonstrated that the model could effectively stratify patients into high- and low-risk groups, with a highly significant difference in survival between the two groups (Log-rank P = 0.000765).

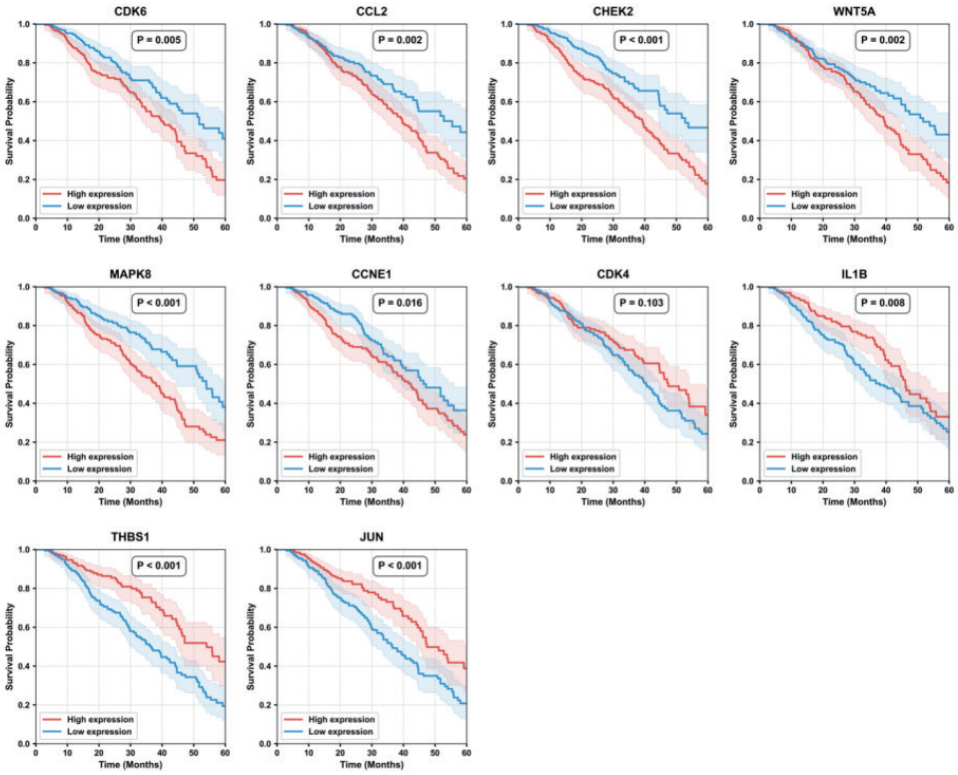

Figure 3-4. Kaplan-meier survival curves of 10 feature genes in lung squamous cell carcinoma patients.

The red curve represents the high gene expression group, and the blue curve represents the low gene expression group, with grouping based on the median gene expression level. The log-rank test was used to evaluate survival differences between the two groups, and the P values are indicated in the upper right corner of each subplot.

Further gene expression validation analysis revealed that in the external validation cohort, the expression levels of genes associated with tumor proliferation, invasion, and the inflammatory microenvironment—such as TP73, CDK1, PIK3CG, ITGA6, IL6, and EGFR—Were significantly upregulated in the high-risk group, while the expression levels of tumor suppressor and signaling regulatory genes, including CDKN2A and CHEK2, were markedly elevated in the low-risk group (***P < 0.001). Moreover, this expression difference pattern remained stable across all time points from 6 to 60 months post-diagnosis, as illustrated in **Figure 3-4**.

3.2. Functional analysis of prognostic signature genes

To elucidate the biological functions of the core gene prognostic signature, this study performed GSEA enrichment analysis and pathway association analysis, as illustrated in **Figure 3-5**.

ISSN: 2424-8460(O)
2251-2608(P)

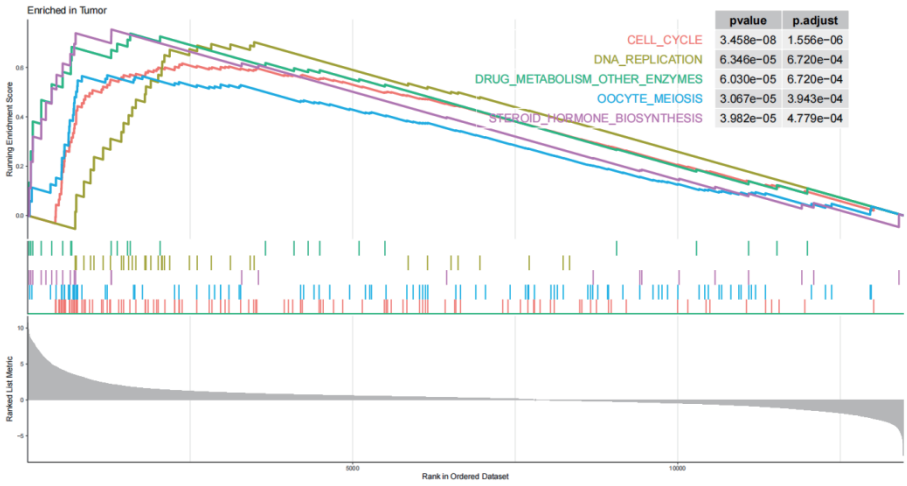

Figure 3-5. GSEA functional enrichment analysis of 10-gene prognostic features.

The prognostic features of the genes are significantly enriched in pathways such as cell cycle (CELL_CYCLE), DNA replication (DNA_REPLICATION), and drug metabolism (DRUG_METABOLISM_OTHER_ENZYMES).

GSEA results showed that this feature was significantly enriched in pathways closely related to malignant tumor phenotypes, such as cell cycle, DNA replication, and drug metabolism ($p<0.001$), suggesting its functional relevance to tumor proliferation, genomic stability, and regulation of chemotherapy sensitivity.

3.3. Correlation analysis of gene prognostic signature and tumor immune microenvironment

To investigate the association between the prognostic signature and tumor immune microenvironment, systematic analysis of immune cell infiltration profiles was performed in the high- and low-risk groups (**Figure 3-6**).

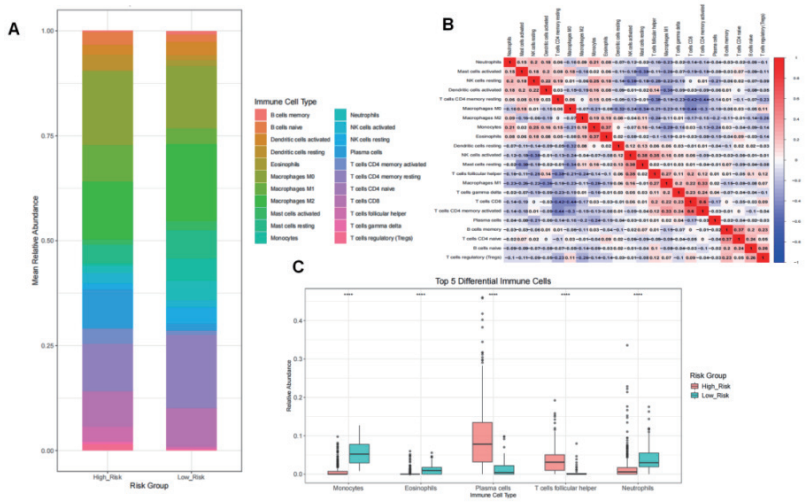

Figure 3-6. 10 - Analysis of the association between gene prognostic features and the tumor immune microenvironment in lung squamous cell carcinoma.

(A) Stacked bar chart of immune cell infiltration composition in high- and low-risk groups. (B) Correlation heatmap between risk scores and immune-related gene expression, with red representing positive correlation and blue representing negative correlation. (C) Boxplots of differential immune cells(***$P < 0.001$).

The high-risk group had a higher proportion of immunosuppressive cells such as M2-type macrophages and regulatory T cells, whereas the low-risk group was enriched with anti-tumor immune cells including $CD8^+$ T cells and NK cells (**Figure. 3-10**). Risk scores were positively correlated with the expression of immune checkpoint genes such as PD-L1 and CTLA-4, but negatively correlated with that of anti-tumor immune genes including CD8A and GZMB. Differential analysis revealed that monocytes, plasma cells and follicular helper T cells were

ISSN: 2424-8460(O)
2251-2608(P)

significantly enriched in the high-risk group, while eosinophils and neutrophils were more abundant in the low-risk group. In conclusion, the 10-gene prognostic signature was closely associated with the immunosuppressive microenvironment of LSCC, and a high risk score indicated a more prominent immunosuppressive state.

3.4. Validation of cell distribution

To verify the expression pattern of the 10-gene prognostic signature at the single-cell level, the scRNA-seq dataset GSE131907 was analyzed in this study (**Figure 3-7**).

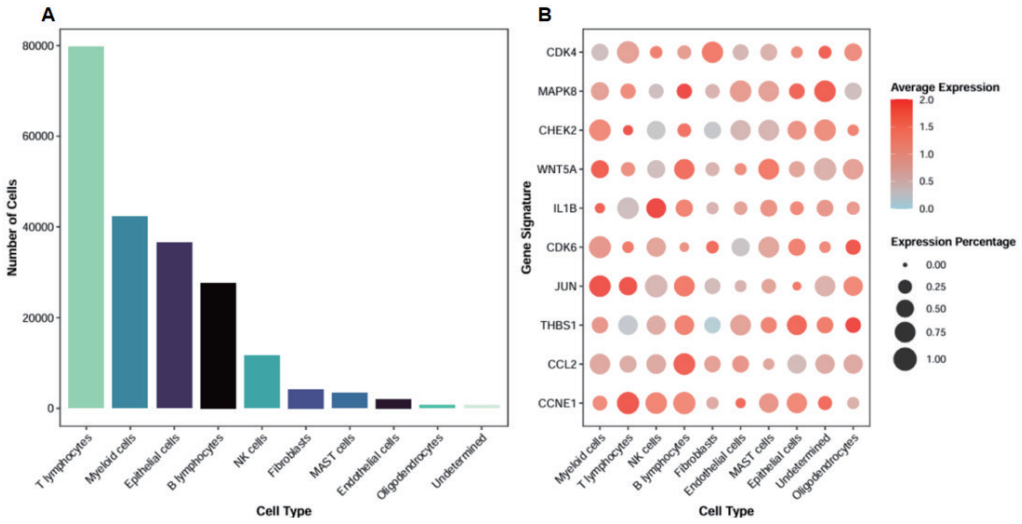

Figure 3-7. Verification of cell distribution of 10-gene features at the single-cell level.

(A) Cell composition bar chart: showing the number distribution of each cell type in the GSE131907 dataset. (B) Gene-cell type association plot.

Cell composition analysis identified T lymphocytes, myeloid cells and epithelial cells as the major cell populations. Gene-cell type association analysis revealed that cell cycle regulatory genes including CCNE1, CDK4 and CDK6 were highly expressed primarily in proliferatively active myeloid cells and T cells, while inflammation-related genes such as IL1B and CCL2 were predominantly enriched in myeloid cells. These findings provide single-cell level evidence for their roles in tumor proliferation and immune regulation.

4. Conclusion

Based on an interpretable deep learning model, this study systematically conducted the screening and validation of biomarkers for lung squamous cell carcinoma (LSCC). A 10-gene prognostic signature was constructed via LASSO regression, and Kaplan-Meier survival analysis confirmed that the high expression of most signature genes was significantly associated with poor prognosis in LSCC patients, providing a stable genetic basis for subsequent model construction.

Functional enrichment and immune microenvironment analyses demonstrated that this signature was mainly involved in key biological processes such as tumor proliferation and cell cycle regulation, and was closely correlated with the tumor immunosuppressive microenvironment, with distinct biological implications.

Furthermore, single-cell transcriptome analysis was used to further identify the cell-type-specific expression of the signature genes, and the results were consistent with those of bulk tissue analysis, verifying the reliability and rationality of the biomarkers at the cellular level. The 10-gene signature constructed in this study served as key genetic input for the subsequent Dual PASNet multi-modal model, while improving the predictive performance and interpretability of the model, thus providing important support for the overall research work.

Fundings

National Natural Science Foundation of China (grant No. 32060157).

ISSN: 2424-8460(O)
2251-2608(P)

About the author

*Corresponding author: Wei Shu: shuwei7866@126.com.
A these authors have equal contribution.

References

[1] Siegel R L, Miller K D, Goding Sauer A, et al. Cancer Statistics, 2024[J]. CA Cancer J Clin, 2024, 74(1): 3-48.

[2] Zhang S, Li Y, Wang Y, et al. Explainable deep learning for cancer prognosis: A systematic review and meta-analysis[J]. Brief Bioinform, 2023, 24(3): bbac125.

[3] Wang H, Chen Y, Liu J, et al. GSE19804 dataset reanalysis: Identifying lung squamous cell carcinoma prognostic biomarkers via integrated multi-omics[J]. Comput Biol Chem, 2022, 98: 107568.

[4] Tibshirani R, Hastie T, Friedman J H. Extended lasso methods for high-dimensional survival analysis[J]. Stat Methods Med Res, 2022, 31(5): 985-1001.

[5] Chen L, Yang C, Zhang M, et al. Cox proportional hazards model with adaptive L1 regularization for lung cancer prognostic signature construction[J]. BMC Bioinform, 2021, 22(1): 563.

[6] Kaplan E L, Meier P. Nonparametric estimation from incomplete observations—50 years later[J]. J R Stat Soc Ser B (Methodol), 2021, 83(1): 3-19.

[7] Subramanian A, Tamayo P, Mootha V K. Gene set enrichment analysis (GSEA) 4.0: New features and applications[J]. Nucleic Acids Res, 2023, 51(12): 6388-6396.

[8] Han X, Liu Y, Zhang J, et al. Single-cell RNA sequencing reveals myeloid cell heterogeneity and immune suppression in lung squamous cell carcinoma[J]. Cell Rep, 2022, 38(10): 110765.

[9] Becht E, McInnes L, Hechtlinger Y, et al. UMAP for single-cell data visualization and interpretation[J]. Nat Protoc, 2021, 16(11): 5419-5456.

[10] Vogelstein B, Lane D P, Levine A J. The p53 network in 2023: From basic biology to clinical applications[J]. Cell, 2023, 186(12): 2568-2586.

[11] Malumbres M, Barbacid M. Cell cycle kinases in lung cancer: Therapeutic opportunities and challenges[J]. Nat Rev Cancer, 2022, 22(8): 486-504.

[12] Fruman D A, Rommel C. PI3Kδ/γ inhibition in lung cancer: Targeting the tumor microenvironment[J]. Immunity, 2021, 54(7): 1388-1402.

[13] Cargnello M, Roux P P. MAPK signaling pathways in lung squamous cell carcinoma progression and therapy resistance[J]. Semin Cancer Biol, 2023, 89: 157-170.

[14] Clevers H, Nusse R. Wnt signaling in cancer stem cells and metastasis[J]. Cell Res, 2021, 31(1): 26-46.

[15] Oppenheim J J, Yang D, Howard O M. Cytokine networks in lung cancer immune microenvironment[J]. Immunol Rev, 2022, 306(1): 123-138.

[16] Kanchanawong P. Focal adhesion signaling in lung cancer invasion and metastasis[J]. Essays Biochem, 2023, 67(3): 245-256.

[17] Zhao X, Wang Y, Zhang L, et al. Antifolate resistance mechanisms in lung squamous cell carcinoma: Implications for precision therapy[J]. Drug Resist Updat, 2023, 64: 100932.

[18] Hynes R O. Integrins and ECM-receptor interactions in tumor progression[J]. Cold Spring Harb Perspect Biol, 2022, 14(9): a040287.

[19] Lee D K, Cheng R, Carbone C. Apelin signaling in lung cancer: From proliferation to immune regulation[J]. Pharmacol Rev, 2023, 75(2): 312-336.