

Identification and experimental validation of lactylation driven prognostic genes in lung adenocarcinoma stemness

Xue-Hong Bai

General Hospital of Ningxia Medical University

Ya-Bo Xu

General Hospital of Ningxia Medical University

Jing-Ying Long

Ningxia Medical University

Li-Jun Ma

Ningxia Medical University

Wei He

442573717@qq.com

General Hospital of Ningxia Medical University

Research Article

Keywords:

Posted Date: August 26th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10608121/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Identification and experimental validation of lactylation driven prognostic genes in lung adenocarcinoma stemness

Xue-Hong Bai¹, Ya-Bo Xu¹, Jing-Ying Long², Li-Jun Ma^{2*}, Wei He^{1*}

¹ General Hospital of Ningxia Medical University (The First Clinical Medical College of Ningxia Medical University), Yinchuan, 750004, Ningxia, China

² School of Basic Medicine, Ningxia Medical University, Yinchuan, 750004, China

* Correspondence:

Corresponding Author : Lijun Ma

E-mail addresses:MLJ2026@hotmail.com

*Corresponding author: Wei He

E-mail addresses:442573717@qq.com

Abstract

Background: Lactylation, a metabolism-linked post-translational modification, has emerged as a potential regulator of cancer stem cell behavior. This study focused on lung adenocarcinoma (LUAD) and sought to identify lactylation-related genes with prognostic value, using the tumor cell stemness index (mRNA_{Si}) as a quantitative indicator of stemness and validating selected targets experimentally.

Methods: Public LUAD transcriptome datasets were analyzed through an integrated bioinformatics pipeline. WGCNA, consensus clustering, differential expression analysis, machine-learning screening, and multivariate regression were combined to identify candidate prognostic factors and establish a risk-score model, followed by internal and external validation. Functional annotation, immune microenvironment assessment, single-cell RNA sequencing (scRNA-seq), computational molecular docking, and immunohistochemistry were used to explore the biological relevance of the model.

Results: Six prognostic genes were obtained, including KIF23, TOP2A, HSPD1, DTYMK, CAV3, and CYP27A1. The risk signature stratified LUAD patients into high-risk (HRG) and low-risk (LRG) groups, with poorer survival observed in HRG. Gene set enrichment indicated that the 'MYC targets V1' pathway was activated in both HRG and LRG. Immune profiling showed higher tumor purity in HRG.

Single-cell analysis annotated 11 cell populations and revealed higher HSPD1 expression in tumor samples, whereas CYP27A1 was more abundant in normal samples. CERES analysis suggested that LUAD cells showed greater dependency on KIF23, TOP2A, and HSPD1. These genes also showed favorable docking with BRD-K71781559, and IHC supported higher KIF23 and TOP2A protein expression in LUAD tissues.

Conclusion: A six-gene LUAD prognostic panel composed of KIF23, TOP2A, HSPD1, DTYMK, CAV3, and CYP27A1 was developed. This panel may facilitate prognostic stratification and guide future mechanism-based therapeutic exploration.

1. Introduction

Lung adenocarcinoma (LUAD) is a frequent and clinically difficult subtype of lung cancer whose incidence has increased over recent years. Its clinical behavior is marked by invasion, aggressiveness, and limited therapeutic responsiveness[1]. Available treatments include non-invasive surgical resection, chemoradiotherapy, tyrosine kinase inhibitors (TKIs), including gefitinib, anaplastic lymphoma kinase inhibitors, and PD-1/PD-L1-directed immunotherapy[2, 3]. Nevertheless, durable disease control is limited to part of the patient population. The prognosis of LUAD remains unfavorable, with a reported 5-year overall survival (OS) rate of about 17.4%[1]. Identification of additional prognostic genes is therefore important for refined risk evaluation and individualized treatment planning.

Tumor stem cells (TSCs) are central to LUAD biology because they are involved in tumor initiation, metastatic dissemination, and persistent disease. Through self-renewal and multilineage differentiation, these cells sustain tumor expansion[4, 5]. They may also resist cell death, chemotherapy, and radiotherapy, thereby facilitating distant spread[5]. For this reason, TSC-related features are both therapeutically relevant and prognostically meaningful in invasive tumors[6, 7]. A small cancer stem cell (CSC) compartment may drive tumor progression. The mRNA-based stemness index (mRNAsi), generated using one-class logistic regression (OCLR), measures transcriptomic stemness and represents CSC-related activity. Increased mRNAsi values have been linked to pathological grade and features reflecting CSC activity and tumor differentiation[8].

Lactic acid is no longer viewed only as an end product of metabolism. It also functions as a signaling mediator in pathological processes such as inflammation and malignancy[3, 9]. In tumors, enhanced glycolysis and lactate accumulation can support CSC self-renewal and maintenance of stemness[10]. Previous work in liver cancer stem cells (LCSCs) reported greater glycolytic activity, lactate accumulation, and lactylation than in hepatocellular carcinoma (HCC) cells[10]. However, the regulatory role of lactylation in tumor stemness, especially in lung cancer cells, is still not fully clarified. Therefore, this study investigated lactylation-associated regulation of LUAD stemness and plasticity to identify candidate therapeutic targets for LUAD. Transcriptomic and clinical data were collected from the Cancer Genome Atlas (TCGA, <http://portal.gdc.cancer.gov/>) and Gene Expression Omnibus (GEO, <https://www.ncbi.nlm.nih.gov/geo/>) resources. An integrated bioinformatics strategy was used to screen mRNAsi-associated lactylation-related prognostic genes in LUAD. Based on these genes, we constructed and evaluated a prognostic model and classified patients into high-risk (HRG) and low-risk (LRG) groups. Nomogram construction, clinical subgroup assessment, immune infiltration analysis, somatic mutation evaluation, and drug sensitivity prediction were then performed to define group-specific features. Finally, single-cell RNA sequencing (scRNA-seq) analysis was used to map prognostic gene expression across cell populations, providing evidence for LUAD prognostic assessment and therapeutic exploration.

Keywords: Lung adenocarcinoma, Lactylation modification, mRNAsi, prognostic genes, Single-cell RNA sequencing

2. Materials and methods

2.1 Data collection

LUAD transcriptomic profiles, clinical annotations, and survival information were downloaded from the UCSC Xena platform and GEO database. In TCGA-LUAD, 576 tumor specimens with complete follow-up information were available. The GSE31210 dataset, generated on the GPL570 platform, provided expression profiles for 226 tissue samples with survival data. The scRNA-seq dataset GSE131907 (GPL16791), also obtained from GEO, included transcriptomic data from 11 LUAD

specimens and 11 matched control samples.

A list of 308 lactylation-associated genes (LAGs) was obtained from MSigDB through the GSEA portal and is provided in Supplementary Table S1.

2.2 Calculation of mRNAsi

For TCGA-LUAD, the OCLR algorithm was applied to build a machine learning-derived mRNAsi model that captured oncogenic dedifferentiation-associated stemness. Spearman correlation was used to relate the mRNAsi model weight vector to TCGA-LUAD expression profiles. The correlation coefficients were subsequently normalized to a 0-1 range by subtracting the minimum value and dividing by the overall range; values closer to 1 represented stronger stemness potential.

The difference in mRNAsi between LUAD tumor tissues and control samples was tested at $P < 0.05$. LUAD cases were then classified into high-mRNAsi and low-mRNAsi groups according to the median mRNAsi value. Kaplan-Meier (K-M) survival curves were plotted with "survminer" (v 0.5.0)[11] to evaluate survival differences between groups ($P < 0.05$).

2.3 WGCNA

In TCGA-LUAD, gene modules associated with mRNAsi were identified using the "WGCNA" package (v 2.27)[12]. Hierarchical clustering was used to evaluate sample outliers before network construction. A soft-thresholding power suitable for scale-free topology was chosen under $R^2 > 0.90$ while maintaining mean connectivity near zero.

The co-expression adjacency matrix was generated with a minimum module size of 50 genes, and each module was assigned a distinct color. Pearson correlations between mRNAsi and modules were calculated ($|\text{cor}| > 0.30$, $P < 0.05$). Modules showing the strongest positive and negative correlations with mRNAsi were considered hub modules. Gene significance (GS) and module membership (MM) were calculated, and genes with $|\text{GS}| > 0.7$ and $|\text{MM}| > 0.5$ were retained as hub module genes for subsequent analyses.

2.4 Consistency clustering analysis

Using key module gene-expression profiles, LUAD patients from TCGA-LUAD were clustered with "ConsensusClusterPlus" (v 1.64.0)[13], and the number of repetitions per clustering (reps) was set to 50. This analysis assigned patients to molecular clusters. Wilcoxon tests compared mRNAsi across clusters ($P < 0.05$), and K-M analysis was used to assess survival differences among clusters ($P < 0.05$).

To identify mRNA^{Asi}-associated lactylation-modified genes, DE-LRGs between clusters were screened with "limma" (v 3.56.2)[14] using $|\log_2FC| \geq 1.0$ and $\text{adj.P} < 0.05$. The biological functions and pathways of these DE-LRGs were then explored by GO and KEGG enrichment analyses using "clusterProfiler" (v 4.8.3)[15], with $P < 0.05$ defined as the significance threshold.

2.4 Machine learning

Three machine learning methods were used to select feature genes from DE-LRGs in TCGA-LUAD. SVM-RFE was run with the "caret" package (v 7.0-1)[16], and 10-fold cross-validation was applied to determine the gene set with the smallest error. XGBoost was then performed with "XGBoost" (v 1.7.8.1)[17]. RF analysis was conducted using "randomForest" (v 4.7-1.2)[18], and genes with an importance score > 5 were retained. Feature genes selected by all three approaches were finally obtained by intersection.

2.5 Constriction and assessment of risk model

In TCGA-LUAD, feature genes were examined by univariate Cox regression using the "survival" package (v 3.8-3)[11], with $P < 0.05$ used to identify prognosis-related genes. A forest plot was generated with "forestplot" (v 3.1.6)[19] to present these results. Multivariate Cox regression was subsequently applied to build the prognostic risk model, and the prognostic genes identified above were incorporated into the model. Each patient-specific risk score was calculated with the following formula:

$$\text{risk score} = \sum_{i=1}^n (\text{coef}_i * X_i)$$

In this formula, "coef" denotes the multivariate Cox regression coefficient of a given prognostic gene, and "X" represents the corresponding gene-expression value.

For model evaluation, 500 TCGA-LUAD patients were stratified into HRG and LRG according to the median risk score.

Risk distribution curves illustrated risk-score stratification, whereas survival-status scatter plots described patient outcomes across the two risk groups. K-M analysis was used to compare survival between HRG and LRG ($P < 0.05$). Time-dependent ROC curves for 1-, 3-, and 5-year survival were constructed with "survivalROC" (v 1.0.3.1)[20] ($\text{AUC} > 0.6$). Expression differences in prognostic genes between HRG and LRG were also evaluated.

Model robustness was further tested in the independent GSE31210 cohort (n = 226) using the same analytical strategy.

2.6 Clinical characteristics and independent prognostic analysis

A heatmap was used to summarize the distribution of molecular clusters and clinical subgroups, including gender, age, stage, M stage, N stage, and T stage, between HRG and LRG LUAD patients.

In TCGA-LUAD, risk scores were combined with clinical variables. Univariate and multivariate Cox regression analyses ($P < 0.05$) were used to determine independent prognostic predictors for LUAD.

A survival-prediction nomogram for LUAD patients was constructed with "rms" (v 6.7-0)[21]. Prediction accuracy was assessed using calibration curves generated by "rms", decision curves produced with "rmda" (v 1.6)[22], and ROC curves plotted with "timeROC" (v 0.4)[23].

2.7 Immune infiltration and immunologic escape analyses

The immune landscapes of HRG and LRG were compared using CIBERSORT, XCELL, TIMER, and EPIC. Spearman or Pearson correlation analysis was then performed to examine associations between risk scores and immune-cell abundance ($|\text{cor}| > 0.30$, $P < 0.05$).

For assessment of the LUAD immune microenvironment, stromal scores, immune scores, ESTIMATE scores, and tumor purity were calculated in TCGA-LUAD HRG and LRG samples with the "ESTAMATE" package (v 1.0.13)[24]. Intergroup differences were tested at $P < 0.05$.

Immune escape-related genes retrieved from the literature[25] were analyzed for correlations with both risk scores and prognostic genes ($|\text{cor}| > 0.30$, $P < 0.05$).

2.8 Drug sensitivity analysis and GSEA

Drug sensitivity in TCGA-LUAD HRG and LRG patients was evaluated by estimating IC50 values for 198 conventional agents with "oncopredict" (v 1.21)[26]. IC50 differences between groups were tested for each drug ($P < 0.05$).

GSEA was conducted in the training cohort to explore biological functions associated with risk status. DEGs between HRG and LRG were identified using "DESeq2" (v 1.40.2)[27] and ranked by log2FC. The "h.all.v2024.1.Hs.symbols.gmt" collection was used as the background gene set, and enrichment was evaluated using $|\text{NES}| > 1$

and $P < 0.05$.

2.9 Single cell expression analysis of prognostic genes

The GSE131907 dataset was analyzed with "Seurat" (v 5.2.1)[14]. During quality control, low-quality cells and genes were removed, including cells with $>20\%$ ribosomal gene expression or $>3\%$ erythrocyte gene expression. The remaining data were normalized with NormalizeData, followed by highly variable gene selection with FindVariableFeatures and PCA using RunPCA. Significant PCs were selected according to JackStrawPlot ($P < 0.05$). Cells were clustered using FindNeighbors and FindClusters and visualized by tSNE. Cell annotations were assigned with the TISCH database (<http://tisch.comp-genomics.org/home/>), after which prognostic gene expression was examined across immune cell subsets.

2.10 Identification of important genes and drug screening via CERES and molecular docking

CERES is a CRISPR knockout-based metric used to quantify gene dependency in cancer cell lines and to identify cancer-associated genetic vulnerabilities[28]. A lower CERES score indicates that the gene is more likely to be essential for the corresponding cancer cell line.

To prioritize essential candidates among the prognostic genes, CERES scores were calculated across 318 LUAD cell lines, and genes with the lowest scores were selected. Candidate antitumor drugs related to these essential genes were then explored. CTRP and PRISM data from the BEST website (<https://bestwebsite.gallery/>) were used to evaluate correlations between important genes and antitumor drugs ($|\text{cor}| > 0.30$, $P < 0.05$). Drugs positively associated with high expression of these genes were considered sensitive candidates.

Molecular docking was then performed for selected therapeutic candidates. Three-dimensional ligand structures were obtained from PubChem, whereas receptor structures were downloaded from RCSB PDB (<http://www.rcsb.org/>). Binding energy was estimated using the CB-Dock web interface. Binding energies < -5 kcal/mol were interpreted as indicating meaningful affinity between the compound and the protein target.

2.11 Immunohistochemical (IHC) analysis

To assess protein-level expression of prognostic genes in LUAD patients, IHC staining was conducted on tissue specimens. Thirty specimens were collected from the General Hospital of Ningxia Medical University. Written informed consent was obtained from all participants, and the protocol was approved by the hospital ethics committee (approval number: KYLL-2023-177).

Paraffin-embedded tissue sections from 15 tumor patients were deparaffinized with xylene, rehydrated through graded ethanol, and processed for antigen retrieval. Endogenous peroxidase activity was inhibited with 3% hydrogen peroxide. Sections were blocked with 10% universal goat serum to minimize nonspecific binding and then incubated with primary antibodies (Supplementary Table 2). Biotinylated anti-rabbit secondary antibodies were applied for primary-antibody detection. Staining was visualized with diaminobenzidine (DAB), followed by hematoxylin counterstaining. Finally, samples were dehydrated, cleared, mounted, and evaluated by whole-film scanning.

2.12 Statistical analysis

All statistical analyses were performed using R software (v 4.3.0). The Wilcoxon test was applied for two-group comparisons, and the Kruskal-Wallis test was used for multi-group comparisons ($P < 0.05$).

3. Results

3.1 The effect of mRNAsi on LUAD

The Wilcoxon test showed that mRNAsi was markedly higher in LUAD tumor samples than in normal tissues ($P < 2 \times 10^{-16}$) (Fig. 1A). K-M analysis further indicated shorter overall survival in the high-mRNAsi group than in the low-mRNAsi group ($P = 0.045$) (Fig. 1B).

No outlier sample was detected in TCGA-LUAD. The optimal power value was 10 (Fig. 2A). After removal of the gray module, 16 gene modules remained (Fig. 2B). MEbrown was strongly and positively correlated with mRNAsi ($\text{cor} = 0.72$, $P = 4 \times 10^{-93}$), whereas MEyellow showed a strong negative correlation ($\text{cor} = -0.81$, $P = 8 \times 10^{-133}$) (Fig. 2C). Applying $|\text{GS}| > 0.7$ and $|\text{MM}| > 0.5$ yielded 2,557 key module genes (Fig. 2D-E).

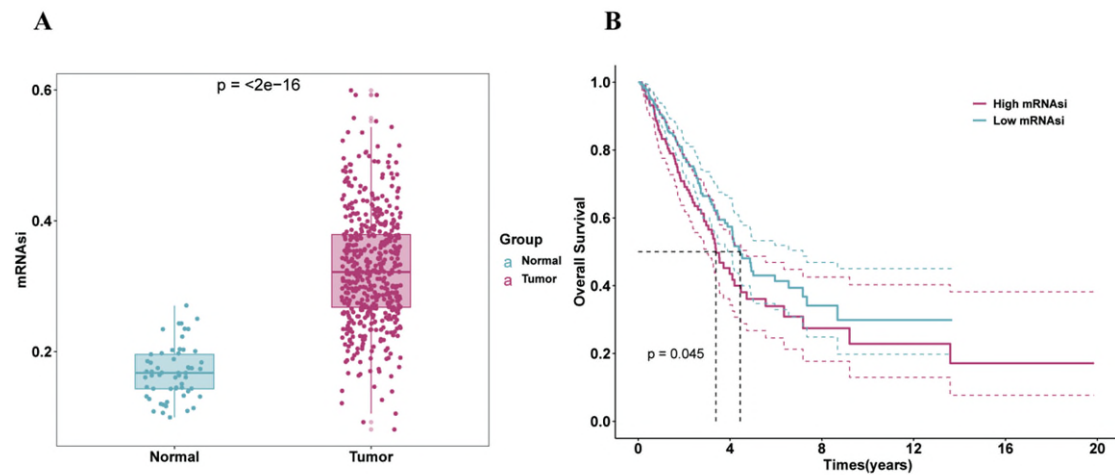


Figure 1. Analysis of the impact of mRNasi on overall survival in LUAD. **(A)** Box plot showing the differences in mRNasi between LUAD tumors and normal samples ($P < 0.05$). **(B)** Kaplan-Meier (K-M) survival analysis comparing overall survival between the high-mRNasi and low-mRNasi groups ($P < 0.05$).

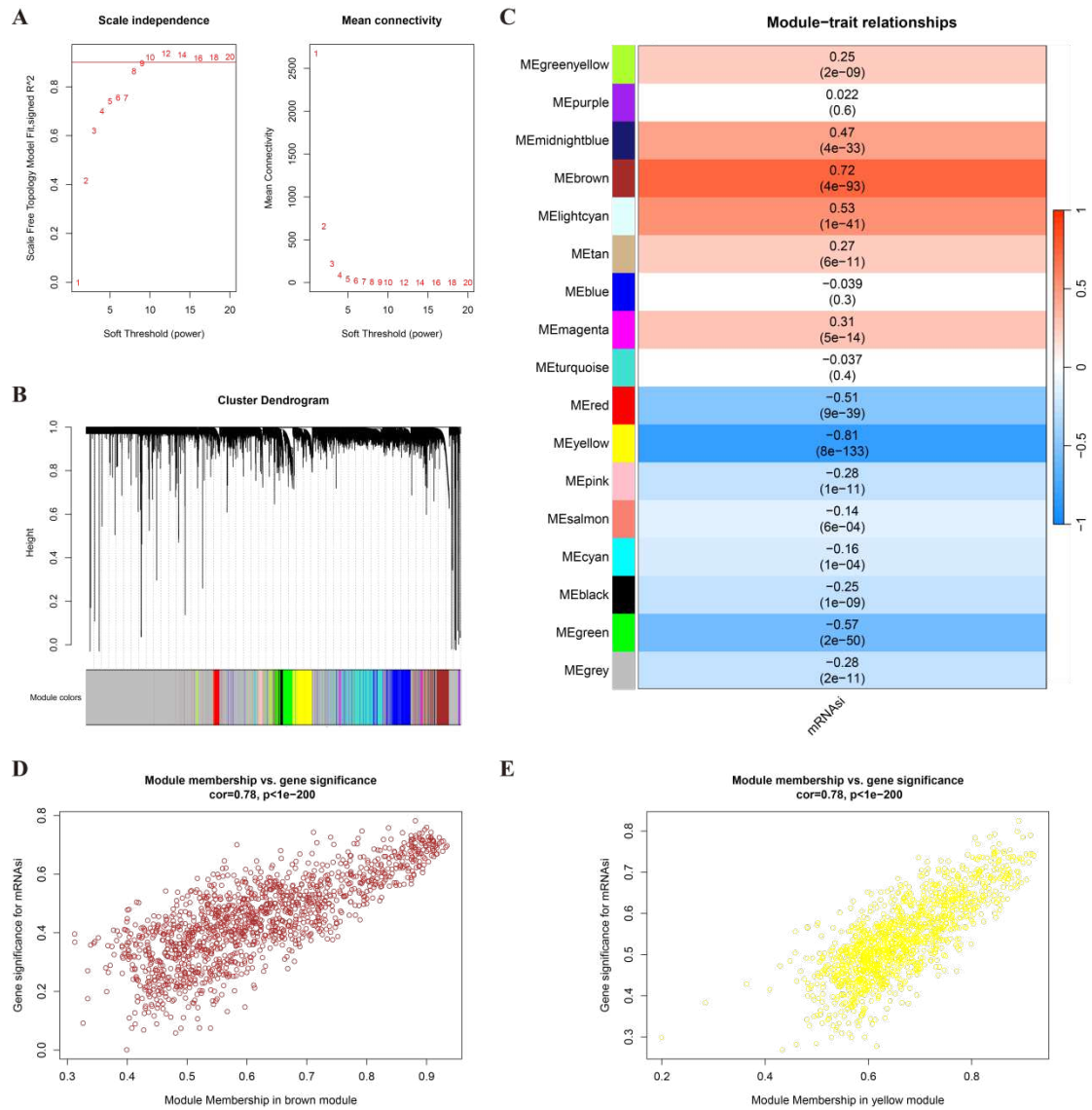


Figure 2. Weighted gene co-expression network analysis (WGCNA). (A) Selection of the optimal soft power for network construction, based on an R^2 value > 0.90 and mean connectivity close to 0. (B) Merged gene modules from the co-expression network, with a total of 16 modules identified (excluding the gray module). (C) Heatmap showing the correlation between gene modules and mRNA_{si} ($|correlation\ coefficient\ (cor)| > 0.30$ and $P < 0.05$). (D-E) Scatter plots depicting the relationship between gene significance (GS) and module membership (MM).

3.2 Identification and functional analyses of 44 DE-LRGs

Consensus clustering separated 509 LUAD patients into two molecular clusters, with 261 cases in cluster 1 and 248 cases in cluster 2 ($K = 2$) (Fig. 3A). Cluster 2 had significantly higher mRNA_{si} than cluster 1 ($P < 2 \times 10^{-16}$) (Fig. 3B). Survival

analysis also differed between clusters; cluster 2 showed poorer OS than cluster 1 ($P = 0.023$) (Fig. 3C).

In cluster 2, 44 DE-LRGs were detected, comprising 25 upregulated and 19 downregulated genes (Fig. 3D). GO enrichment linked these genes to terms such as "positive regulation of T cell mediated immune response to tumor cell" and "sterol catabolic process" (Fig. 3E). KEGG analysis identified enriched pathways including "Pyrimidine metabolism", "Terpenoid backbone biosynthesis", and "p53 signaling pathway" (Fig. 3F).

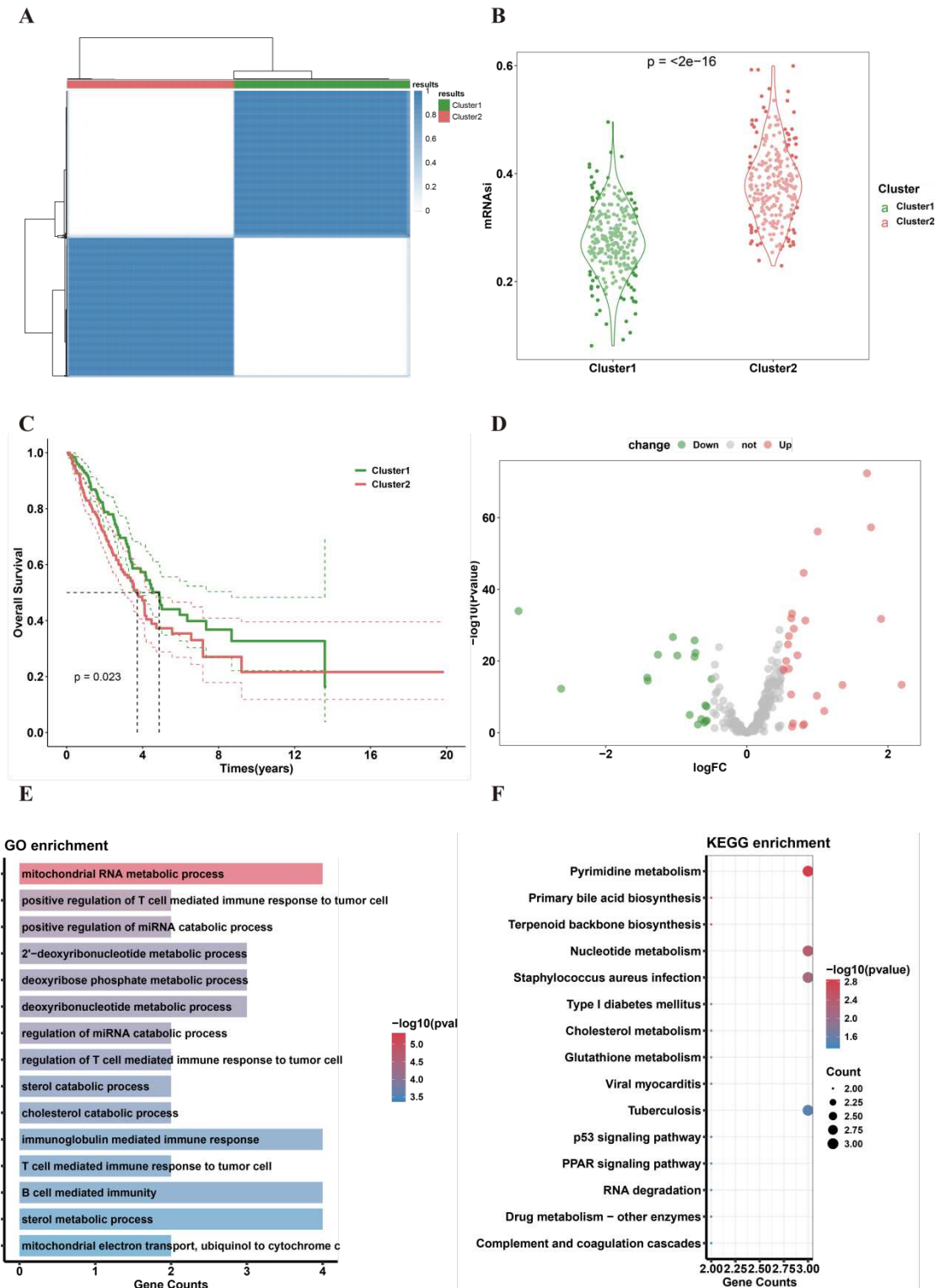


Figure 3. Identification and functional analysis of differentially expressed lactylation-related genes (DE-LRGs). **(A)** Consistency clustering analysis dividing LUAD patients in the TCGA-LUAD dataset into two clusters. **(B)** Comparison of mRNA levels differences between the two clusters. **(C)** Kaplan-Meier (K-M) survival analysis comparing overall survival between the two clusters ($P < 0.05$). **(D)** Volcano

plot of DE-LRGs, with green indicating downregulated genes, red for upregulated genes, and gray for genes with no significant difference. (E) Bar chart showing Gene Ontology (GO) analysis results for DE-LRGs. (F) Bubble plot depicting Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis results for DE-LRGs.

3.3 Construction of well-performing risk model

Among the 44 DE-LRGs, RF selected 18 genes (Fig. 4A-B), SVM selected 22 genes (Fig. 4C), and XGBoost selected 15 genes (Fig. 4D). Intersecting the three algorithm outputs produced 14 feature genes (Fig. 4E).

Six prognostic genes, KIF23, TOP2A, HSPD1, DTYMK, CAV3, and CYP27A1, were then obtained ($P < 0.05$) (Fig. 5A), and these genes were used to construct a multivariate Cox regression-based risk model.

For model evaluation, 500 TCGA-LUAD patients were assigned to HRG (250 samples) or LRG (250 samples) by the median risk score. In GSE31210, the 226 LUAD patients were similarly separated into HRG (113 samples) and LRG (113 samples). HRG patients had lower survival probabilities in both cohorts (Fig. 5B-C) and showed higher mortality with shorter survival time (Fig. 5D). HSPD1, DTYMK, TOP2A, and KIF23 were expressed at higher levels in HRG, whereas CAV3 and CYP27A1 were higher in LRG (Fig. 5D). In TCGA-LUAD, the 1-, 3-, and 5-year AUC values were 0.68, 0.64, and 0.59, respectively (Fig. 5E). In GSE31210, all AUC values were greater than 0.60 (Fig. 5F).

These findings suggest that the risk model can distinguish LUAD patients with different survival risks and may contribute to individualized treatment planning and disease management.

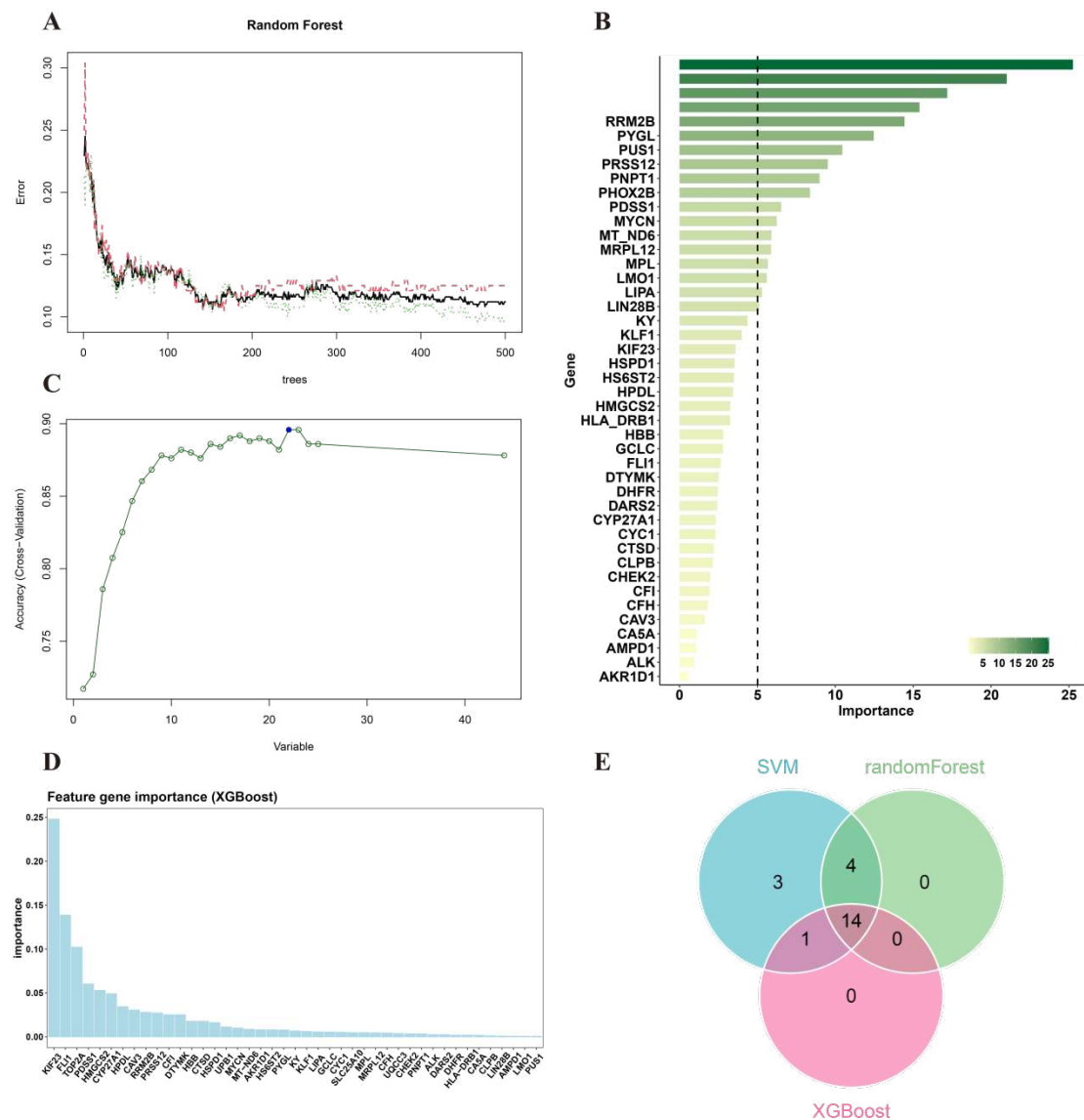


Figure 4. Machine learning analyses. **(A-B)** Results of the Random Forest (RF) algorithm: (A) Error rate and tree visualization; (B) Feature importance ranking. **(C)** Results of the Support Vector Machine-Recursive Feature Elimination (SVM-RFE) analysis. **(D)** Results of the XGBoost algorithm. **(E)** Venn diagram illustrating the overlap of feature selections from RF, SVM-RFE, and XGBoost.

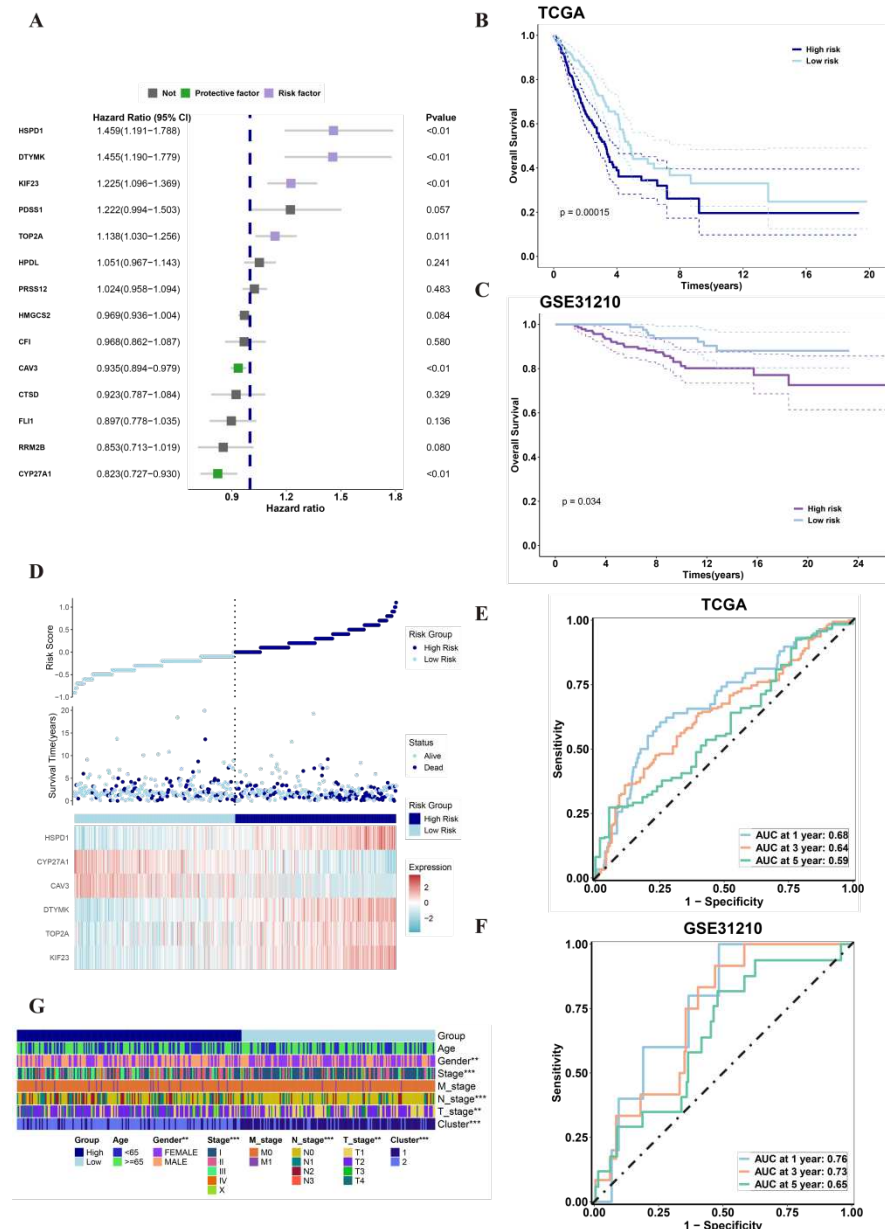


Figure 5. Construction and validation of the risk model. **(A)** Forest plot identifying 6 prognostic genes through univariate Cox regression analysis. **(B-C)** Kaplan-Meier (K-M) survival curves comparing overall survival (OS) between HRG and LRG patients in the **(B)** TCGA-LUAD dataset and **(C)** GSE31210 dataset. **(D)** Risk curves, survival status distribution, and heatmap in the TCGA-LUAD dataset. **(E-F)** 1-year, 3-year, and 5-year Receiver Operating Characteristic (ROC) curves for the risk model in **(E)** TCGA-LUAD dataset and **(F)** GSE31210 dataset. **(G)** Heatmap showing the distribution of various clusters and clinical characteristic subgroups in HRG and LRG patients from the TCGA-LUAD dataset.

3.4 Establishment good nomogram based on independence prognostic factors

The heatmap displayed the distribution of molecular clusters and clinical characteristic subgroups in LUAD patients (Fig. 5G).

T stage and risk score were confirmed as independent prognostic factors for LUAD (Fig. 6A-B). A nomogram was constructed (Fig. 6C), and the AUC values for 1-, 3-, and 5-year prediction were 0.73, 0.68, and 0.59, respectively (Fig. 6D). Calibration curves showed good consistency between predicted and observed outcomes (Fig. 6E). Decision curve analysis indicated that the nomogram produced greater net benefit than either single factor alone (Fig. 6F), supporting its prognostic value in LUAD.

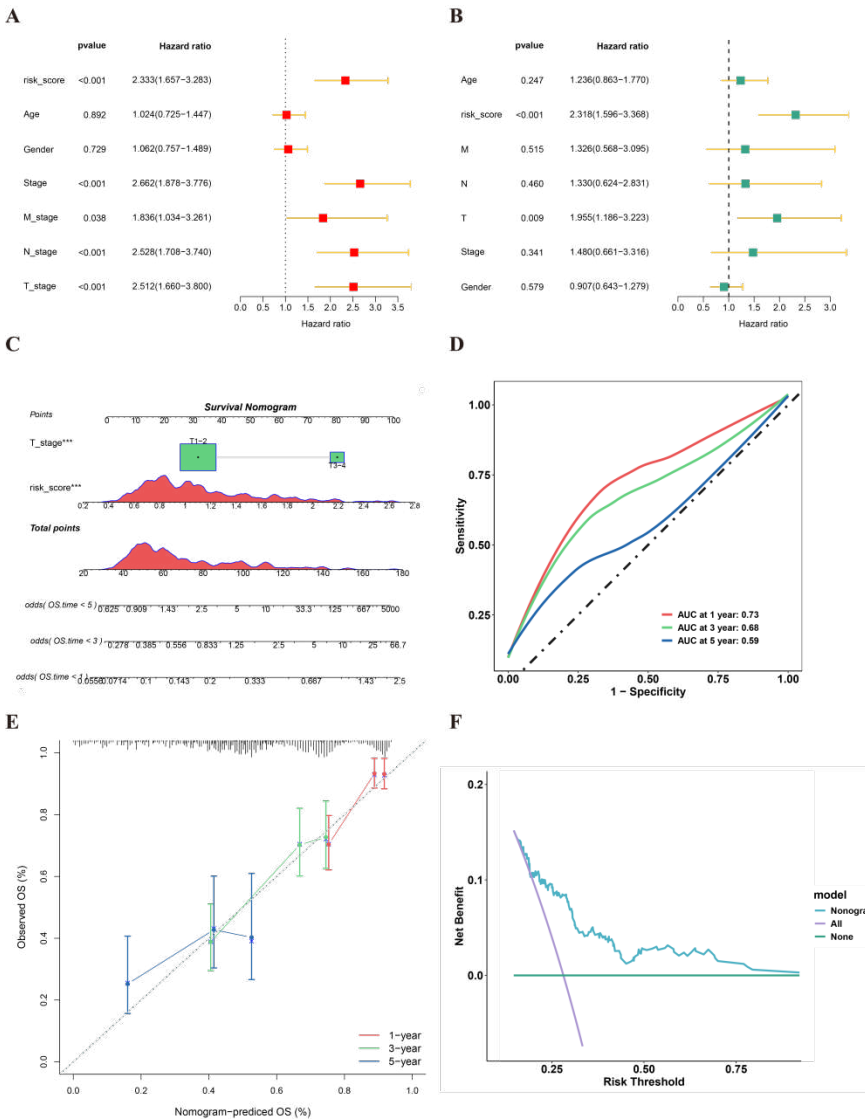


Figure 6. Assessment of the clinical utility of the nomogram. (A-B) Forest plots displaying (A) univariate and (B) multivariate Cox regression analysis of clinical parameters and risk scores. (C) Nomogram predicting the 1-year, 3-year, and 5-year

probabilities of OS outcome. (D) ROC curves for predicting 1-year, 3-year, and 5-year outcomes in LUAD patients. (E) Calibration curve for assessing the accuracy of the nomogram. The gray dashed diagonal line represents an ideal model. (F) Decision curve analysis (DCA) for evaluating the clinical utility of the nomogram model.

3.5 Investigating the immune infiltration, immune escape, and functional pathways of LUAD patients

Risk score showed negative correlations with most infiltrating immune cells, including Fibroblasts, M2 macrophages, and NKT cells, indicating that higher-risk patients may have a more immunosuppressive microenvironment (Fig. 7A). ESTIMATE analysis demonstrated lower stromal, immune, and ESTIMATE scores in HRG ($P < 0.05$), together with higher tumor purity ($P < 0.05$) (Fig. 7B). The top 10 drugs with significant IC50 differences are presented in Fig. 7C. For instance, Doramapimod_1042 had a higher IC50 in HRG than in LRG, suggesting greater sensitivity among patients with lower risk scores. HSPD1, DTYMK, TOP2A, and KIF23 were positively correlated with most immune escape-related genes, including CEP55 and BOLA3, while CAV3 and CYP27A1 were negatively correlated with most of these genes (Fig. 7D). GSEA revealed between-group differences in pathways including "E2F targets", "G2M checkpoint", and "MYC targets V1" (Fig. 7E), suggesting distinct molecular and immune features associated with LUAD progression and immune escape.

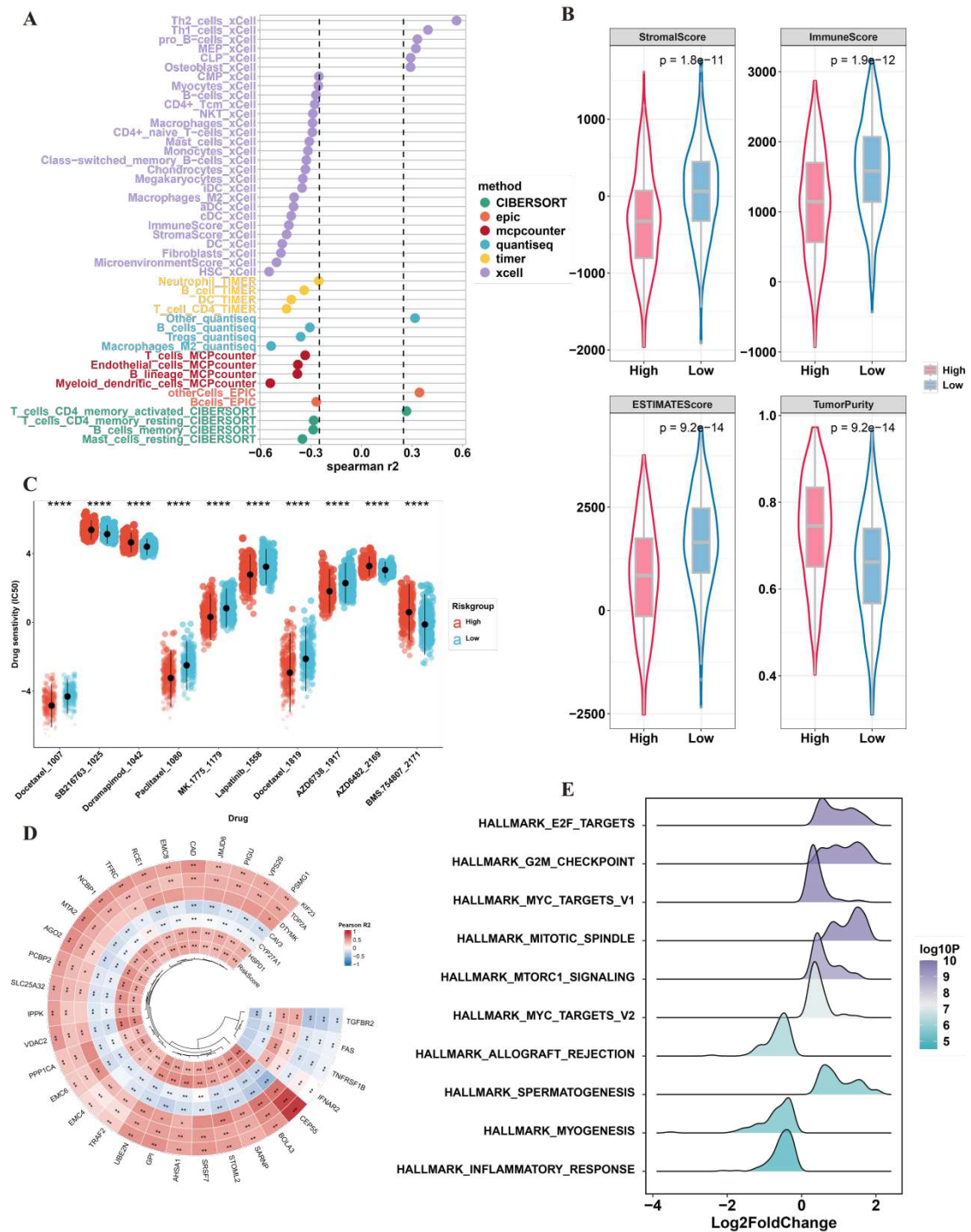


Figure 7. Immune microenvironment and drug sensitivity analysis in LUAD. **(A)** Correlation coefficients between immune cell proportions and risk scores, determined using the CIBERSORT, XCELL, TIMER, and EPIC algorithms. **(B)** ESTIMATE analysis comparing stromal, immune, ESTIMATE scores, and tumor purity between HRG and LRG groups ($P < 0.05$). **(C)** Top 10 significantly differential drugs based on IC50 between HRG and LRG groups ($P < 0.05$). **(D)** Correlation analysis between

immune escape-related genes and prognostic genes, as well as risk scores. (E) Gene set enrichment analysis (GSEA) highlighting significant pathway differences between HRG and LRG groups.

3.6 Exploration the expression of prognostic genes in single-cell level

The scRNA-seq analysis identified 11 annotated cell types, including B cells, CD4Tconv, CD8T, CD8 T exhausted (CD8Tex), dendritic cells (DC), endothelial cells, epithelial cells, fibroblasts, mast cells, monocytes/macrophages, and plasma cells (Fig. 8A). HSPD1 expression was higher in LUAD tumor samples than in normal samples, whereas CYP27A1 was more abundant in normal samples (Fig. 8B). tSNE plots showed the cellular distribution of the six prognostic genes across annotated cell populations (Fig. 8C).

KIF23, TOP2A, and HSPD1 had the lowest CERES scores (Fig. 8D) and were therefore considered important genes. Three antitumor drugs positively associated with high gene expression were selected from two databases: BRD-K71781559, L-168049, and trimebutine (Fig. 8E). Docking analysis showed favorable binding between HSPD1 and BRD-K71781559, with a binding energy of -10.4 kcal/mol and key interactions involving residues M506 and D504 (Fig. 9A-B). TOP2A also interacted with BRD-K71781559, with a binding energy of -10.1 kcal/mol and contacts involving residues R98 and N95 (Fig. 9A, 9C). Additional docking results are shown in Supplementary Figure S1.

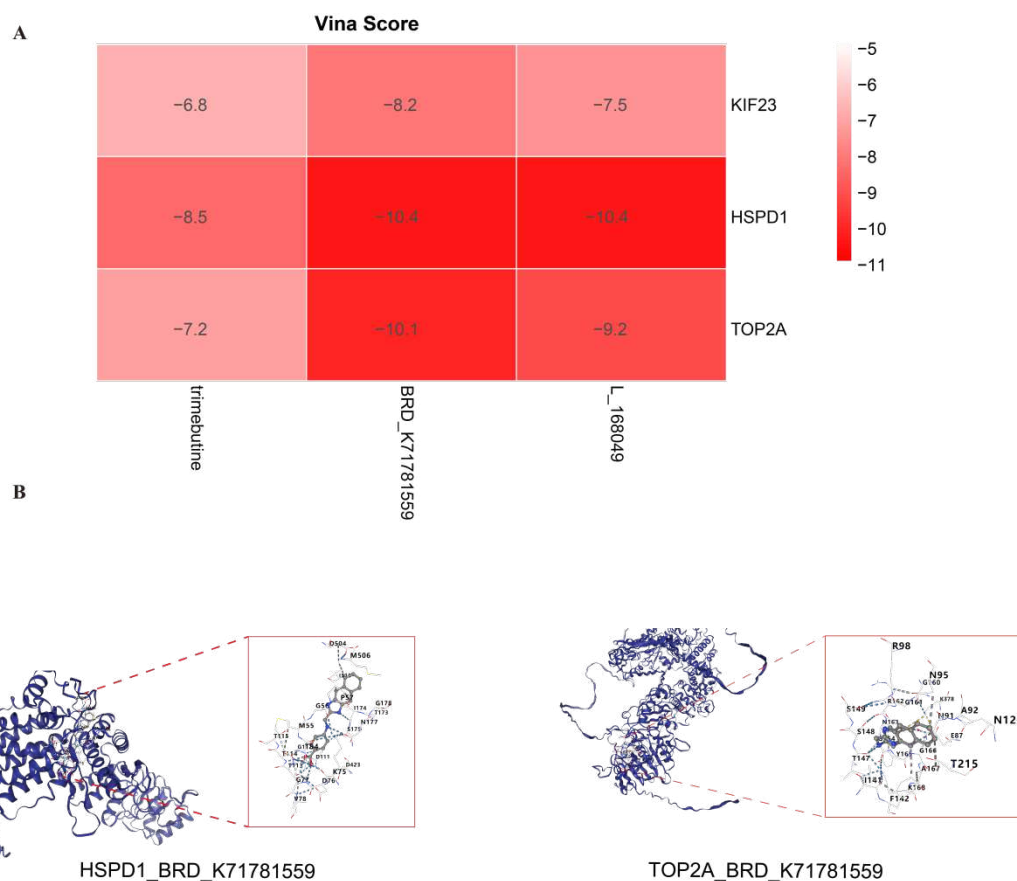


Figure 9. Molecular docking analysis between prognostic genes and antitumor drugs.

(A) Binding energy results between antitumor drugs and prognostic genes. (B) Detailed molecular docking interactions between the prognostic genes and antitumor drugs.

3.7 Exploration the expression of prognostic genes in protein level

IHC staining showed elevated expression of KIF23 and TOP2A in LUAD tissues (Fig. 10).

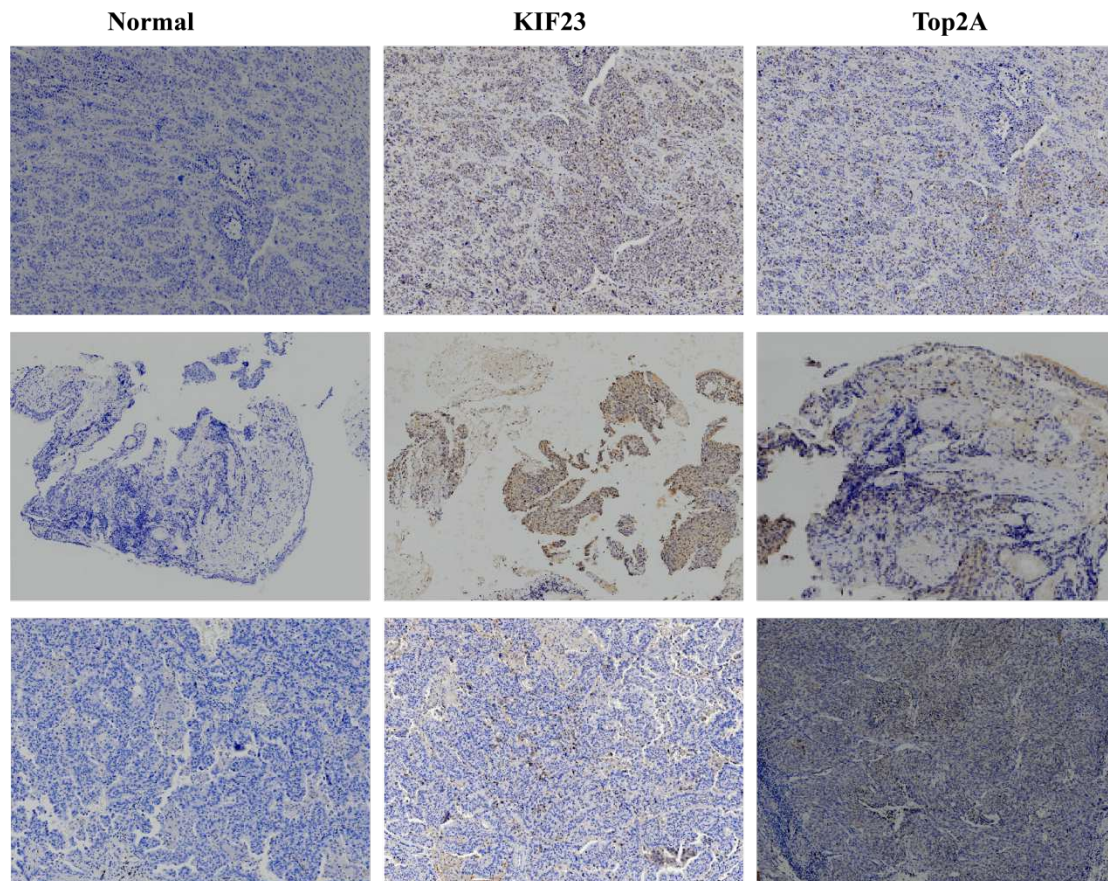


Figure 10. Immunohistochemical analysis of prognostic genes KIF23 and TOP2A.

4. Discussion

LUAD accounts for roughly 40% of lung cancer cases and represents the most frequent malignant histological subtype worldwide. Lung cancer causes approximately 2.1 million deaths annually and contributes to nearly one-fifth of all cancer-related deaths[29]. CSCs are a specialized tumor cell population involved in tumor initiation, progression, establishment, and recurrence[30]. Their heterogeneity, invasive potential, and therapeutic resistance are important drivers of poor prognosis[31]. These properties also help explain the high mortality of cancers such as lung, liver, colorectal, and gastric malignancies[32].

In LUAD, stemness-related gene programs have been shown to influence tumor progression and patient outcome. Using mRNAseq, Zhang et al.[33] reported that CSC-associated expression patterns increased with pathological stage and predicted worse prognosis. Lactylation has also been linked to LUAD biology. For example,

SLC4A7 may restrain tumor progression through regulation of lactate transport, protein lactylation, and stemness-related characteristics, indicating its potential as a therapeutic target in LUAD[34]. These findings collectively support a functional connection between CSC activity, lactylation, and LUAD progression.

How lactylation controls CSC properties in LUAD has not been fully resolved. In the present study, we addressed this issue by analyzing lactylation-associated genes and identifying potential therapeutic targets in LUAD. Six prognostic genes were selected: KIF23, TOP2A, HSPD1, DTYMK, CAV3, and CYP27A1. Compared with low-risk patients, those in the high-risk group had consistently poorer survival and more aggressive clinical features. These findings indicate that the six-gene panel has prognostic stratification value and may reflect biologically meaningful vulnerabilities in LUAD.

Kinesin Family Member 23 (KIF23) is a kinesin superfamily motor protein involved in malignant-cell proliferation and migration. Prior lung cancer studies reported elevated KIF23 expression and linked it to prognosis, suggesting therapeutic relevance[35]. KIF23 overexpression has also been characterized as an independent prognostic factor in lung cancer[35, 36]. In experimental models, RNA interference-mediated KIF23 depletion suppressed lung tumor formation in vivo and induced apoptosis in lung cancer cell lines[37]. KIF23 may also participate in hypoxia-induced tumor cell dehydration within the lung tumor immune microenvironment, and increased expression has been associated with unfavorable prognosis[38]. These observations are consistent with the present results.

Topoisomerase IIA (TOP2A) modulates DNA topology by relieving positive and negative supercoils, thereby supporting chromosome condensation, chromatid segregation, DNA replication, and transcription[39-41]. Aberrant TOP2A overexpression has been reported in multiple malignancies, including pancreatic, breast, prostate, colon, gastroesophageal, esophageal, hepatoblastoma, and malignant schwannoma tumors[42-45]. In LUAD, TOP2A is frequently upregulated and negatively associated with prognosis; high expression correlates with poorer overall and recurrence-free survival[46].

Yang et al. found that TOP2A and KIF23 were associated with disease stage and overall survival in LUAD[47]. In our analysis, both genes were highly expressed in HRG, supporting their association with LUAD progression and adverse prognosis. Single-cell analysis further indicated that KIF23, TOP2A, and HSPD1 had the lowest CERES scores, implying potential dependence of LUAD cells on these genes in cell-cycle regulation and cellular metabolism[48, 49]. The IHC results, which showed increased KIF23 and TOP2A protein expression in LUAD tissues, further support their roles in proliferation, genomic instability, and poor clinical outcome.

Heat shock protein family D member 1 (HSPD1), also known as HSP60, is a molecular chaperone that functions in multiple cellular compartments and extracellular contexts[50]. Increasing evidence implicates HSPD1 in cancer development, progression, and chemoresistance[51, 52]. High HSPD1 expression has been associated with shorter overall survival in LUAD patients, consistent with our findings[53]. Deoxythymidylate kinase (DTYMK) catalyzes phosphorylation of deoxythymidylate (dTMP) to produce deoxythymidine diphosphate (dTDP)[54]. Zhang et al.[8] reported negative associations between DTYMK expression and both recurrence-free survival (RFS) and OS in LUAD, which supports our results. DTYMK was also increased in the high-risk group, emphasizing its prognostic relevance. CYP27A1, a sterol 27-hydroxylase of the cytochrome P450 oxidase family, is mainly expressed in liver tissue and participates in sterol hydroxylation, vitamin D3 metabolism, and cholesterol homeostasis[55, 56]. Yin et al.[57] found that reduced CYP27A1 expression predicted poor LUAD prognosis and was related to immune pathways such as macrophage activation, inflammatory leukocyte activation, and mast cell activation. Our observation that CYP27A1 was lower in high-risk patients is consistent with this evidence. Although CAV3 has been less studied in LUAD, our results suggest that it may have prognostic relevance and merits further investigation.

GSEA indicated enrichment of E2F and MYC target pathways in HRG. Both pathways are tightly linked to cell proliferation[58-60]. Their concurrent activation may indicate rapid tumor-cell division and metabolic dysregulation in high-risk patients. Such pathway activity suggests an active proliferative state and may promote

malignant behaviors, including invasion and metastasis[60-63]. Therefore, pathway activation in HRG may partly explain the aggressive growth pattern of high-risk LUAD.

Immune infiltration analysis showed that risk scores were negatively correlated with most immune-cell populations, suggesting a more immunosuppressive tumor microenvironment in high-risk patients. NKT cells were particularly negatively associated with the risk score. These cells contribute to antitumor immunity by recognizing and killing tumor cells and by coordinating immune responses[64, 65]. In high-risk patients, reduced abundance or impaired activity of NKT cells may weaken tumor immune recognition[66], promote immune escape, and facilitate LUAD progression.

Through integrated bioinformatics analysis, this study identified six prognostic biomarkers: KIF23, TOP2A, HSPD1, DTYMK, CAV3, and CYP27A1. IHC staining supported protein-level expression of KIF23 and TOP2A in clinical samples. The six-gene risk signature separated LUAD patients into distinct prognostic categories, with markedly worse survival in the high-risk group. Beyond prognostic modeling, immune microenvironment assessment, drug-response prediction, and GSEA provided complementary evidence for precision LUAD management. The addition of scRNA-seq analysis further helped characterize cellular heterogeneity associated with tumor progression.

Several limitations remain. First, the relatively modest cohort size may affect the generalizability of the results. Second, functional validation in vivo, particularly animal experiments, was not performed. Future work should include larger cohorts and mechanistic experiments to confirm these findings.

References

1. Denisenko TV, Budkevich IN, Zhivotovsky B. Cell death-based treatment of lung adenocarcinoma. *Cell death & disease*. 2018; 9: 117.
2. Wen G, Niu S, Mei S, Wang S. The cancer stem cells characteristics analysis of LGR5 + cells that influence lung cancer risk by using single cell RNA-seq analysis.

Scientific reports. 2025; 15: 16085.

3. Xu L, Li K, Li J, Liu L, Xu F, Xu Y, et al. MiR-21/Sonic Hedgehog (SHH)/PI3K/AKT Pathway is Associated with NSCLC of Primary EGFR-TKI Resistance. *Oncologie*. 2022; 24: 579--90.

4. Wang X, Chen Y, Wang X, Tian H, Wang Y, Jin J, et al. Stem Cell Factor SOX2 Confers Ferroptosis Resistance in Lung Cancer via Upregulation of SLC7A11. *Cancer research*. 2021; 81: 5217-29.

5. Guo L, Mohanty A, Singhal S, Srivastava S, Nam A, Warden C, et al. Targeting ITGB4/SOX2-driven lung cancer stem cells using proteasome inhibitors. *iScience*. 2023; 26: 107302.

6. Masciale V, Banchelli F, Grisendi G, Samarelli AV, Raineri G, Rossi T, et al. The molecular features of lung cancer stem cells in dedifferentiation process-driven epigenetic alterations. *The Journal of biological chemistry*. 2024; 300: 107994.

7. Pan Z, Liu H, Chen J. [Lung Cancer Stem-like Cells and Drug Resistance]. *Zhongguo fei ai za zhi = Chinese journal of lung cancer*. 2022; 25: 111-7.

8. Zhang Y, Wang H, Liu Y, Yang J, Zuo X, Dong M, et al. Comprehensive analysis of DTYMK in pan-cancer and verification in lung adenocarcinoma. *Bioscience reports*. 2022; 42.

9. Liu X, Zhang Y, Li W, Zhou X. Lactylation, an emerging hallmark of metabolic reprogramming: Current progress and open challenges. *Frontiers in cell and developmental biology*. 2022; 10: 972020.

10. Feng F, Wu J, Chi Q, Wang S, Liu W, Yang L, et al. Lactylome Analysis Unveils Lactylation-Dependent Mechanisms of Stemness Remodeling in the Liver Cancer Stem Cells. *Advanced science (Weinheim, Baden-Wurttemberg, Germany)*. 2024; 11: e2405975.

11. Shi Y, Wang Y, Dong H, Niu K, Zhang W, Feng K, et al. Crosstalk of ferroptosis regulators and tumor immunity in pancreatic adenocarcinoma: novel perspective to mRNA vaccines and personalized immunotherapy. *Apoptosis : an international journal on programmed cell death*. 2023; 28: 1423-35.

12. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC bioinformatics*. 2008; 9: 559.

13. Shu J, Jiang J, Zhao G. Identification of novel gene signature for lung adenocarcinoma by machine learning to predict immunotherapy and prognosis. *Frontiers in immunology*. 2023; 14: 1177847.

14. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic acids research*. 2015; 43: e47.

15. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics : a journal of integrative biology*. 2012; 16: 284-7.

16. Mahmoudian M, Venäläinen MS, Klén R, Elo LL. Stable Iterative Variable Selection. *Bioinformatics (Oxford, England)*. 2021; 37: 4810-7.

17. Kui B, Pintér J, Molontay R, Nagy M, Farkas N, Gede N, et al. EASY-APP: An artificial intelligence model and application for early and easy prediction of severity

in acute pancreatitis. *Clinical and translational medicine*. 2022; 12: e842.

18. Wang Y, Zhuang H, Jiang XH, Zou RH, Wang HY, Fan ZN. Unveiling the key genes, environmental toxins, and drug exposures in modulating the severity of ulcerative colitis: a comprehensive analysis. *Frontiers in immunology*. 2023; 14: 1162458.
19. Fang Y, Huang S, Han L, Wang S, Xiong B. Comprehensive Analysis of Peritoneal Metastasis Sequencing Data to Identify LINC00924 as a Prognostic Biomarker in Gastric Cancer. *Cancer management and research*. 2021; 13: 5599-611.
20. Zhang S, Sun L, Cai D, Liu G, Jiang D, Yin J, et al. Development and Validation of PET/CT-Based Nomogram for Preoperative Prediction of Lymph Node Status in Esophageal Squamous Cell Carcinoma. *Annals of surgical oncology*. 2023; 30: 7452-60.
21. Wang H, Li Q, Wang H, Song W. Construction and validation of a line chart for gestational diabetes mellitus based on clinical indicators. *Lipids in health and disease*. 2024; 23: 349.
22. Liu C, He Y, Luo J. Application of Chest CT Imaging Feature Model in Distinguishing Squamous Cell Carcinoma and Adenocarcinoma of the Lung. *Cancer management and research*. 2024; 16: 547-57.
23. Lu L, Wang H, Fang J, Zheng J, Liu B, Xia L, et al. Overexpression of OAS1 Is Correlated With Poor Prognosis in Pancreatic Cancer. *Frontiers in oncology*. 2022; 12: 944194.
24. Yoshihara K, Shahmoradgoli M, Martínez E, Vegesna R, Kim H, Torres-Garcia W, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nature communications*. 2013; 4: 2612.
25. Lawson KA, Sousa CM, Zhang X, Kim E, Akthar R, Caumanns JJ, et al. Functional genomic landscape of cancer-intrinsic evasion of killing by T cells. *Nature*. 2020; 586: 120-6.
26. Maeser D, Gruener RF, Huang RS. oncoPredict: an R package for predicting in vivo or cancer patient drug response and biomarkers from cell line screening data. *Briefings in bioinformatics*. 2021; 22.
27. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology*. 2014; 15: 550.
28. Meyers RM, Bryan JG, McFarland JM, Weir BA, Sizemore AE, Xu H, et al. Computational correction of copy number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells. *Nature genetics*. 2017; 49: 1779-84.
29. Zha L, Matsu-Ura T, Sluka JP, Murakawa T, Tsuta K. Morphological basis of the lung adenocarcinoma subtypes. *iScience*. 2024; 27: 109742.
30. Capp JP. Cancer Stem Cells: From Historical Roots to a New Perspective. *Journal of oncology*. 2019; 2019: 5189232.
31. Kuşoğlu A, Biray Avcı Ç. Cancer stem cells: A brief review of the current status. *Gene*. 2019; 681: 80-5.
32. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians*. 2018; 68: 394-424.

33. Zhang Y, Tseng JT, Lien IC, Li F, Wu W, Li H. mRNAsi Index: Machine Learning in Mining Lung Adenocarcinoma Stem Cell Biomarkers. *Genes*. 2020; 11.
34. Yan H, He Q, Gao Y, He X, Luo H, Shao L, et al. SLC4A7 suppresses lung adenocarcinoma oncogenesis by reducing lactate transport and protein lactylation. *International journal of oncology*. 2025; 66.
35. Kato T, Wada H, Patel P, Hu HP, Lee D, Ujiie H, et al. Overexpression of KIF23 predicts clinical outcome in primary lung cancer patients. *Lung cancer (Amsterdam, Netherlands)*. 2016; 92: 53-61.
36. Vikberg AL, Vooder T, Lokk K, Annilo T, Golovleva I. Mutation analysis and copy number alterations of KIF23 in non-small-cell lung cancer exhibiting KIF23 over-expression. *OncoTargets and therapy*. 2017; 10: 4969-79.
37. Iltzsche F, Simon K, Stopp S, Pattschull G, Francke S, Wolter P, et al. An important role for Myb-MuvB and its target gene KIF23 in a mouse model of lung adenocarcinoma. *Oncogene*. 2017; 36: 110-21.
38. Guo L, Li H, Li W, Tang J. Construction and investigation of a combined hypoxia and stemness index lncRNA-associated ceRNA regulatory network in lung adenocarcinoma. *BMC medical genomics*. 2020; 13: 166.
39. Nitiss JL. DNA topoisomerase II and its growing repertoire of biological functions. *Nature reviews Cancer*. 2009; 9: 327-37.
40. Wang JC. Cellular roles of DNA topoisomerases: a molecular perspective. *Nature reviews Molecular cell biology*. 2002; 3: 430-40.
41. Thakurela S, Garding A, Jung J, Schübeler D, Burger L, Tiwari VK. Gene regulation and priming by topoisomerase II α in embryonic stem cells. *Nature communications*. 2013; 4: 2478.
42. Heestand GM, Schwaederle M, Gatalica Z, Arguello D, Kurzrock R. Topoisomerase expression and amplification in solid tumours: Analysis of 24,262 patients. *European journal of cancer (Oxford, England : 1990)*. 2017; 83: 80-7.
43. Hooks KB, Audoux J, Fazli H, Lesjean S, Ernault T, Dugot-Senant N, et al. New insights into diagnosis and therapeutic options for proliferative hepatoblastoma. *Hepatology (Baltimore, Md)*. 2018; 68: 89-102.
44. Kolberg M, Høland M, Lind GE, Ågesen TH, Skotheim RI, Hall KS, et al. Protein expression of BIRC5, TK1, and TOP2A in malignant peripheral nerve sheath tumours--A prognostic test after surgical resection. *Molecular oncology*. 2015; 9: 1129-39.
45. Pei YF, Yin XM, Liu XQ. TOP2A induces malignant character of pancreatic cancer through activating β -catenin signaling pathway. *Biochimica et biophysica acta Molecular basis of disease*. 2018; 1864: 197-207.
46. Kou F, Sun H, Wu L, Li B, Zhang B, Wang X, et al. TOP2A Promotes Lung Adenocarcinoma Cells' Malignant Progression and Predicts Poor Prognosis in Lung Adenocarcinoma. *Journal of Cancer*. 2020; 11: 2496-508.
47. Yang X, Feng Q, Jing J, Yan J, Zeng Z, Zheng H, et al. Identification of differentially expressed genes associated with lung adenocarcinoma via bioinformatics analysis. *General physiology and biophysics*. 2021; 40: 31-48.
48. Zhang J, Zhang H, Wang Y, Wang Q. MCM2-7 in Clear Cell Renal Cell

- Carcinoma: MCM7 Promotes Tumor Cell Proliferation. *Frontiers in oncology*. 2021; 11: 782755.
49. Zhang S, Yu S, Wang J, Cheng Z. Identification of PSMD7 as a prognostic factor correlated with immune infiltration in head and neck squamous cell carcinoma. *Bioscience reports*. 2021; 41.
50. Aluksanasuwan S, Sueksakit K, Fong-Ngern K, Thongboonkerd V. Role of HSP60 (HSPD1) in diabetes-induced renal tubular dysfunction: regulation of intracellular protein aggregation, ATP production, and oxidative stress. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*. 2017; 31: 2157-67.
51. Yun CW, Kim HJ, Lim JH, Lee SH. Heat Shock Proteins: Agents of Cancer Development and Therapeutic Targets in Anti-Cancer Therapy. *Cells*. 2019; 9.
52. Fucarino A, Pitruzzella A. Role of HSP60/HSP10 in Lung Cancer: Simple Biomarkers or Leading Actors? *Journal of oncology*. 2020; 2020: 4701868.
53. Aluksanasuwan S, Somsuan K, Ngoenkam J, Chutipongtanate S, Pongcharoen S. Potential association of HSPD1 with dysregulations in ribosome biogenesis and immune cell infiltration in lung adenocarcinoma: An integrated bioinformatic approach. *Cancer biomarkers : section A of Disease markers*. 2024; 39: 155-70.
54. Lam CW, Yeung WL, Ling TK, Wong KC, Law CY. Deoxythymidylate kinase, DTYMK, is a novel gene for mitochondrial DNA depletion syndrome. *Clinica chimica acta; international journal of clinical chemistry*. 2019; 496: 93-9.
55. Gottfried E, Rehli M, Hahn J, Holler E, Andreesen R, Kreutz M. Monocyte-derived cells express CYP27A1 and convert vitamin D3 into its active metabolite. *Biochemical and biophysical research communications*. 2006; 349: 209-13.
56. Sawada N, Sakaki T, Ohta M, Inouye K. Metabolism of vitamin D(3) by human CYP27A1. *Biochemical and biophysical research communications*. 2000; 273: 977-84.
57. Yin Y, He M, Huang Y, Xie X. Transcriptomic analysis identifies CYP27A1 as a diagnostic marker for the prognosis and immunity in lung adenocarcinoma. *BMC immunology*. 2023; 24: 37.
58. Llombart V, Mansour MR. Therapeutic targeting of "undruggable" MYC. *EBioMedicine*. 2022; 75: 103756.
59. Du J, Liao W, Wang H, Hou G, Liao M, Xu L, et al. MDIG-mediated H3K9me3 demethylation upregulates Myc by activating OTX2 and facilitates liver regeneration. *Signal transduction and targeted therapy*. 2023; 8: 351.
60. Chida K, Oshi M, Roy AM, Yachi T, Nara M, Yamada K, et al. E2F target score is associated with cell proliferation and survival of patients with hepatocellular carcinoma. *Surgery*. 2023; 174: 307-14.
61. Park SM, Choi EY, Bae DH, Sohn HA, Kim SY, Kim YJ. The LncRNA EPEL Promotes Lung Cancer Cell Proliferation Through E2F Target Activation. *Cellular physiology and biochemistry : international journal of experimental cellular physiology, biochemistry, and pharmacology*. 2018; 45: 1270-83.
62. Yıldırım-Buharalıoğlu G. Lysine Demethylase 6B Regulates Prostate Cancer Cell

Proliferation by Controlling c-MYC Expression. *Molecular pharmacology*. 2022; 101: 106-19.

63. Harrington CT, Sotillo E, Dang CV, Thomas-Tikhonenko A. Tilting MYC toward cancer cell death. *Trends in cancer*. 2021; 7: 982-94.

64. Bae EA, Seo H, Kim IK, Jeon I, Kang CY. Roles of NKT cells in cancer immunotherapy. *Archives of pharmacal research*. 2019; 42: 543-8.

65. Burks J, Olkhanud PB, Berzofsky JA. The role of NKT cells in gastrointestinal cancers. *Oncoimmunology*. 2022; 11: 2009666.

66. Song P, Li W, Guo L, Ying J, Gao S, He J. Identification and Validation of a Novel Signature Based on NK Cell Marker Genes to Predict Prognosis and Immunotherapy Response in Lung Adenocarcinoma by Integrated Analysis of Single-Cell and Bulk RNA-Sequencing. *Frontiers in immunology*. 2022; 13: 850745.

Clinical trial number

Not applicable

Acknowledgements

We extend our thanks to all colleagues who have provided assistance and support throughout the research process. Their collaboration and encouragement have been invaluable to our work.

Author Contributions

Conceptualization, Li-jun Ma.; methodology, Ya-Bo Xu .and Xue-Hong Bai; software, Jing-Ying Long.; validation, Li-jun Ma and Wei He; formal analysis, Ya-Bo Xu.; investigation, Li-jun Ma; resources, Wei He; data curation, Wei He.; writing—original draft preparation, Xue-Hong Bai.; writing—review and editing, Xue-Hong Bai. and Wei He; visualization, Xue-Hong Bai.; supervision, Wei He.; project administration, Li-jun Ma; funding acquisition, Li-jun Ma. And Wei He. All authors have read and agreed to the published version of the manuscript.

Funding statement

This work was supported by the Ningxia Natural Science Foundation (Grant NO. 2024AAC03686, 2026AAC030241), the Key research project of Ningxia Natural

Science Foundation (NO.2023BEG03032, 2026BEG02030), scientific research project of Ningxia Medical University (Special talent start project, NO. XT2023004) and the Open Research Fund of Key Laboratory of Endocrine Glucose & Lipids Metabolism and Brain Aging, Ministry of Education (Open Research Fund NO. 2024JYBZDSYS013).

Conflict of interest statement

The authors declare that they have no conflicts of interest with respect to the research, authorship, and/or publication of this article.

Data Availability

The datasets analyzed during the current study are available in the following public repositories: TCGA-LUAD (<https://portal.gdc.cancer.gov/projects/TCGA-LUAD>), GSE31210 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE31210>), GSE131907 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE131907>), MSigDB (<https://www.gsea-msigdb.org/gsea/msigdb>), TISCH (<http://tisch.comp-genomics.org>), DepMap including CERES dataset (<https://depmap.org/portal/download/>), PubChem (<https://pubchem.ncbi.nlm.nih.gov/>), and RCSB PDB (<https://www.rcsb.org/>).

Ethics approval and consent to participate

This study was approved by the Ethics Committee of the General Hospital of Ningxia Medical University (approval number: KYLL-2023-177). All experimental protocols were conducted in accordance with the institutional guidelines and regulations. All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical

736 standards.

737 **Consent to publish**

738 Written informed consent for publication was obtained from all participants.

739

740 **Supplementary Materials**

741 **Table S1.** Gene list of 308 Lactylation-related genes (LRGs).

742 **Table S2.** Primary antibodies used in Immunohistochemical (IHC) analysis.

743 **Figure S1.** Results of molecular docking.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryTableS1.csv](#)
- [SupplementaryTableS2.docx](#)