

RESEARCH ARTICLE

Development of six immune-related lncRNA signature prognostic model for smoking-positive lung adenocarcinoma

Dajie Zhou¹  | Jing Wang¹  | Xiangdong Liu²

¹Department of Clinical Laboratory Center, Yantai Yuhuangding Hospital, Yantai, China

²Department of Clinical Laboratory, Shandong Provincial Hospital Affiliated to Shandong First Medical University, Jinan, China

Correspondence

XiangDong Liu, Department of Clinical Laboratory, Shandong Provincial Hospital Affiliated to Shandong First Medical University, Jinan 250021, Shandong, China.

Email: x_liu@126.com

Funding information

Shandong Province Key R&D Program, Grant/Award Number: 2019GSF108201

Abstract

Background: Smoking is one of the most hazardous risk factors for the development of lung adenocarcinoma (LUAD). Many survival and prognosis-related biomarkers were discovered using database mining. However, the precision of immune-related long noncoding RNAs (lncRNAs) predictions is insufficient. We identified a novel signature to improve the estimate of smoking-related LUAD prognosis.

Methods: The Cancer Genome Atlas database (TCGA) was used to obtain the LUAD lncRNA expression profiles. The smoking-related LUAD cohort was randomly split into discovery and validation cohorts. To determine the risk score, use the LASSO Cox regression technique on the prognostic immune-related lncRNA. The risk signature has been developed.

Results: A total of 643 immune-related lncRNAs were identified as potential candidates for a risk signature. Finally, six immune-related lncRNAs (AL359915.2, AP000695.1, HSPC324, TGFB2-AS1, AC026355.1, and AC002128.2) were identified and used to carry out risk signature, which showed a close association with overall survival in the discovery cohort. We classified patients as high risk or low risk based on a median risk score of 1.0783. In the discovery cohort, overall survival was marginally longer in the low-risk group than in the high-risk category ($p = 2.28e08$). The area under the curves (AUC) for 1-, 3-, and 5-year survival was 0.67, 0.7, and 0.82, respectively. Furthermore, we successfully validated and combined cohorts using this risk profile. We discovered a strong positive connection between HSPC324 and VIPR1 as a possible novel biomarker for smoking-related LUAD development in our study.

Conclusions: Our research has established a six immune-lncRNA signature that may be used to predict the prognosis of smoking-related LUAD with great accuracy.

KEYWORDS

immune, lncRNA, prognostic biomarker, smoking-positive lung adenocarcinoma, TCGA

Dajie Zhou and Jing Wang contributed equally to this work.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2022 The Authors. *Journal of Clinical Laboratory Analysis* published by Wiley Periodicals LLC.

1 | INTRODUCTION

Lung cancer is becoming the most common cancer in the world. With 2.1 million diagnosed persons and 1.8 million fatalities, lung cancer accounts for 26 percent of new cancer cases and 47 percent of cancer-related mortality.^{1,2} The 5-year survival rate for lung cancer survivors was typically low (10–20 percent in most nations).³ Smoking is believed to be the most major risk factor for lung cancer. Continued smoking increases the cancer death rate, the risk of second primary malignancies, and the side effects of therapy.⁴

lncRNAs are noncoding RNAs having a length more than 200 nucleotides. The irregular expression of some lncRNAs is emerging as a significant component of cancer development due to their critical function in carcinogenesis and cancer proliferation.^{5,6} Many studies have shown that lncRNAs have tumorigenic value, including lung cancer.^{7,8} However, the remaining lncRNA markers for LUAD prognosis must be refined further. Many systems biology techniques have been developed in order to categorize lncRNA biomarkers and build lncRNA signatures.^{9,10}

According to new studies, the immune system plays an important role in cancer beginning and progression.¹¹ Furthermore, various studies on lncRNAs show that they play an important role in cancer immunity, such as inhibiting metastasis of tumors.¹² As a result, new immune-related triggers must be established in order to enhance the generation of anticancer immunotherapy.

Biomarkers that may effectively predict cancer prognosis and patient survival aid in tumor treatment. We used the TCGA gene expression profiles and clinical data to create a prognostic and predictive immune-related lncRNA prognostic signature for smoking-related LUAD. Finally, a six-immune-related lncRNA profile linked with smoking-related LUAD pathogenesis, total survival, and recurrence prediction was developed and confirmed.

2 | METHODS

2.1 | Data collection and processing

We obtained the "OncoSG, Nat Genet 2020" data collection from the cBioportal database,¹³ divided it into a smoking and nonsmoking group, and examined the influence of smoking on patient survival. Meanwhile, gene expression profiles and clinical-pathological features of lung adenocarcinoma patients were obtained from TCGA (<https://tcga-data.nci.nih.gov/tcga/>). We retrieved lncRNA and mRNA expression data from gene expression profiles. We removed samples from patients who did not smoke, and smoking data were not generated. We named the nonsmoking adenocarcinoma study whether the patient had never smoked or had smoked 100 cigarettes in their lives. As smoking-associated adenocarcinoma, samples from previous and current users were mixed together.^{14,15} Table 1 shows the pathological clinic aspects such as survival time, survival status, age, gender, stage, TNM, and smoking history. The patient profile

data and clinical features of lung adenocarcinoma are publicly available and available upon open access.

2.2 | Immune-related lncRNA recognition

We obtained lncRNA expression data from the mRNA expression profile data. Immune-related genes were collected from the Molecular Signatures Database (<http://www.broadinstitute.org/gsea/msigdb/index.jsp>) using the keywords IMMUNE SYSTEM PROCESS (m13664) and IMMUNE RESPONSE (m19817).¹⁶ Following that, we used Pearson's correlation analysis to differentiate immune-related lncRNAs based on the correlation coefficient and *p*-value ($|\text{correlation coefficient}| > 0.6$ and $p < 0.01$). The survival time and status were then coupled with immune-related lncRNA expression data (407 cases).

TABLE 1 Clinic pathological features of 407 smoking-related lung adenocarcinoma (LUAD) patients

	Discovery cohort (n = 285)	Validation cohort (n = 122)
Status		
Dead	196	86
Alive	89	36
Age		
<65	126	58
≥65	159	64
Gender		
Female	149	72
Male	136	50
Path diagnosis		
Lung adenocarcinoma	285	122
Tumor stage		
Stage I/II	226	95
Stage III/IV	53	26
Unknown	6	1
T		
T1/2	243	104
T3/4	42	18
N		
N0/1	245	104
N2/3	36	15
Unknown	4	3
M		
M0	181	75
M1	12	9
Unknown	92	38
Smoking history		
All	285	122

The complete sample of smoking-related lung adenocarcinoma was randomly divided into a computer-generated allocation series based on the discovery cohort ($n = 285$, 70%) and validation cohort ($n = 122$, 30%).¹⁷

2.3 | Prognosis of signature construction dependent on the discovery cohort

We used univariate Cox proportional hazards regression analysis to categorize immune-related lncRNAs that were substantially connected to smoking-related LUAD patient survival using data from the discovery cohort. The LASSO regression approach was then used to select the best lncRNAs. The multivariate Cox proportional hazards regression analysis was used to find the independent prognostic lncRNAs of complete survival. Subsequently, to build up a risk score model, we used independent prognostic lncRNAs. Each patient's risk score was obtained by multiplying the lncRNA expression level by its corresponding coefficient.¹⁸ Based on the median evaluation of the risk scores, we divided the discovery cohort into two groups: high risk and low risk. The total survival rates of high-risk and low-risk groups were compared. To examine the time-dependent prediction value of the risk signature, the "SurvivalROC" program is used to generate the receiver operating characteristic (ROC) area under curve (AUC). Furthermore, the survival variance stratified by clinicopathological characteristics was compared between high-risk and low-risk groups.

2.4 | Prognostic signature confirmation through the use of validation and combine cohorts

Similarly, based on the discovery cohort overhead risk score model, we split the validation and merge cohort into high-risk and low-risk groups. In high-risk and low-risk groups, we compared overall survival and overall survival stratified by clinicopathological characteristics.

2.5 | Analysis of immune cell infiltration

CIBERSORT was used to obtain 22 tumor-infiltrating immune cells (TIICs) gene expression matrices of high-risk and low-risk groups for our investigation. Then, using spearman analysis, the correlation between the risk score and immune infiltration was computed. $p < 0.05$ showed significant correlation.

2.6 | Functional of GESA enrichment

In order to expose the potential function of the high-risk and low-risk groups, a gene set enrichment analysis was conducted (GSEA).¹⁹ $FDR < 0.05$ indicated significant functional enrichment.

2.7 | Identification of HSPC324 and VIPR1 as potential biomarkers for smoking-related LUAD progression

The edgeR program was used to compare the expression of lncRNA and mRNA in smoking LUAD vs normal tissues. Venn diagrams were then utilized to show common lncRNAs and mRNAs. To confirm the low expression of VIPR1 in smoking LUAD, GSE31210 was utilized as an external validation data set.

2.8 | Statistical analysis

We conducted a statistical study with GraphPad Prism 8.0 (GraphPad Software Inc.). The overall survival difference between the groups was determined using the Kaplan–Meier and log-rank test methods. Two-sided is both statistical analyses. $p < 0.05$ was deemed to be statistically significant.

3 | RESULTS

3.1 | Data acquisition

OncoSG Nat Genet 2020 has 305 cases, four of which have an unclear smoking status and three of which have uncertain survival durations. There were 111 smokers and 187 non-smokers in the room. We discovered that the nonsmoking group outlived the smoking group ($p < 0.05$) (Figure 1A). Simultaneously, we looked at the levels of "immune-gene signature" content in both smoking and nonsmoking groups.²⁰ Eventually, immune proliferation ($p = 0.0037$) (Figure 1B) and NK cell content ($p = 0.0105$) (Figure 1C) were found to be higher in the smoking group. Therefore, it is proved that immune microenvironment is involved in the development of smoking-related LUAD.

The Cancer Genome Atlas database provided information on 407 smoking-positive LUAD patients, including gene expression data, survival status, TNM stage, age, and gender. Three hundred thirty-one immune-related genes were removed from the Molecular Signatures Database. 643 immune-related lncRNAs have been discovered based on $|\text{correlation coefficient}| > 0.6$ and $p < 0.01$. (Table S1).

3.2 | Construction of prognostic signature based on the discovery cohort

Immune-related lncRNAs have been found to be highly linked with overall survival in univariate studies (Table 2). The LASSO regression identified ten optimal lncRNAs: AL359915.2, AP000695.1, HSPC324, AC079949.1, AL442125.2, AL445493.3, TGFB2-AS1, AC026355.1, AC002128.2, and AL121772.33.2 (Figure 2A,B). The top ten lncRNAs were then submitted to multivariate Cox regression

analysis in order to categorize the prognosis-related lncRNAs. AL359915.2, AP000695.1, HSPC324, TGFB2-AS1, AC026355.1, and AC0021282 were identified as independent predictive lncRNAs

for smoking-related LUAD in the multivariate investigation (Table 2). The risk score for each patient was computed as follows: $(0.2432 \times \text{AL359915.2}) + (0.5378 \times \text{AP000695.1}) + (-0.51 \times \text{HSPC324}) +$

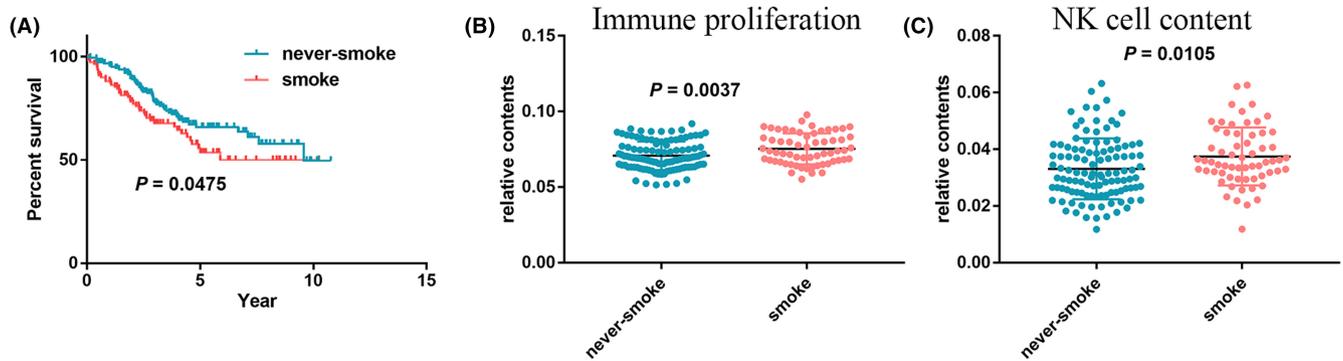


FIGURE 1 Effect of various smoking status on survival. (A) The overall survival of patients in the nonsmoking category is longer than those in the smoking category. ($p < 0.05$) (B) Immune proliferation was higher in the smoking group. ($p = 0.0037$) (C) NK cell content was found to be higher in the smoking group. ($p = 0.0105$)

TABLE 2 Cox regression analysis of immune-related lncRNA

ID	Univariate Cox regression analysis				Multivariate Cox regression analysis				
	HR	95% CI lower	95% CI higher	p Value	Coef	HR	95% CI lower	95% CI higher	p Value
AP000695.1	1.945	1.449	2.612	0.000	0.537	1.712	1.238	2.366	0.001
AL121772.3	1.119	1.027	1.219	0.009					
AL445493.3	1.064	1.013	1.117	0.012					
AC002128.2	1.293	1.047	1.597	0.016	0.372	1.45	1.129	1.864	0.003
AC048341.1	1.203	1.034	1.401	0.016					
TGFB2-AS1	1.097	1.013	1.188	0.021	0.085	1.089	0.989	1.2	0.082
AC091185.1	1.3	1.038	1.629	0.022					
ABALON	1.3889	1.042	1.85	0.024					
HSPC324	0.553	0.327	0.933	0.026	-0.51	0.6	0.367	0.98	0.041
AC026355.1	0.828	0.699	0.98	0.028	-0.213	0.808	0.687	0.949	0.009
AC079949.1	1.258	1.02	1.553	0.031					
AL442125.2	1.117	1.004	1.243	0.04					
AL359915.2	1.233	1.004	1.515	0.045	0.243	1.275	0.964	1.687	0.088

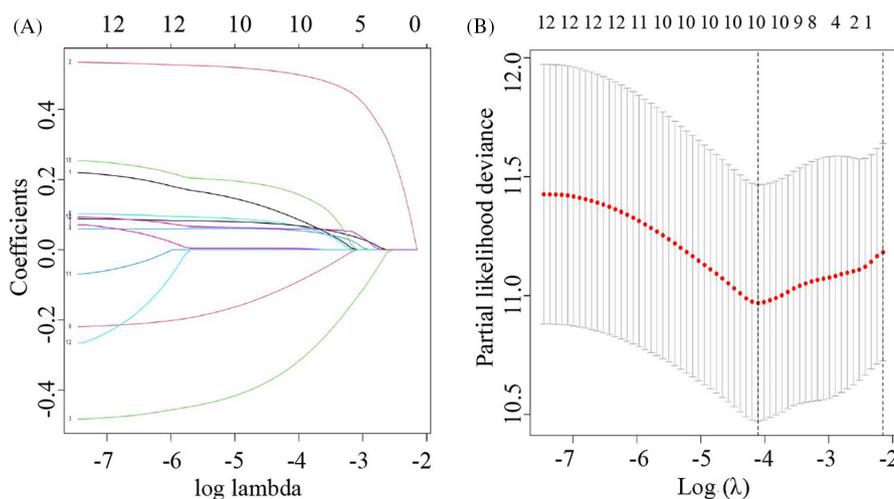


FIGURE 2 Immune-related filtering of genes using regression with LASSO. (A) LASSO coefficient profiles for 13 relevant lncRNAs in univariate Cox regression analysis. For higher lambda values, coefficient profiles diminish. (B) Cross-validation for choosing the LASSO model tuning parameters. Vertical lines are plotted according to the minimum criteria and the 1-standard error criterion, based on the optimal data. The left vertical line reflects the eventually defined ten lncRNAs

$(0.0859 \times \text{TGFB2-AS1}) + (-0.2131 \times \text{AC026355.1}) + (0.3721 \times \text{AC002128.2})$. Based on the median score of the risk signature, 142 and 143 patients were classified as high-risk and low-risk groups, respectively.

Figure 3A depicts the distributions of risk scores, whereas Figure 3B depicts the distributions of survival status. Figure 3C depicts the pattern expression of these six prognostic lncRNAs in high-risk and low-risk groups. Overall survival in the low-risk category

was somewhat higher than in the high-risk group ($p = 2.289 \times 10^{-8}$) (Figure 3D). Overall survival discrepancies between high-risk and low-risk groups were investigated further using distinct clinico-pathological criteria. After removing patients with missing tumor stage, gender, age, or TNM, a total of 187 cases remained in the discovery cohort. As shown in Figure 3E, the low-risk ($n = 97$) group had slightly better overall survival than the high-risk ($n = 90$) group for cases aged 65 years ($p = 0.0051$) and older ($p = 0.0003$), male

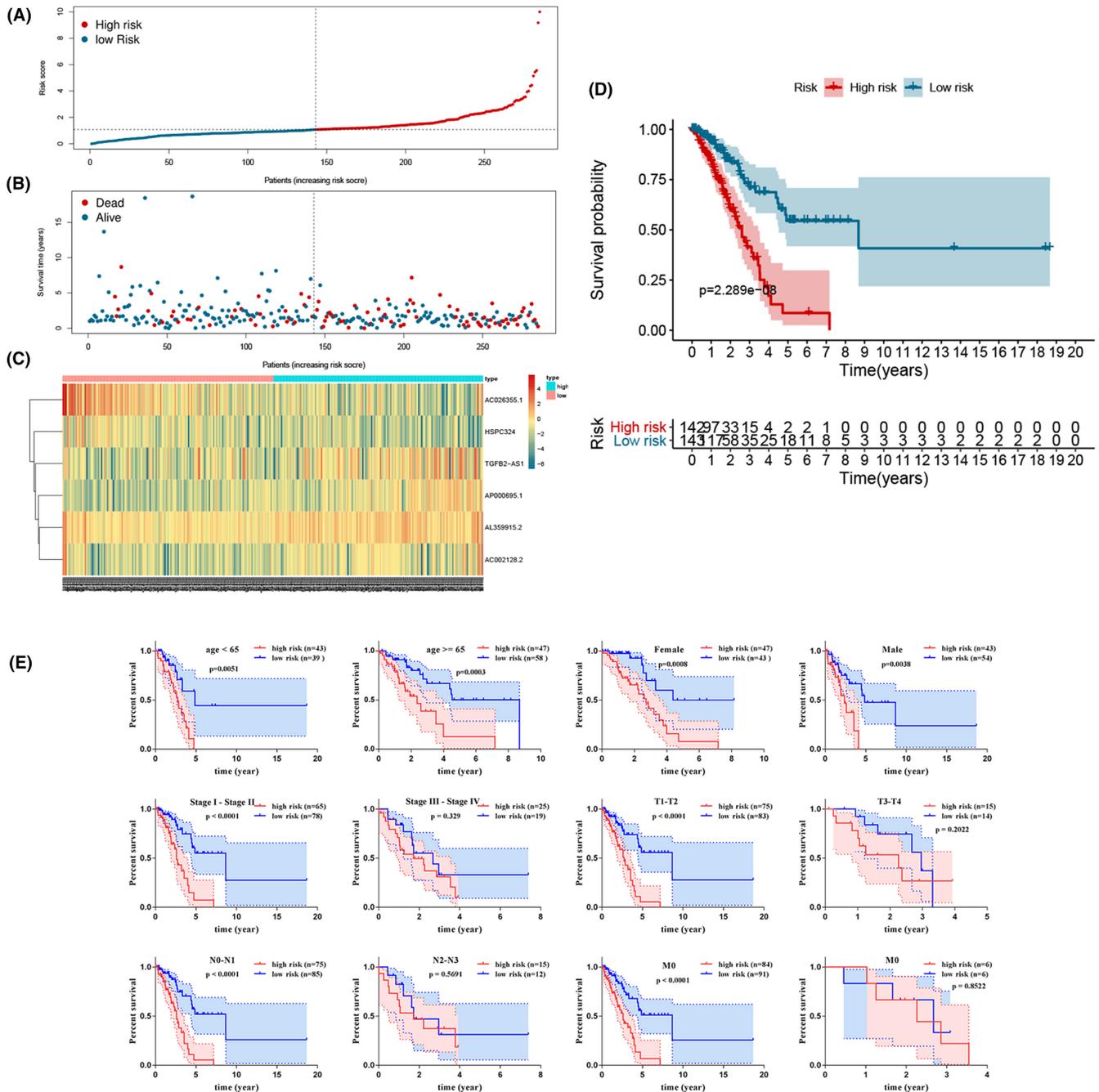


FIGURE 3 Formation of the prognostic signature based on the discovery cohort of The Cancer Genome Atlas database (TCGA). (A) The distribution of scores on risk. (B) Distributions in both high- and low-risk categories' overall survival status. (C) Heatmap of the pattern of six-prognostic immune-lncRNA signature expression between categories of high and low risk. (D) The overall survival of patients in the low-risk category is longer than those in the high-risk category. (E) The overall survival disparity in the TCGA discovery cohort between the low- and high-risk categories stratified by age, gender, stage, and TNM

($p = 0.0038$) and female ($p = 0.0008$), stage I-II ($p < 0.0001$), T1-T2 ($p < 0.0001$), N0-N1 ($p < 0.0001$), and M0 grade ($p < 0.0001$). However, there was no significant difference in overall survival between the high-risk and low-risk groups for patients with stage III-IV ($p = 0.329$), T3-T4 ($p = 0.2022$), N2-N3 ($p = 0.5691$), or M1 ($p = 0.8522$).

3.3 | Validation of the prognostic six-lncRNA signature

The validation cohort included 60 and 62 people who were classified as high risk or low risk based on the discovery cohort cut-off value. Figure 4A depicts the distributions of risk scores,

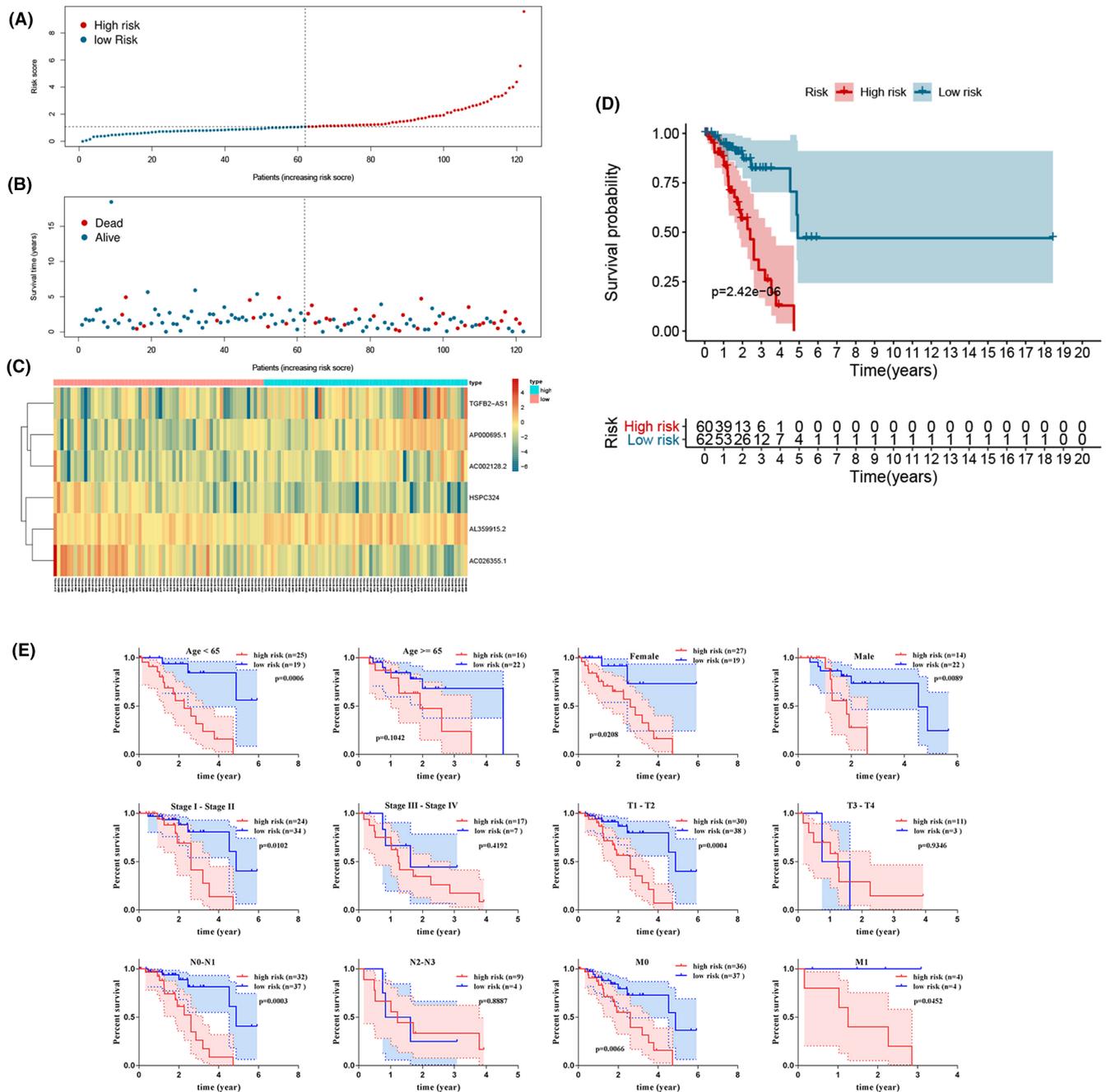


FIGURE 4 Validation of the prognostic signature based on the discovery cohort. (A) The distribution of risk scores in the validation cohort. (B) Distributions in both high- and low-risk categories' overall survival status in the validation cohort. (C) Heatmap of the pattern of six-prognostic immune-lncRNA signature expression between categories of high and low risk in the validation cohort. (D) Patients' overall survival in the low-risk category is longer than those in the high-risk category in the validation cohort. (E) The overall survival disparity in the The Cancer Genome Atlas database (TCGA) validation cohort between the low- and high-risk categories stratified by age, gender, stage, and TNM

longer overall survival than the high-risk ($n = 41$) category for patients age <65 ($p = 0.0006$), in both male ($p = 0.0089$) and female cases ($p = 0.0208$), patients at stage I-II ($p = 0.0102$), T1-T2 ($p = 0.0004$), N0-N1 ($p = 0.0003$), both M0 ($p = 0.0066$) and M1 ($p = 0.0452$). However, there was minimal difference in overall survival between high-risk and low-risk patients above the age ≥ 65 ($p = 0.1042$), stage III-IV ($p = 0.4192$), T3-T4 ($p = 0.9346$), and N2-N3 ($p = 0.8887$) (Figure 4E).

Two hundred and six and 201 persons in the combined cohort were classified as high risk or low risk based on the discovery cohort cut-off value. Figure 5A depicts the distributions of risk scores, whereas Figure 5B depicts the survival status. The heatmap was used to compare the expression of prognostic lncRNAs in high-risk and low-risk groups (Figure 5C). Smoking-related LUAD patients in the low-risk category have a little longer overall survival duration than those in the high-risk category ($p = 9.462e-08$) (Figure 5D). Similarly, after excluding patients with missing information in tumor stage, gender, age, or TNM, 275 cases remained in the combined cohort. The findings

found that the low-risk ($n = 142$) category had longer overall survival than the high-risk ($n = 133$) category for patients age <65 ($p = 0.0002$), age ≥ 65 ($p = 0.0004$) in both male ($p = 0.0002$) and female cases ($p = 0.0003$), patients at stage I-II ($p < 0.0001$), T1-T2 ($p < 0.0001$), N0-N1 ($p < 0.0001$), both M0 ($p < 0.0001$). However, there was little apparent disparity in overall survival for patients stage III-IV ($p = 0.2588$), T3-T4 ($p = 0.0723$), N2-N3 ($p = 0.8041$) and M1 ($p = 0.3346$) in the high-risk and low-risk categories (Figure 5E).

The ROC analysis was primarily utilized to assess the sensitivity and accuracy of the six-lncRNA markers in estimating 1-year, 3-year, and 5-year overall survival. The 1-year survival AUC was 0.67 [95% CI, 0.586–0.754], 3-year 0.7 [95% CI, 0.616–0.780], and 5-year 0.82 [95% CI, 0.744–0.903], suggesting that the six-lncRNA markers were extremely sensitive and specific in the discovery cohort (Figure 6A). The 1-year survival AUC was 0.64 [95% CI, 0.470–0.811], 3-year 0.76 [95% CI, 0.636–0.878] and 5-year 0.88 [95% CI, 0.761–0.991], suggesting that the six-lncRNA was particularly sensitive and specific in the validation cohort (Figure 6B).

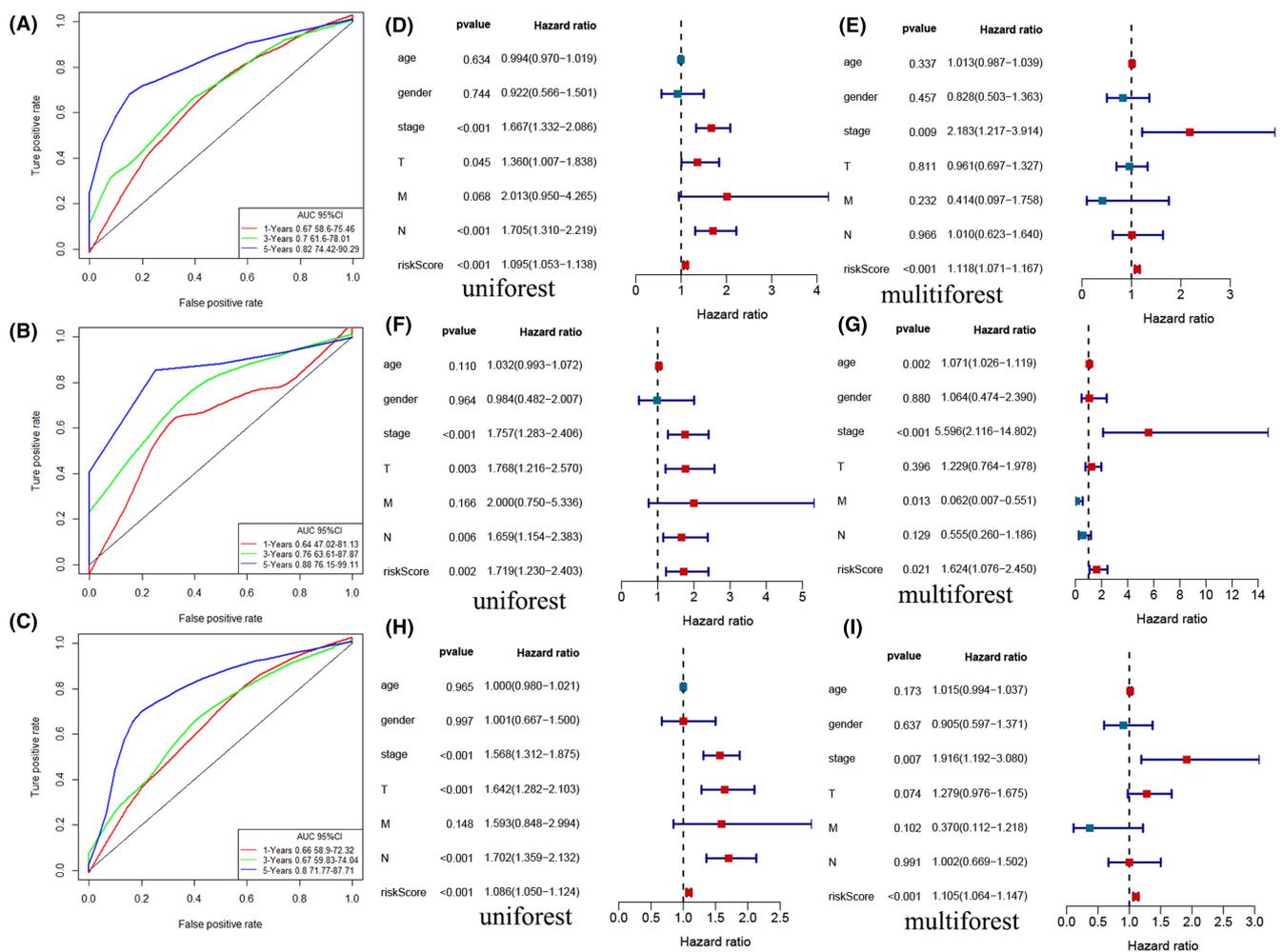


FIGURE 6 Time-dependent receiver operating characteristic curves for 1, 3, and 5 years based on the signature of the six-immune lncRNA. (A) Discovery cohort. (B) Validation cohort. (C) Combine cohort. Univariate analysis and multivariate analysis revealed independent prognostic factors. (D, E) discovery cohort. (F, G) validation cohort. (H, I) combine cohort

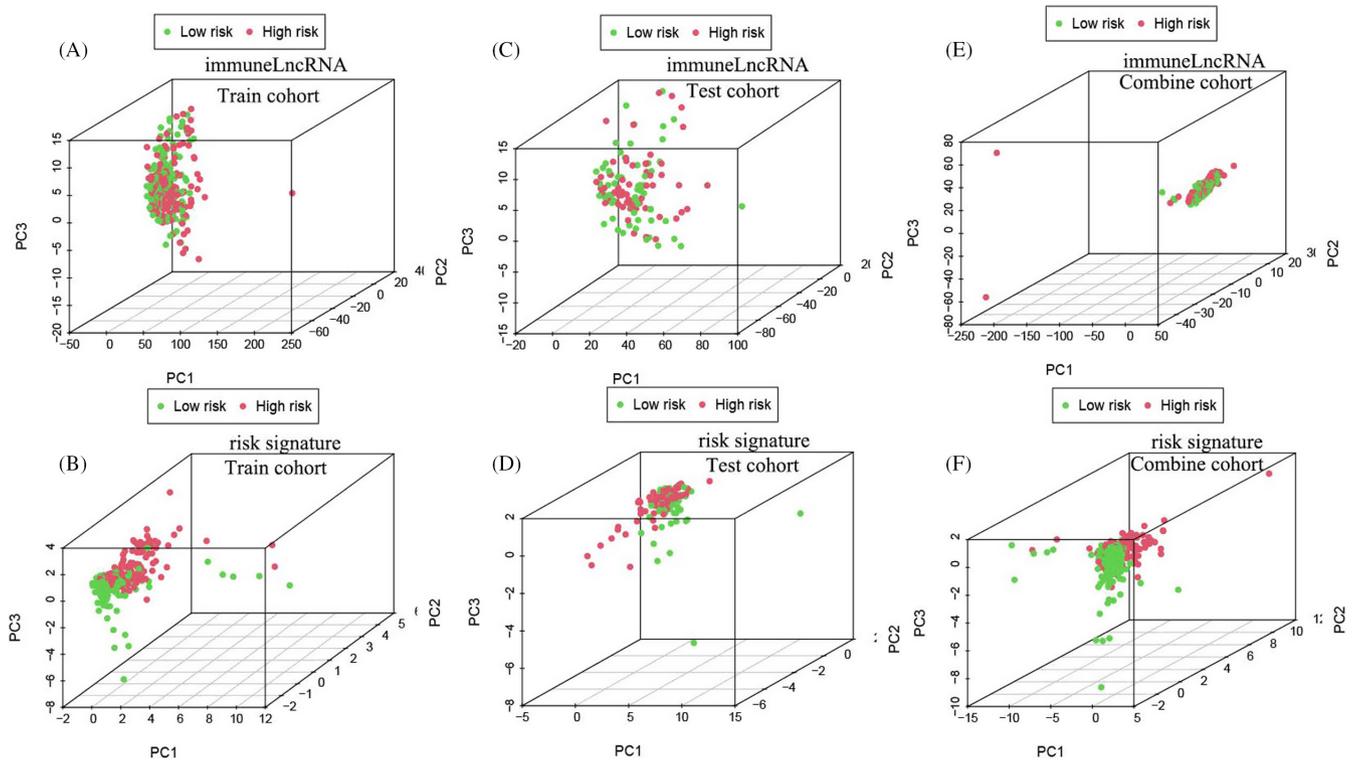


FIGURE 7 Principal component analysis (PCA) based on the seven immune-related lncRNAs showed that the low-risk group and high-risk group tended to separate into two sides. (A, B) Training set. (C, D) Validation set. (E, F) Combination set

The 1-year survival AUC was 0.66 [95% CI, 0.589–0.723], 3-year 0.67 [95% CI, 0.598–0.740] and 5-year 0.8 [95% CI, 0.717–0.877], suggesting that the six-lncRNA was particularly sensitive and specific in the combine cohort (Figure 6C). More specifically, both the discovery cohort ($p < 0.001$, HR = 1.118, 95% CI = 1.071–1.167) (Figure 6D,E), the validation cohort ($p = 0.021$, HR = 1.624, 95% CI = 1.076–2.450) (Figure 6F,G), and combine cohort ($p < 0.001$, HR = 1.105, 95% CI = 1.064–1.147) (Figure 6H,I) multiforest results showed that the risk signature was an independent prognostic factor.

Principal component analysis of the high-risk and low-risk groups reveals that they can be separated using six-lncRNA signatures. (Figure 7A–F).

3.4 | Immune cell infiltration and immune correlations of the prognostic model

CIBERSORT algorithm was used to screen out samples with CIBERSORT output p value less than 0.05 for further research. In the end, high-risk ($N = 195$) and low-risk ($N = 184$) samples were chosen for the CIBERSORT study (Table S2). The results show that Dendritic cells resting ($p = 0.027$) and Mast cells resting ($p = 0.013$) were infiltrated differently in high-risk and low-risk categories (Figure 8A). Furthermore, to determine whether the immune prognostic model accurately reflected the state of the tumor immune

microenvironment, we analyzed the relationship between risk scores and immune cell infiltration. The risk score was substantially connected to Macrophages M0 cells ($R = 0.11$, $p = 0.027$) (Figure 8B), Mast cells resting ($R = -0.2$, $p = 0.001$) (Figure 8C).

3.5 | GSEA functional enrichment analysis

Gene set enrichment analysis revealed several important mechanisms implicated in cancer development and immune-related cancer incidence among the high-risk and low-risk categories. Then, GSEA studied high-risk and low-risk categories detached from the Molecular Signatures Database. In the high-risk category, the possible Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways that were primarily substantially enriched in ADHERENS JUNCTION, CELL CYCLE, DNA REPLICATION, ECM RECEPTOR INTERACTION, FOCAL ADHESION, GAP JUNCTION, PATHWAYS IN CANCER et al ($FDR < 0.05$) (Figure 9A), but the significant enriched KEGG pathways in the low-risk category were ALPHA LINOLENIC ACID METABOLISM, ARACHIDONIC ACID METABOLISM, ARGININE AND PROLINE METABOLISM, BETA-ALANINE METABOLISM, DRUG METABOLISM CYTOCHROME P450, FATTY ACID METABOLISM, GLYCINE SERINE AND THREONINE METABOLISM, LINOLEIC ACID METABOLISM, METABOLISM OF XENOBIOTICS BY CYTOCHROME P450, PROPANOATE METABOLISM et al ($FDR < 0.05$) (Figure 9B).

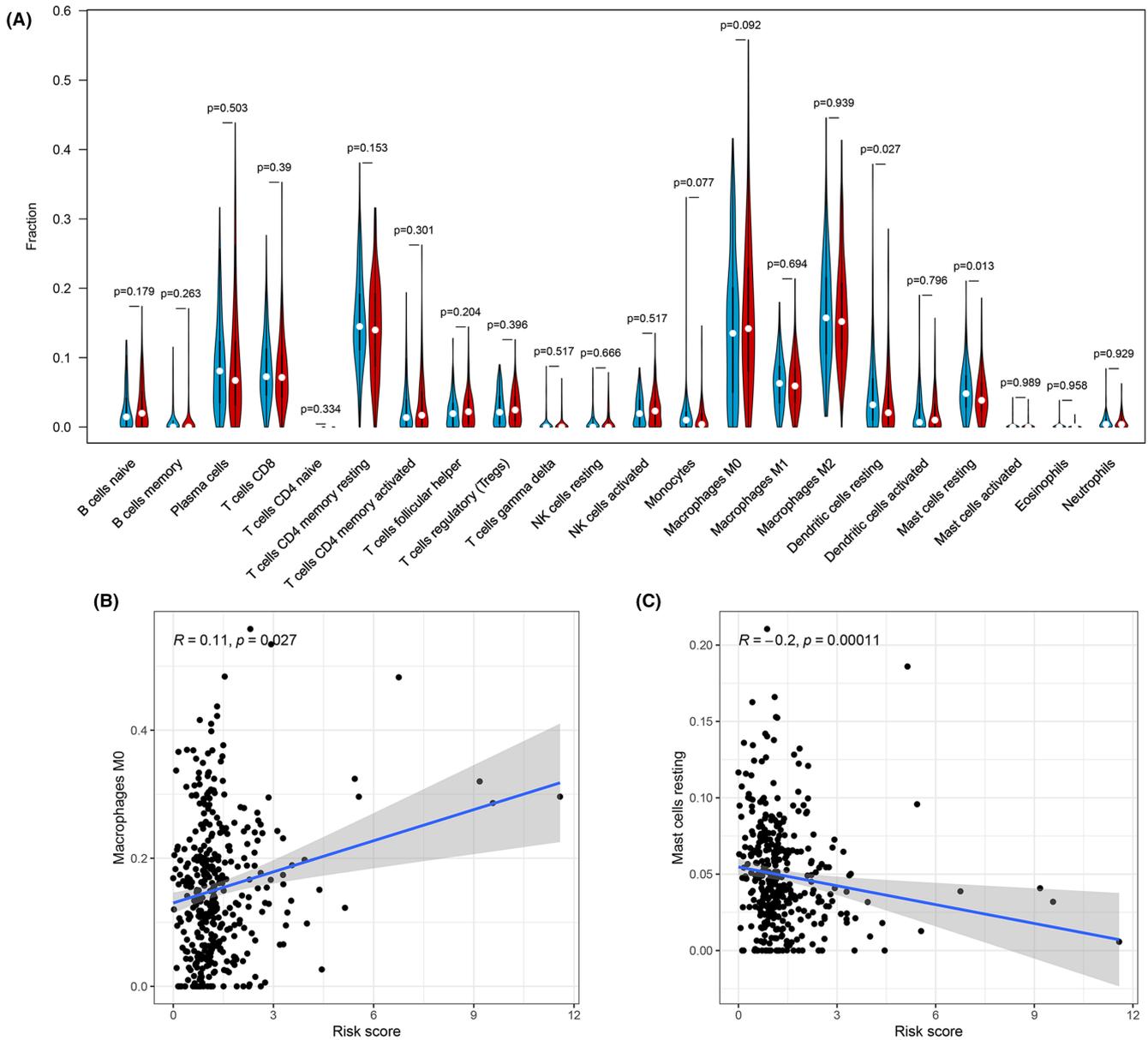


FIGURE 8 Immune cell infiltration and immune correlations of the prognostic model. (A) Violin plot comparing the proportions of TIICs between high-risk and low-risk group. (B) Risk score was significantly related to Macrophages M0 cells. (C) Risk score was significantly related to Mast cells resting

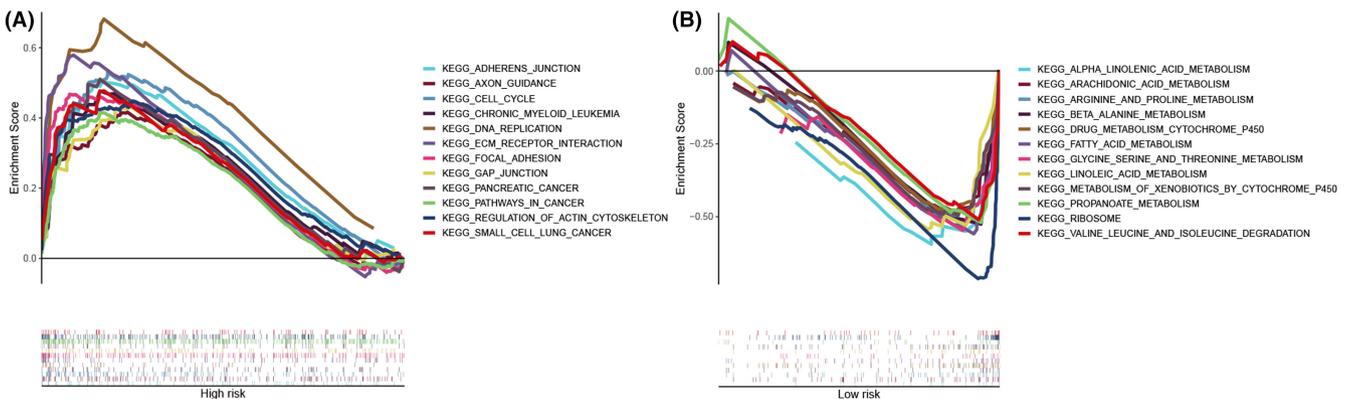


FIGURE 9 Functional annotation between the high- and low-risk categories. (A) High-risk group. (B) Low-risk group

3.6 | Identification of tightly correlation HSPC324 and VIPR1 as candidate biomarkers for smoking-positive LUAD progression

When smoking-related LUAD was compared to normal tissues, 124 lncRNAs and 1030 mRNAs were found to be differentially expressed ($|\log_2\text{-fold change}| \geq 2.0$ and $p < 0.05$) (Figure 10A,B). Common lncRNAs in differentially expressed lncRNAs and risk signature lncRNAs are represented by Venn diagrams (Figure 10C). Similarly, we choose common mRNAs in differentially expressed mRNAs and six risk signature lncRNAs related mRNAs (Figure 10D; Table 3). As a consequence, we identified the lncRNA HSPC324 and the mRNA VIPR1 as possible target genes that are highly positively linked. The expression of HSPC324 and VIPR1 in smoking LUAD was much lower than in normal tissue, according to our findings ($p < 0.0001$) (Figure 10E,F). High expression HSPC324 and VIPR1 had considerably longer overall survival times than low expression (Figure 10G,H) ($p < 0.05$). VIPR1 was shown to be substantially reduced in smoking-positive LUAD than in normal tissue in GSE31210 (Figure 10I).

4 | DISCUSSION

Non-small cell lung cancer, the most frequent kind of adenocarcinoma, continues to be the largest cause of cancer-related mortality worldwide. In clinical treatment, several therapeutic techniques, like as chemotherapy, immunotherapy, and others, are used.²¹⁻²³ The majority of patients are in the late stages of metastasis at the time of diagnosis, leading in delayed treatment.²⁴ We examined the OncoSG Nat Genet 2020 data and discovered that smoking is associated with immune response and can boost the production of certain immune

markers. As a result, finding more effective prognostic factors to reliably predict patients' survival with smoking-related LUAD is critical.

Many lncRNA prognostic indicators for various malignancies have been identified in recent years.²⁵ Qu et al. classify a promising and potential four-lncRNA predictive model in predicting the survival of patients with stage I-III clear cell renal cell carcinoma.⁹ For stage I-II LUAD patients without adjuvant treatment, Peng et al. described a two-lncRNA prognostic signature consisting of C1orf132 and TMPO-AS1.²⁶ Zhang et al. assume that the immune-related lncRNA model has an important survival prediction benefit and tests the response to hepatocellular carcinoma immunotherapy.²⁷ A growing number of studies to identify diagnostic or prognostic biomarkers for LUAD.²⁸⁻³⁰

The diverse heterogeneity in patients' clinical effects with LUAD associated with smoking calls for novel biomarkers for prognosis. For smoking-related LUAD, a few immune-related lncRNA prognostic markers have been carefully defined. To our knowledge, we first built a six immune-related lncRNA predictive signature for smoking-related LUAD in this publication. Unlike the traditional technique, the LASSO algorithm may choose the variables with the greatest relevance.³¹ After multivariate Cox regression analyses, six immune-related lncRNAs (AL359915.2, AP000695.1, HSPC324, TGFB2-AS1, AC026355.1, and AC002128.2) played imperative roles in this study, and they constructed our risk signature model. Jafarzadeh et al.³² found that ectopic expression of lncRNA HSPC324 involves many cancer progression, such as repressed proliferation and migration, increased apoptosis in lung adenocarcinoma cells. Ling et al.³³ indicated that TGFB2-AS1 regulates lung adenocarcinoma progression via mediate mir-340-5p expression to target EDNRB expression. Panagiotis et al.³⁴ revealed that upregulated TGF- β /Smad-mediated transcription and TGF- β -target genes resulting in the depleting of TGFB2-AS1. Liu et al.³⁵ observed that TGFB2-AS1 depletion expression also

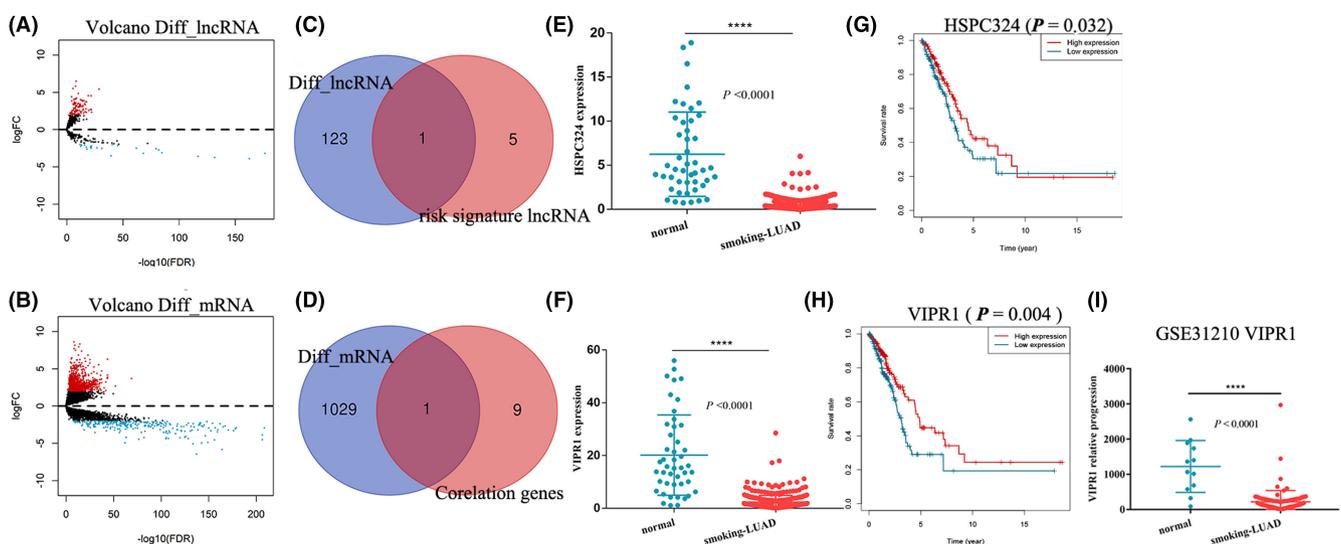


FIGURE 10 Expression levels of the differentially expressed genes in smoking-associated lung adenocarcinoma (LUAD) to normal tissues. (A) lncRNA. (B) mRNA. Venn diagram shows the intersection genes. (C) lncRNA. (D) mRNA. The expression of HSPC324 and VIPR1 in smoking LUAD and normal tissue. (E) HSPC324. (F) VIPR1. Overall survival analysis of HSPC324 and VIPR1 in smoking-positive LUAD. (G) HSPC324. (H) VIPR1. The expression of VIPR1 in GSE31210 dataset. (I) The expression of VIPR1 in smoking-positive LUAD and normal tissue

Immune Gene	lncRNA	Cor	p Value	Regulation
KMT2A	AL359915.2	0.654867	5.74E-56	Positive
SEMA7A	AP000695.1	0.739221	3.02E-78	Positive
VIPR1	HSPC324	0.629577	1.29E-50	Positive
NOTCH4	HSPC324	0.610873	5.73E-47	Positive
TGFB2	TGFB2-AS1	0.893317	2.55E-156	Positive
DPP8	AC026355.1	0.635793	6.92E-52	Positive
TRAF6	AC026355.1	0.604853	7.62E-46	Positive
ZEB1	AC002128.2	0.636743	4.40E-52	Positive
KMT2A	AC002128.2	0.74286	2.11E-79	Positive
DPP8	AC002128.2	0.70636	1.27E-68	Positive
ATP6V0A2	AC002128.2	0.611893	3.68E-47	Positive
TRAF6	AC002128.2	0.723092	2.38E-73	Positive
ACVR2A	AC002128.2	0.62785	2.86E-50	Positive

TABLE 3 Risk signature lncRNA-related immune genes

prevents HepG2 cell proliferation and migration and induces apoptosis. This six immune lncRNA risk signature was powerfully correlated with the overall survival of smoking-related LUAD patients and could also predict 1-year, 3-year, and 5-year overall survival in both discovery, validation, and combine cohorts. The AUC for 5-year overall survival in the discovery cohort was 0.82, 0.88 in the validation cohort, and 0.80 in the combine cohort, indicating that this risk signature has significant predictive potential. Univariate and multivariate analyses of Cox regression showed that independent prognostic variables were the stage and risk signature model. This risk signature was in conclusion, a valid prognostic model of clinical relevance.

Pathway enrichment indicated that the different pathways potentially affected smoking-related LUAD progression in high-risk or low-risk patients. We suppose the lncRNA's possible function using the mRNA expression data of the same category of patients. To find KEGG pathway enrichment, we calculated whole mRNA expression data of high-risk or low-risk categories in GSEA software. As a result, most of the high-risk group are enriched in cancer-related pathways, such as ADHERENS JUNCTION, CELL CYCLE, DNA REPLICATION, ECM RECEPTOR INTERACTION, FOCAL ADHESION, GAP JUNCTION, PATHWAYS IN CANCER, and most of the low-risk groups are enriched in metabolism-related pathways et al. So, we think that based on this risk model, the high-risk group is more conducive to cancer progression.

lncRNAs regulate gene expression by interacting with DNA, RNA, and protein and affect RNA splicing, stability, and translation, et al.³⁶ In our result, immune-related lncRNA HSPC324 positively correlated with mRNA VIPR1. Both HSPC324 and VIPR1 were low expression in smoking-related LUAD. Deeply analysis of the biological functions of HSPC324 and VIPR1 through vivo or vitro experiments might provide novel mechanism insights for the carcinogenesis of smoking-related LUAD.

Although we have established and verified the risk model, our research has limitations. Firstly, we were seeking to use the GEO database (<https://www.ncbi.nlm.nih.gov/geo/>) as the validation package. As a validation package, no qualifying listing lncRNA dataset

was used to our best advantage. As a result, the predictive value of six immune-lncRNA signatures was only assessed by the discovery, validation, and combine cohorts, randomly divided by the TCGA dataset. Secondly, the underlying molecular mechanism of these prognostic lncRNAs in smoking-related LUAD is unclear. Thirdly, we only analyzed the correlations between the six-lncRNA signature with clinicopathological parameters, such as age, gender, and stage of smoking-related LUAD. There were no other clinical features such as radioresistance, chemoresistance, as well as EGFR status, PD-1/PD-L1 status, et al. Further studies on more detailed clinical knowledge must be done to investigate the clinical importance of the six-lncRNA signature in LUAD linked to smoking. Consequently, we have proved that six immune-lncRNA markers can accurately predict the prognosis of smoking-related LUAD, which is very promising in clinical practice.

AUTHOR CONTRIBUTIONS

Jing Wang, Dajie Zhou, and Xiangdong Liu involved in conception and design. Jing Wang and Dajie Zhou involved in data selection and assembly and manuscript writing. Jing Wang involved in data analysis and interpretation. All authors made final approval of manuscript.

ACKNOWLEDGMENTS

Shandong Province Key R&D Program (No. 2019GSF108201).

CONFLICT OF INTEREST

There is no conflict of interest in this manuscript.

DATA AVAILABILITY STATEMENT

On request, the data used to validate the results of this analysis are available from the corresponding author.

ORCID

Dajie Zhou  <https://orcid.org/0000-0002-4608-0486>

Jing Wang  <https://orcid.org/0000-0003-3923-5643>

REFERENCES

- Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. *CA Cancer J Clin.* 2019;69(1):7-34.
- The Lancet. GLOBOCAN 2018: counting the toll of cancer. *Lancet.* 2018;392(10152):985.
- Allemani C, Matsuda T, Di Carlo V, et al. Global surveillance of trends in cancer survival 2000–14 (CONCORD-3): analysis of individual records for 37 513 025 patients diagnosed with one of 18 cancers from 322 population-based registries in 71 countries. *Lancet.* 2018;391(10125):1023-1075.
- Park E, Perez G, Regan S, et al. Effect of sustained smoking cessation counseling and provision of medication vs shorter-term counseling and medication advice on smoking abstinence in patients recently diagnosed with cancer: a randomized clinical trial. *JAMA.* 2020;324(14):1406-1418.
- Peng Y, Tang D, Zhao M, Kajiyama H, Kikkawa F, Kondo Y. Long non-coding RNA: a recently accentuated molecule in chemoresistance in cancer. *Cancer Metastasis Rev.* 2020;39(3):825-835.
- Poursheikhani A, Bahmanpour Z, Razmara E, et al. Non-coding RNAs underlying chemoresistance in gastric cancer. *Cell Oncol.* 2020;43(6):961-988.
- Dong H, Wang R, Jin X, Zeng J, Pan J. LncRNA DGCR5 promotes lung adenocarcinoma (LUAD) progression via inhibiting hsa-mir-22-3p. *J Cell Physiol.* 2018;233(5):4126-4136.
- Shahabi S, Kumaran V, Castillo J, et al. LINC00261 is an epigenetically regulated tumor suppressor essential for activation of the DNA damage response. *Can Res.* 2019;79(12):3050-3062.
- Qu L, Wang Z, Chen Q, et al. Prognostic value of a long non-coding RNA signature in localized clear cell renal cell carcinoma. *Eur Urol.* 2018;74(6):756-763.
- Li J, Chen Z, Tian L, et al. LncRNA profile study reveals a three-lncRNA signature associated with the survival of patients with oesophageal squamous cell carcinoma. *Gut.* 2014;63(11):1700-1710.
- Wu J, Zhao Y, Zhang J, Wu Q, Wang W. Development and validation of an immune-related gene pairs signature in colorectal cancer. *Oncoimmunology.* 2019;8(7):1596715.
- Xu M, Xu X, Pan B, et al. LncRNA SATB2-AS1 inhibits tumor metastasis and affects the tumor immune cell microenvironment in colorectal cancer by regulating SATB2. *Mol Cancer.* 2019;18(1):135.
- Chen J, Yang H, Teo ASM, et al. Genomic landscape of lung adenocarcinoma in East Asians. *Nat Genet.* 2020;52(2):177-186.
- Tessema M, Yingling CM, Liu Y, et al. Genome-wide unmasking of epigenetically silenced genes in lung adenocarcinoma from smokers and never smokers. *Carcinogenesis.* 2014;35(6):1248-1257.
- Li X, Li J, Wu P, et al. Smoker and non-smoker lung adenocarcinoma is characterized by distinct tumor immune microenvironments. *Oncoimmunology.* 2018;7(10):e1494677.
- Barbie DA, Tamayo P, Boehm JS, et al. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature.* 2009;462(7269):108-112.
- Li W, Chen Q, Huang T, Wu P, Shen L, Huang Z. Identification and validation of a prognostic lncRNA signature for hepatocellular carcinoma. *Front Oncol.* 2020;10:780.
- Tang X, Li Y, Liang S, et al. Development and validation of a gene expression-based signature to predict distant metastasis in locoregionally advanced nasopharyngeal carcinoma: a retrospective, multicentre, cohort study. *Lancet Oncol.* 2018;19(3):382-393.
- Zhang L, He M, Zhu W, et al. Identification of a panel of mitotic spindle-related genes as a signature predicting survival in lung adenocarcinoma. *J Cell Physiol.* 2020;235(5):4361-4375.
- Sobhani N, Corona SP, Roviello G, et al. Immune-gene signature: a new tool for patient selection for checkpoint inhibitors? *Future Oncol.* 2020;16(19):1327-1330.
- Malhotra J, Jabbar SK, Aisner J. Current state of immunotherapy for non-small cell lung cancer. *Transl Lung Cancer Res.* 2017;6(2):196-211.
- Schiller JH, Harrington D, Belani CP, et al. Comparison of four chemotherapy regimens for advanced non-small-cell lung cancer. *N Engl J Med.* 2002;346(2):92-98.
- Arbour K, Riely G. Systemic therapy for locally advanced and metastatic non-small cell lung cancer: a review. *JAMA.* 2019;322(8):764-774.
- Bodor JN, Kasireddy V, Borghaei H. First-line therapies for metastatic lung adenocarcinoma without a driver mutation. *J Oncol Pract.* 2018;14(9):529-535.
- Zheng J, Zhao Z, Wan J, et al. N-6 methylation-related lncRNA is potential signature in lung adenocarcinoma and influences tumor microenvironment. *J Clin Lab Anal.* 2021;35(11):e23951.
- Peng F, Wang R, Zhang Y, et al. Differential expression analysis at the individual level reveals a lncRNA prognostic signature for lung adenocarcinoma. *Mol Cancer.* 2017;16(1):98.
- Zhang Y, Zhang L, Xu Y, Wu X, Zhou Y, Mo J. Immune-related long noncoding RNA signature for predicting survival and immune checkpoint blockade in hepatocellular carcinoma. *J Cell Physiol.* 2020;235(12):9304-9316.
- Cui Y, Fang W, Li C, et al. Development and validation of a novel signature to predict overall survival in "Driver Gene-negative" Lung Adenocarcinoma (LUAD): results of a multicenter study. *Clin Cancer Res.* 2019;25(5):1546-1556.
- Shukla S, Evans J, Malik R, et al. Development of a RNA-Seq based prognostic signature in lung adenocarcinoma. *J Natl Cancer Inst.* 2017;109(1):djw200.
- Chen Y, Zhang X, Li J, Zhou M. Immune-related eight-lncRNA signature for improving prognosis prediction of lung adenocarcinoma. *J Clin Lab Anal.* 2021;35(11):e24018.
- Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw.* 2010;33(1):1-22.
- Jafarzadeh M, Tavallaie M, Soltani BM, Hajipoor S, Hosseini SM. LncRNA HSPC324 plays role in lung development and tumorigenesis. *Genomics.* 2020;112(3):2615-2622.
- Ling Z, Wen Z, Tang Z, et al. LncRNA TGFB2-AS1 regulates lung adenocarcinoma progression via act as a sponge for miR-340-5p to target EDNRB expression. *Am J Transl Res.* 2020;12(7):3813-3821.
- Papoutsoglou P, Tsubakihara Y, Caja L, et al. The TGFB2-AS1 lncRNA regulates TGF- β signaling by modulating corepressor activity. *Cell Rep.* 2019;28(12):3182-98.e11.
- Liu W, Huai R, Zhang Y, et al. Down-regulation expression of TGFB2-AS1 inhibits the proliferation, migration, invasion and induces apoptosis in HepG2 cells. *Genes Genomics.* 2019;41(8):951-959.
- Statello L, Guo C-J, Chen L-L, Huarte M. Gene regulation by long non-coding RNAs and its biological functions. *Nat Rev Mol Cell Biol.* 2021;22(2):96-118.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Zhou D, Wang J, Liu X. Development of six immune-related lncRNA signature prognostic model for smoking-positive lung adenocarcinoma. *J Clin Lab Anal.* 2022;36:e24467. doi:[10.1002/jcla.24467](https://doi.org/10.1002/jcla.24467)