Am J Transl Res 2022;14(10):7308-7323
www.ajtr.org /ISSN:1943-8141/AJTR0141757

Original Article

Epithelial-mesenchymal transition-related lncRNAs associated with prognosis and immune cell infiltration in lung adenocarcinoma

Quncheng Zhang[1,2*], Xuan Wu[1,2,3*], Ya Sun[4], Li Yang[1,2,3], Ziqi Wang[1,2,3], Yuanjian Yang[1,2], Xingru Zhao[1,2], Xiaoju Zhang[1,2]

[1]Department of Respiratory and Critical Care Medicine, Zhengzhou University People's Hospital, China; [2]Henan Provincial People's Hospital, Zhengzhou, China; [3]Academy of Medical Science, Zhengzhou University, Zhengzhou, China; [4]Xinxiang Medical University, Xinxiang, China. [*]Equal contributors.

Received January 14, 2022; Accepted May 11, 2022; Epub October 15, 2022; Published October 30, 2022

Abstract: Background: Lung adenocarcinoma (LUAD) remains the most common type of lung cancer and is associated with distant metastasis and poor prognosis. Epithelial-mesenchymal transition (EMT) plays crucial roles in carcinogenesis, embryogenesis, and wound healing. EMT-related molecules may be adopted for early diagnosis and prognosis of cancer and targeting them may constitute an attractive strategy for treatment. This study aims to identify the EMT-related long non-coding RNAs (lncRNAs) and develop a risk signature to accurately predict the prognosis of LUAD patients. Methods: The RNA-seq data and corresponding clinical profiles were obtained from LUAD cohort of The Cancer Genome Atlas (TCGA) database. EMT-related lncRNAs significantly associated with prognosis were identified by Pearson correlation analysis and univariate regression analysis. Subsequently, an EMT-related prognostic risk signature was developed through LASSO and multivariate regression analyses. Kaplan Meier and receiver operating characteristic curve analysis were implemented to assess the predictive performance of the signature. The nomogram was constructed to predict the 1-year, 3-year, and 5-year overall survival of LUAD patients. Additionally, enrichment analyses were carried out to identify probable biologic processes and cellular pathways involved in the signature. The correlation of immune cell infiltration and risk score was also evaluated by CIBERSORT algorithm. Finally, we constructed a ceRNA network to further study possible downstream targets and molecular mechanisms of EMT-related lncRNAs in LUAD. Results: Eight EMT-related lncRNAs were identified to develop a prognostic risk signature in LUAD. Patients with high-risk scores had worse survival outcomes than those with low-risk scores. The signature showed robust predictive potential, and was verified to be an independent prognostic factor. Moreover, the risk score based on the signature was significantly correlated with immune cell infiltration in LUAD. Conclusions: We established and validated a prognostic signature that reflects the tumor microenvironment characteristics and predicts the outcomes for LUAD.

Keywords: EMT-related lncRNAs, prognosis, signature, immune cell infiltration, lung adenocarcinoma

Introduction

As the second most commonly diagnosed cancer, lung cancer was the top cause of cancer-related deaths both in men and women in 2021 [1]. Millions of new cases are diagnosed with lung cancer every year, and frustratingly, the disease is always associated with poor prognosis and high cost. Lung adenocarcinoma (LUAD) remains the most diagnosed histologic subtype of lung cancer, comprising around 40% of all cases. Despite advances in treatment strategies over the past few years, the prognosis of LUAD patients is still unsatisfactory [2]. Genetic alterations are crucial in biologic pathways during carcinogenesis, and further induce the recurrence of cancer and reduce the survival rate [3].

Thus, it is essential to identify molecular markers that, together, constitute a prognostic risk signature.

Defined as the interaction between basement membrane and polarized epithelial cells, epithelial-mesenchymal transition (EMT) occurs

during wound healing, fibrosis, organ development, and embryogenesis, and cells are morphologically dedifferentiated from an epithelial to a mesenchymal phenotype, including a low expression level of E-cadherin and high expression level of N-cadherin, vimentin, and cellular proteases [4]. More notably, the dysregulation of EMT is involved in carcinogenesis and progression of cancers, including the resistance to anti-tumor drugs, invasion, and metastasis [5]. EMT is regulated by a variety of genes that originate from the tumor stroma to affect the progression of cancer [6]. Next-generation sequencing enables the comprehensive genomic atlas of cancer to be extensively studied. Specific genomic constitutents, including long non-coding RNA (lncRNA), mRNA, and microRNA, provide insight into tumor molecular characteristics [7]. The lncRNAs are implicated in tumor progression and may have prognostic value in cancer [8, 9]. Notably, EMT-related lncRNAs play critical roles in the processes [10]. For example, ADAMTS9-AS1 has been reported to be related to lymph node metastasis and prognosis in colorectal cancer [11]. Upregulation of PTAR can facilitate EMT and metastasis in the progression of ovarian cancer [12].

Given the critical roles of EMT and lncRNAs in carcinogenesis and progression of cancer, our study aims to explore the prognostic value of EMT-related lncRNAs in LUAD. We downloaded the RNA-seq data from The Cancer Genome Atlas (TCGA) database to develop and validate a risk signature with EMT-related lncRNAs for LUAD. A lncRNA-mediated ceRNA network was also constructed to further explore the downstream targets and molecular mechanisms of EMT-related lncRNAs in LUAD.

Materials and methods

Data collection

The 200 genes related to EMT were obtained from the MSigDB v7.2. The RNA-seq expression information and corresponding clinical profiles of LUAD samples were obtained from the TCGA database. Totally, 492 LUAD samples with complete survival data and OS more than 30 days were included and randomly divided into training and testing datasets in a 3:1 ratio.

Correlation analysis

The EMT-related lncRNAs were determined using Pearson's correlation analysis (correlation coefficient $|R|>0.5$ and $P<0.01$).

Construction of a prognostic signature

We used univariate Cox analysis to determine the EMT-related lncRNAs that were closely related to OS using the "survival" package in the training dataset ($P<0.05$). The candidate lncRNAs were defined as risk-related ones with hazard ratios (HRs) HRs>1 and protective ones (HRs<1). The "glmnet" package was used to perform LASSO regression analysis with the candidate lncRNAs. As a type of linear regression, LASSO is a form of compression estimation that can produce a more refined signature by setting up a penalty function, that can compress certain coefficients by defining others as zero, gaining the advantage of subset shrinkage. Subsequently, 20 lncRNAs were analyzed by multivariate Cox regression analysis to determine the independent prognostic factors using the "survminer" package. Finally, eight EMT-related lncRNAs were identified to construct a prognostic signature. The formula based on the regression coefficient and expression of each lncRNA is shown below.

Risk Score = 0.033460557 * AP000695.6 + 0.065258036 * AGAP2-AS1 + 0.19135319 * RP11-342K6.1 + (-0.069681386) * LINC01128 + (-0.151082575) * CRNDE + (-0.016143386) * TRG-AS1 + (-0.005174436) * RP11-116018.1 + (-0.118042351) * LINC00892.

Verification of the prognostic signature

The risk score of each sample was calculated using the formula above. Then the samples were classified into the high and low-risk groups according to the median risk score. A scatter plot was applied to show the correlation between the survival status and risk score. Kaplan-Meier survival curve analysis was used to compare the survival of patients between the two risk groups. The predictive accuracy of the signature was assessed using receiver-operating characteristic (ROC) curves. ROC curves were used to evaluate the effectiveness of the risk score, and area under the curve

(AUC) was calculated by the "survivalROC" R package.

Correlation of clinicopathologic characteristics and risk scores

Chi-square test was performed to explore the correlation of risk scores with clinicopathologic characteristics, including age (<65/≥65 years), gender (female/male), tumor stage (stage I/II/III/IV), TNM stage (TI/T2/T3/T4, N0/N1/N2/N3, and M0/M1).

Establishment and verification of the nomogram

To predict the probable 1, 3, and 5-year OS of LUAD patients, we constructed a prognostic nomogram model using the "rms" R package based on traditional clinical variables related to prognosis and risk score. Then, calibration plots and time-dependent ROC curves were performed to assess the performance of the prognostic nomogram.

Functional enrichment analysis

The "limma" R package was used to determine the differentially expressed genes (DEGs) between the two risk groups using |logFC|>0.5 and P<0.05. Subsequently, the "clusterProfiler" R package was used to perform Gene Ontology (GO) enrichment analysis of DEGs to identify biologic processes, cellular components, and molecular functions. Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis was used to explore the potential enriched signaling pathways. We also performed the Gene Set Enrichment Analysis (GSEA) to explore biologic pathways and functions related to low and high-risk groups.

Correlation of the signature and immune cell infiltration

To identify the correlation between the risk group and immune cell infiltration, the CIBERSORT algorithm was performed to define the proportion of 28 immune cells acquired from gene expression profiles. CIBERSORT is a method for characterizing cell composition of complex tissues from their gene expression profiles. Wilcoxon rank-sum test was used to determine immune cells that had different infiltration level between the high and low-risk groups. Pearson correlation coefficient was performed to explore the relationship between each signature lncRNA and immune cells.

Construction of the ceRNA network and functional enrichment analysis

The miRcode database was downloaded from the website (http://www.mirco de.org/), and used to determine miRNAs that interacted with the lncRNAs. Finally, 18 pairs of interactions in eight miRNAs and five lncRNAs were identified. Subsequently, the correlation of the miRNAs and their target mRNAs was identified by TargetScan, and miRTarBase, and 217 mRNAs in total were determined. Cytoscape software was used to visualize the ceRNA network. The target mRNAs were used to construct a PPI network with the help of STRING website. We obtained the hub genes from the above PPI network using CytoHubba plugin of Cytoscape. Finally, GO and KEGG analyses were used to unravel the main functions of the target mRNAs.

Statistical analysis

All the data analysis was done with R software (version 4.0.2). Kaplan-Meier curve analysis was performed to analyze the survival. LASSO, univariate, and multivariate Cox regression analyses were performed to assess prognostic significance. ROC curve analyses and AUC were applied to determine the predictive capacity of the prognostic risk signature. P<0.05 was considered significant.

Results

Construction of a prognostic signature in the training dataset

The flowchart of this study is shown in **Figure 1**. We identified 200 genes related to EMT from MSigDB v7.2 and 178 EMT-related lncRNAs using Pearson correlation analysis from the TCGA LUAD cohort. Univariate Cox regression analysis was done to determine the candidate lncRNAs with the significant prognostic values, in total, 54 EMT-related lncRNAs were included (Figure S1). Then, the included lncRNAs were entered into an OS-based LASSO Cox regression model with penalty parameter tuning conducted by 10-fold cross-validation to detect the best penalty parameter lambda. 20 lncRNAs were identified for further study (**Figure 2A-C**). Eight lncRNAs were finally determined through multivariate Cox regression analysis to con-

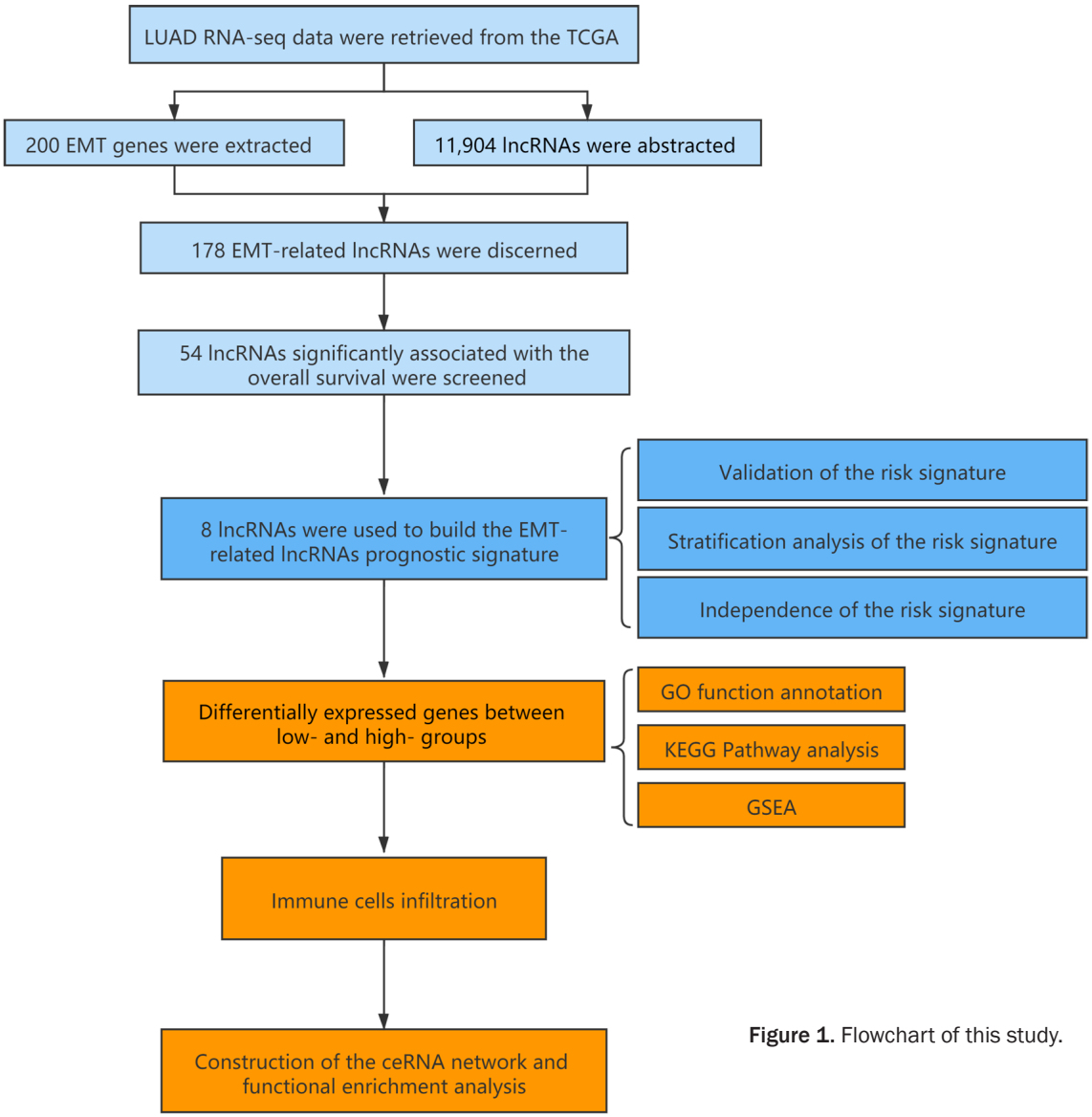

Figure 1. Flowchart of this study.

struct the risk signature (**Figure 2D**). Three lncRNAs (AP000695.6, AGAP2-AS1, and RP11-342K6.1) with HRs>1 were considered as risk factors, while the other five (LINC01128, CRNDE, TRG-AS1, RP11-116018.1, and LINC-00892) with HRs<1 were regarded as protective lncRNAs. Subsequently, a Sankey diagram (**Figure 2E**) was used to show the correlation of the lncRNAs and the EMT-related genes as well as the risk types.

Evaluation and validation of the EMT-related lncRNA signature

A scatter plot was made to present OS status of LUAD patients based on the risk scores in the training dataset (**Figure 3A-C**). The results showed the high-risk group was related to more deaths. The scatter plot and heatmap presented consistent results in the testing dataset (**Figure 3D-F**), indicating a robust and stable predictive capacity of the signature.

Additionally, Kaplan-Meier curve analysis demonstrated that the low-risk group had significantly higher OS and PFS rates than the high-risk group both in the training (**Figure 4A**, **4B**) and testing datasets (**Figure 4D**, **4E**). ROC analysis based on the risk signature showed AUC values of 1, 3, and 5-year survival of 0.655, 0.691, and 0.664 in the training dataset (**Figure 4C**), and 0.829, 0.705, and 0.723 in the

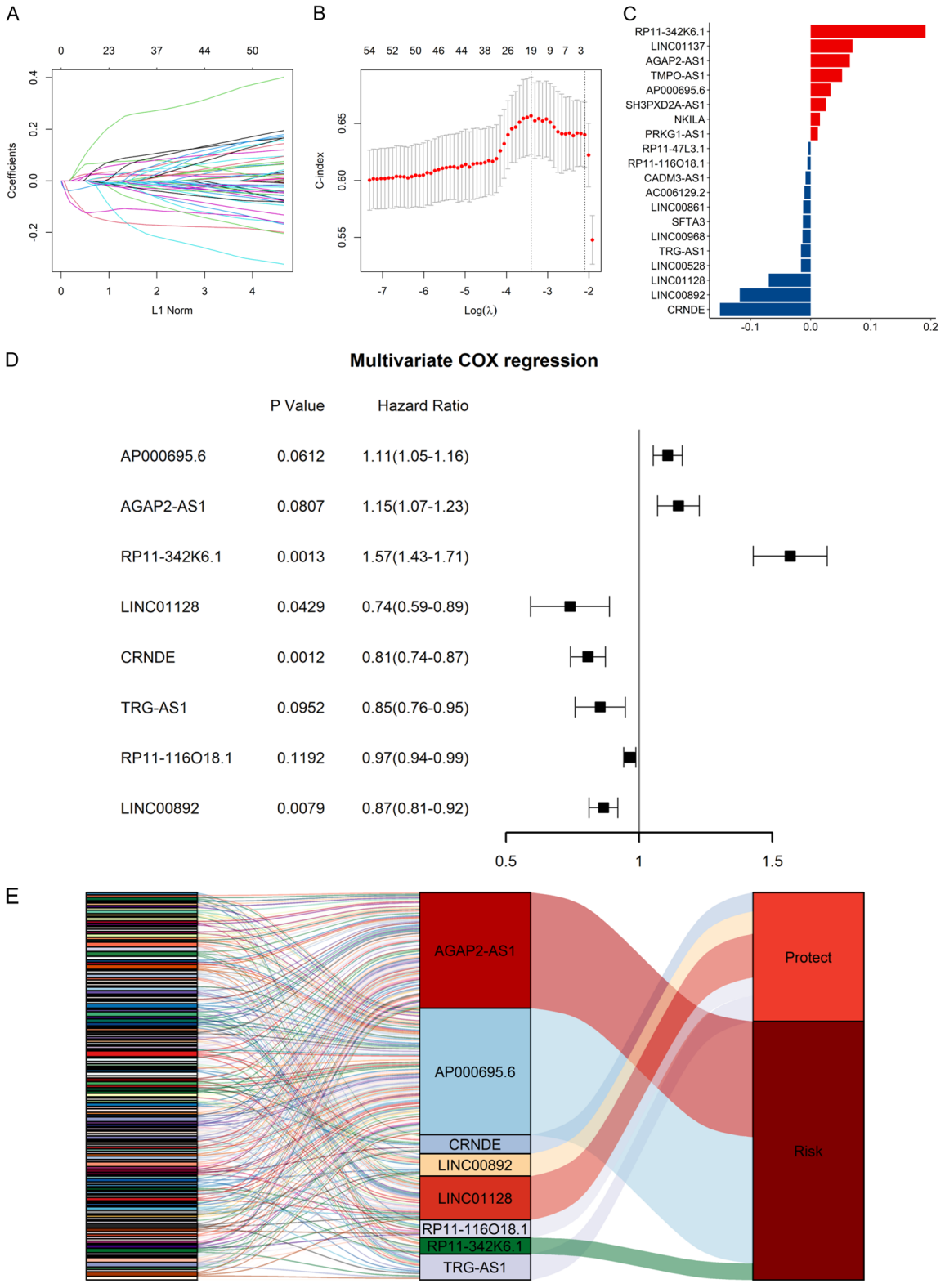

Figure 2. LASSO coefficient profiles of 54 EMT-related prognostic lncRNAs. Each curve corresponds to a lncRNA (A). Selection of the optimal parameter in LASSO regression with 10-fold cross-validation (B). Regression coefficient of each signature lncRNA (C). Multivariate Cox regression analysis of the eight signature lncRNAs (D). Sankey diagram of lncRNAs and EMT-related genes as well as risk types (E).

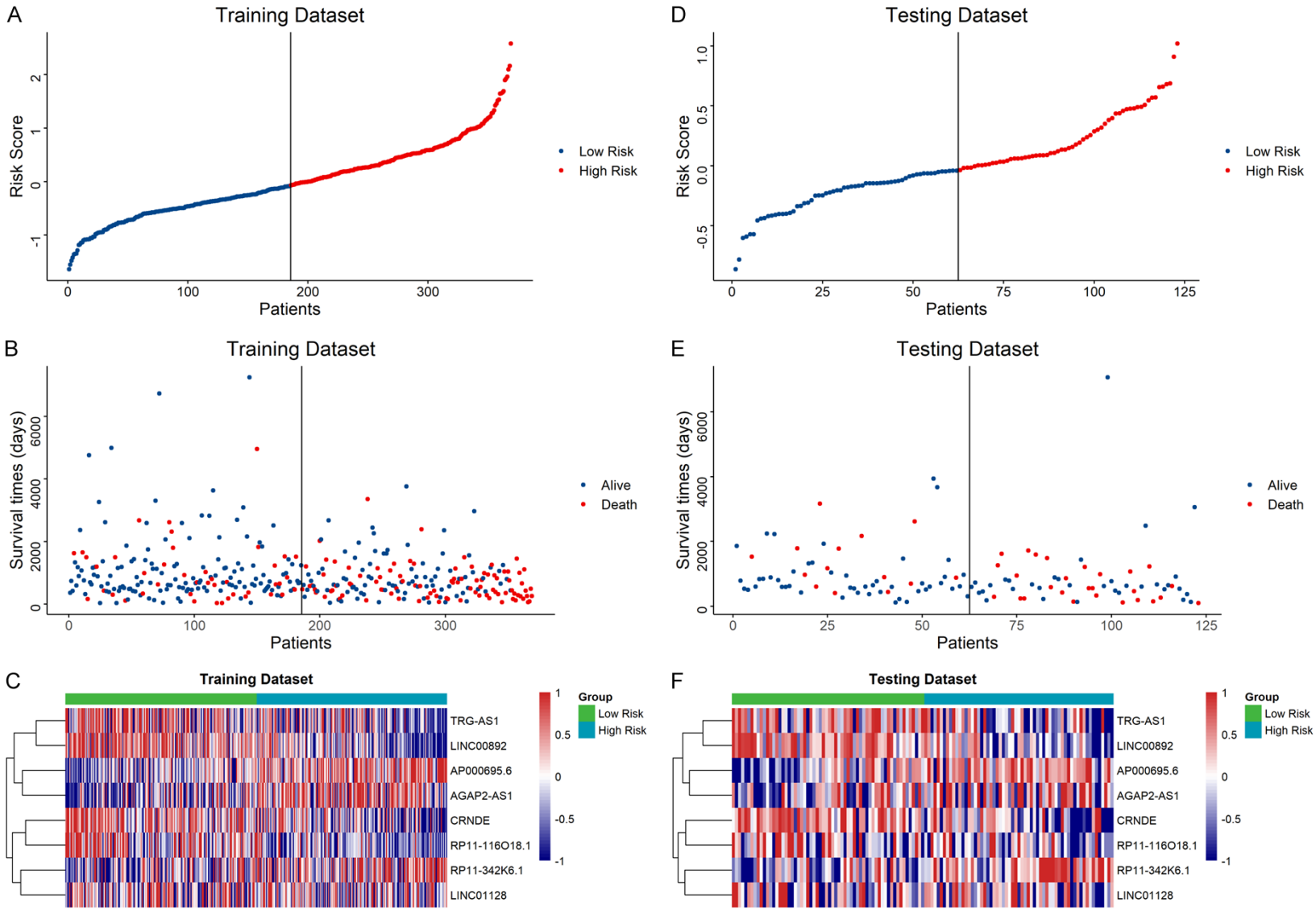

Figure 3. Distributions of risk scores and survival status of LUAD samples in the training (A, B) and testing datasets (D, E). The heatmaps of the signature lncRNAs in the training (C) and teating (F) datasets.

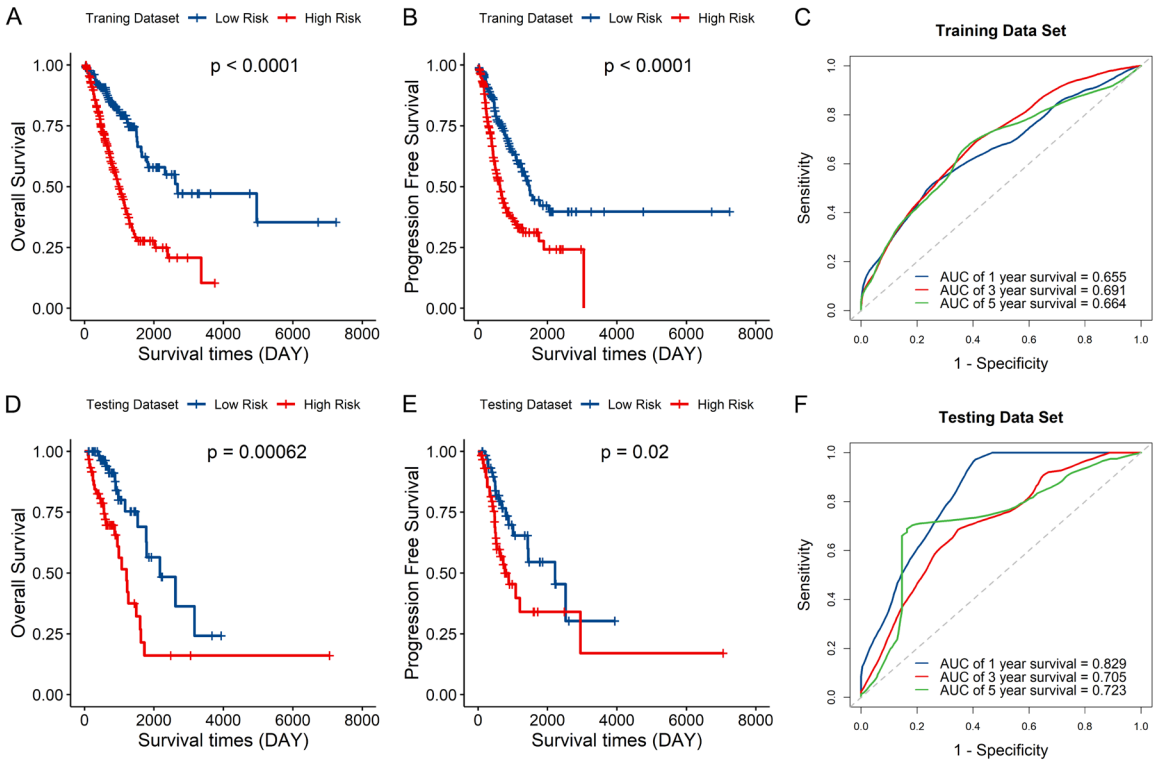

Figure 4. Kaplan-Meier curves showed that the high-risk group had a worse overall survival and progression-free survival than the low-risk group both in the training (A, B) and testing datasets (D, E). Receiver operating characteristic (ROC) curves of the signature for predicting the 1, 3, and 5-year survival in the training (C) and testing datasets (F).

testing dataset (**Figure 4F**). The results revealed that the EMT-related lncRNA prognostic signature had a robust reliability and accuracy in predicting the prognosisof LUAD.

Correlation of the EMT-related prognostic signature and clinicopathologic features

To evaluate the correlation of the risk scores and EMT, we used the chi-square test in the training (**Figure 5A**) and testing (**Figure 5B**) datasets. Results showed that the high-risk group had significantly higher EMT scores than the low-risk group both in the training and testing datasets (P<0.05), therefore the high-risk group confers metastasis and invasion properties on cancer cells by enhancing resistance to apoptotic stimuli, aggressiveness, and mobility. Similarly, the risk scores of patients with the ages ≥65 years, stages III-IV, T3-T4, and N2-N3, were significantly higher than those with the ages <65 years, and stages I-II, N0-N1, T1-T2, respectively (P<0.05) (**Figure 5C**, **5E-G**). However, no significant difference was observed in the risk scores between the M0 and M1 patients, or gender subgroups (**Figure 5D**, **5H**).

To further assess the clinical value and application of the predictive signature, Kaplan-Meier survival curve analysis was performed to compare discrepancies between the risk groups and clinicopathologic features, including age (<65/≥65 years, **Figure 6A**, **6B**), gender (female/male, **Figure 6C**, **6D**), tumor stage (stage I/II/III/IV, **Figure 6E**, **6F**), and TNM stage (T1/T2/T3/T4, N0/N1/N2/N3, and M0/M1, **Figure 6G-L**). The survival rates of patients between the two risk groups were significantly different in age, gender, tumor stage, and TNM stage (P<0.05). The high-risk group had poorer survival outcomes than the low-risk group in almost all the subgroups. However, no significant difference was found in N stage (N2-N3), probably because of the small sample size. The results suggest the high-risk group tended to be associated with more advanced tumor stage, invasion, and metastasis than the low-risk group.

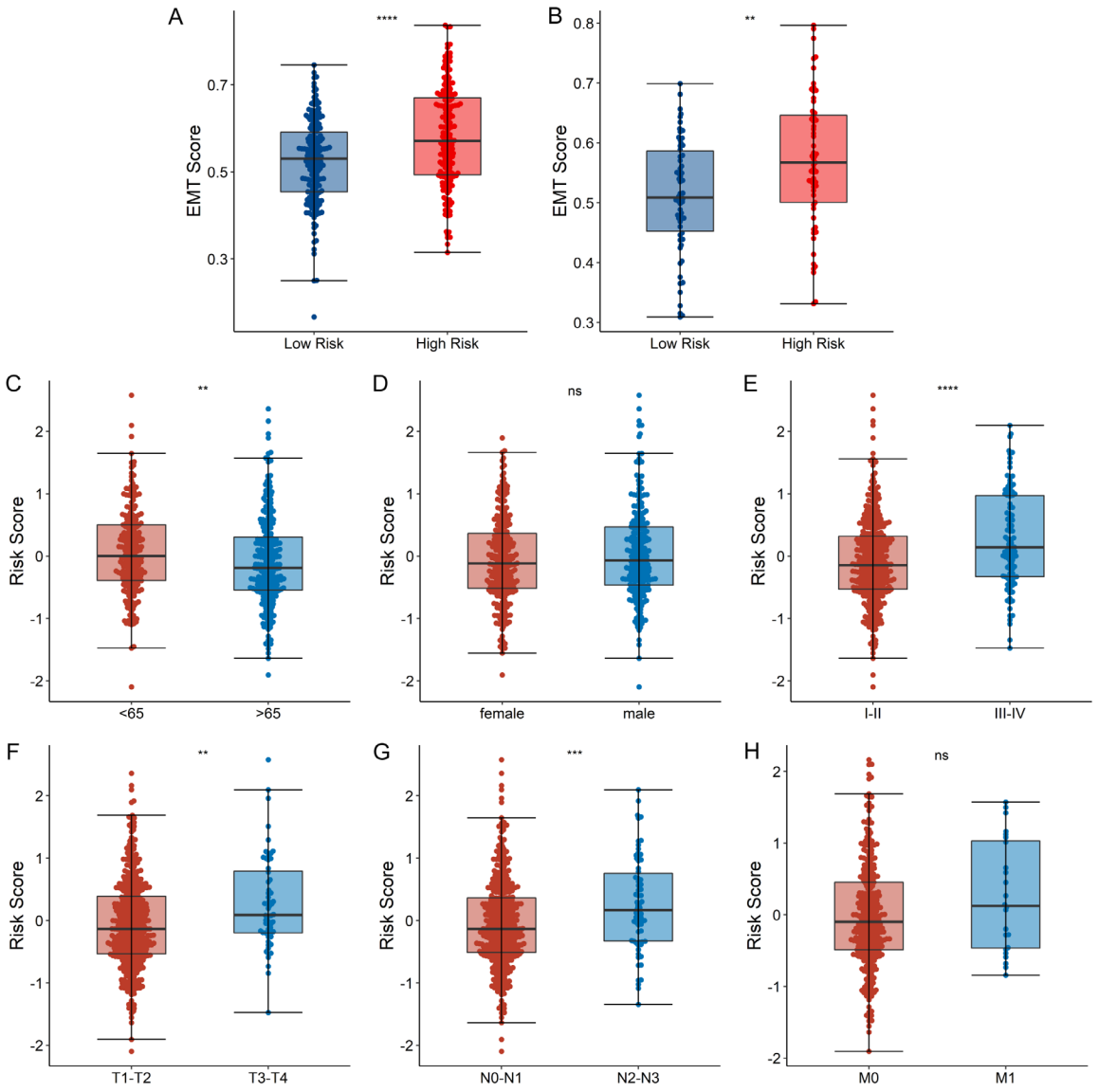

Figure 5. The correlation of the risk scores and EMT in the training (A) and testing (B) datasets. Patients with different clinicopathologic features, including age (C), tumor stage (E), T stage (F), and N stage (G), but not gender (D) and M stage (H), had different levels of risk scores.

Independent prognostic value of the signature in LUAD

Univariate Cox regression analysis indicated that risk score, tumor stage, and TNM stage were significantly correlated with the OS of LUAD patients (P<0.05, **Figure 7A**). Multivariate Cox regression analysis showed that the risk score was an independent prognostic factor for LUAD patients after adjusting for these clinical values, although tumor stage was also independently prognostic (P<0.05, **Figure 7B**). For a more accurate evaluation of the signature, we conducted a nomogram model based on the traditional independent clinical variables and risk score (**Figure 7C**). The calibration curves at 1-year, 3-years, and 5-years presented high consistency between the predicted and actual survival rates (**Figure 7D-F**). ROC curves showed the AUCs for 1-year, 3-year and 5-year OS were 0.724, 0.753, and 0.686, respectively (**Figure 7G**). The results suggested a robust predictive capacity of the prognostic signature.

Principal component analysis (PCA) and functional annotation of EMT-related lncRNAs

We performed PCA to determine the differences between the two risk groups according to EMT genes, EMT-related lncRNAs and overall

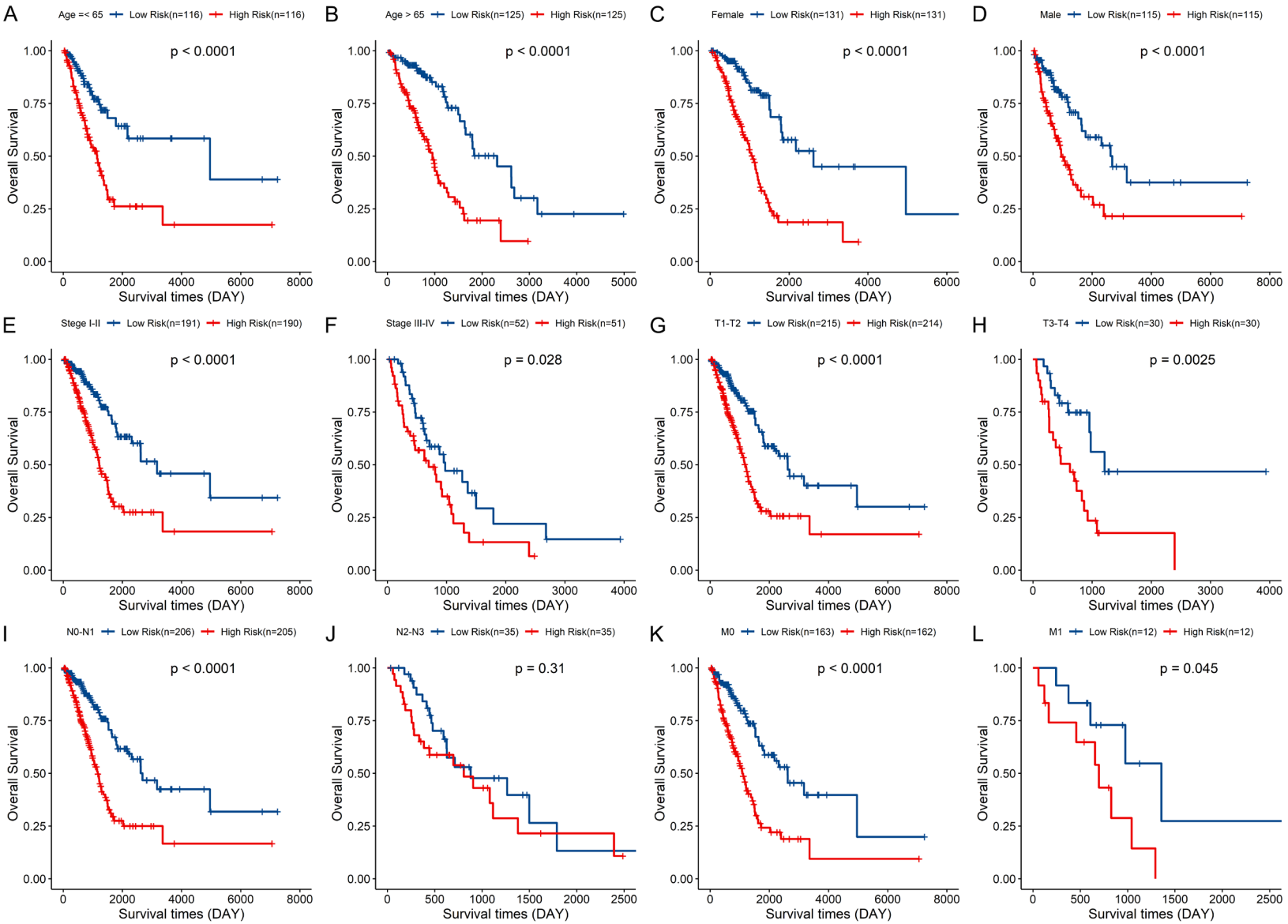

Figure 6. Kaplan-Meier curves showed that the signature retained its prognostic value in of LUAD samples with different clinicopathologic features, including age (A, B), gender (C, D), tumor stage (E, F), T stage (G, H), N0-N1 stage (I), and M stage (K, L), but not N2-N3 stage (J).

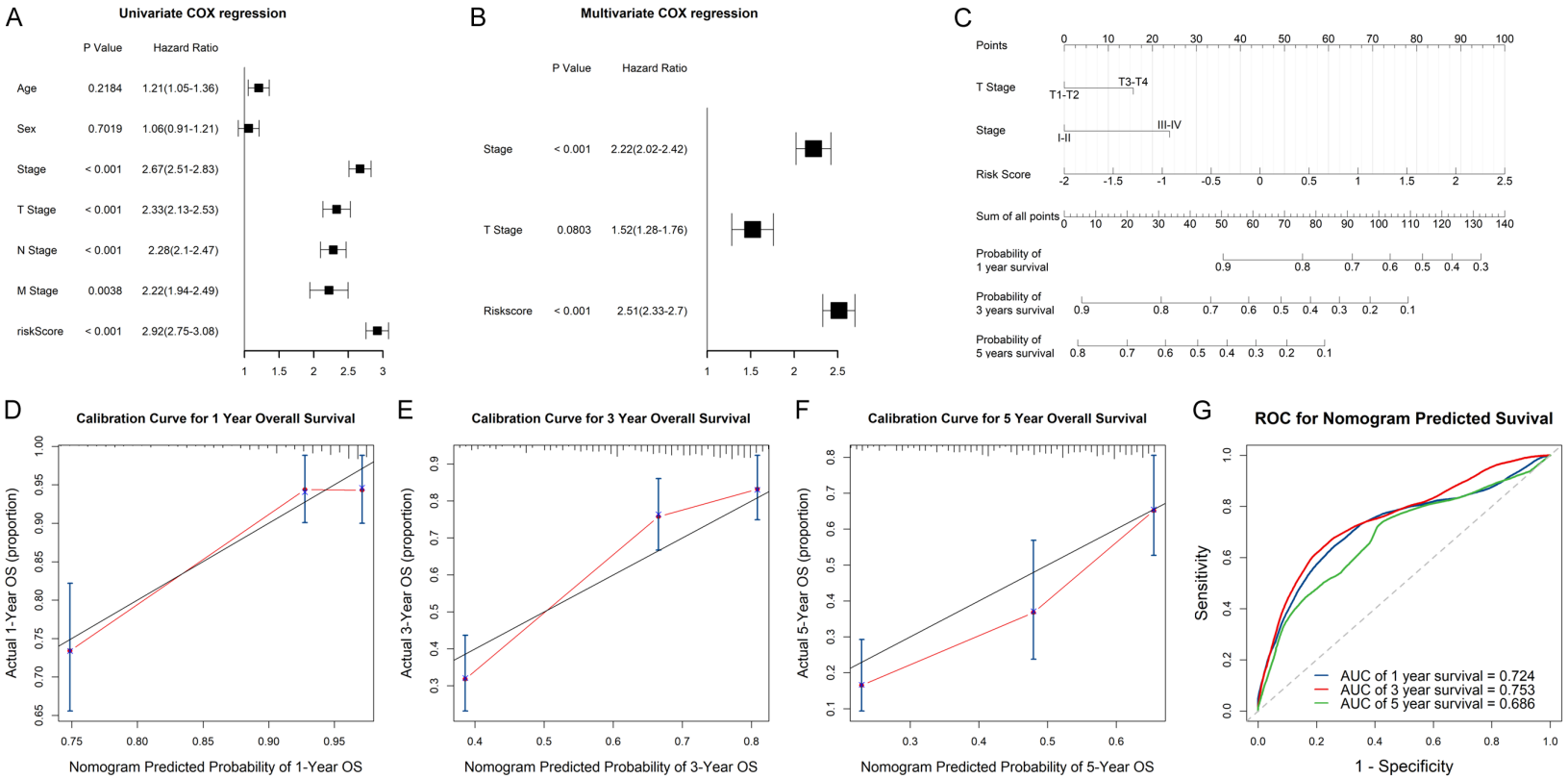

Figure 7. Univariate and multivariate Cox regression analyses revealed that the risk score was an independent prognostic predictor in the training dataset (A, B). Nomogram based on risk score, tumor stage, and T stage (C). Calibration plots of the nomogram for predicting the probability of OS at 1, 3, and 5-year (D-F). Time-dependent receiver operating characteristic (ROC) curves for the Nomogram to predict 1, 3, and 5-year overall survival (G).

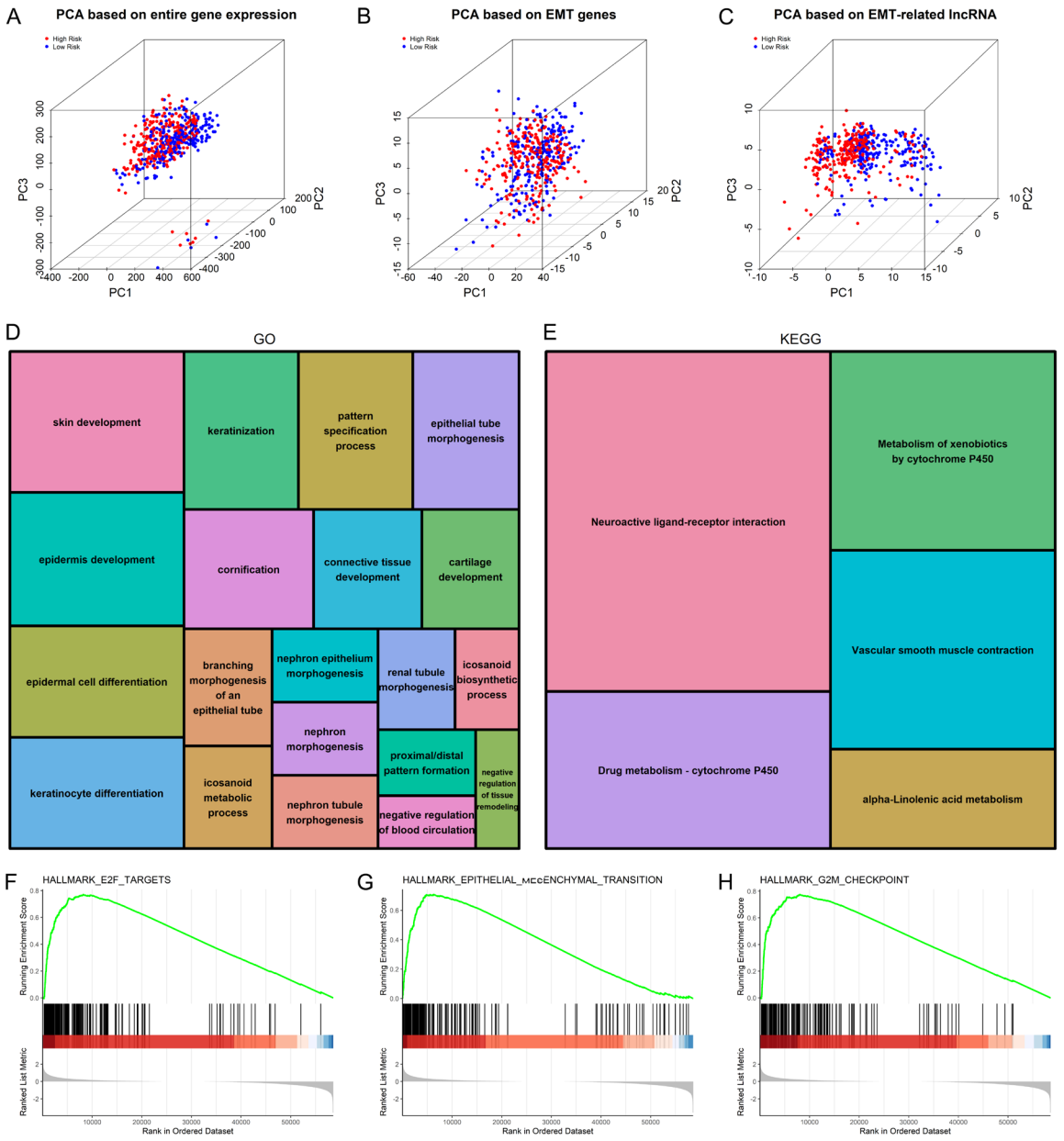

Figure 8. Principal component analysis (PCA) based on entire set (A), EMT genes (B), and EMT-related lncRNA (C). Gene ontology (GO) biologic process analysis (D), Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis (E), and Gene Set Enrichment Analysis (GSEA) (F-H) showing the differentially expressed genes (DEGs) between the two risk groups.

gene expression profiles. The results suggest that the lncRNA signature can distinguish between the two risk groups (**Figure 8A-C**).

GO analyses showed that the DEGs between the two risk groups were enriched in the following GO terms: keratinocyte differentiation, epidermal development, T cell receptor complex plasma membrane signaling receptor complex, and endopeptidase regulator activity (**Figure 8D**). KEGG analysis demonstrated that metabolism of xenobiotics by cytochrome P450, drug metabolism-cytochrome P450, neuroactive ligand-receptor interaction, and vascular smooth muscle contraction were significantly enriched (**Figure 8E**). Finally, GSEA was used to explore the probable biologic processes and pathways in the two risk groups. The biologic pathways including E2F targets, EMT, and G2M checkpoint were identified with the altered genes in the high-risk group (**Figure 8F-H**). Above all, cancer-related pathways and EMT

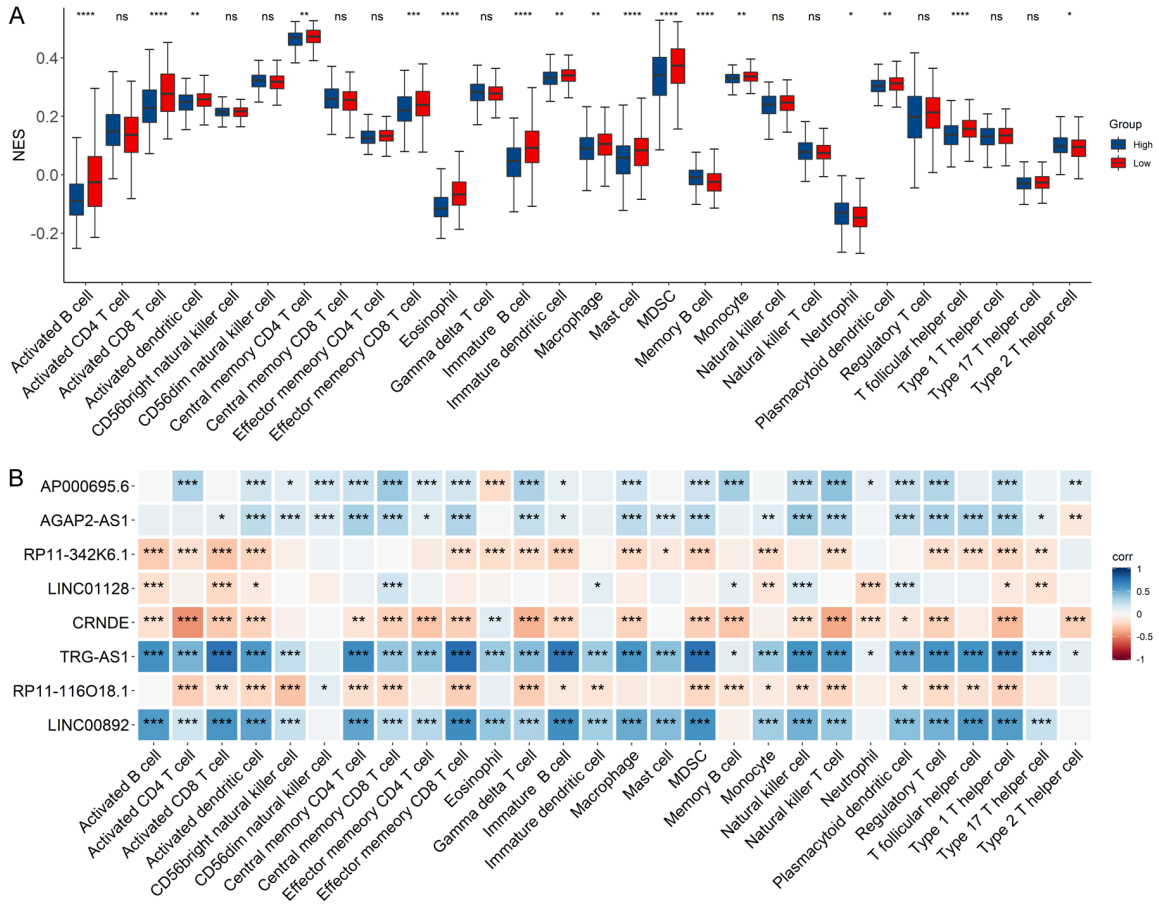

Figure 9. Correlations between the signature and each immune cell infiltration score (A). Correlations between each signature gene and each immune cell infiltration score (B).

were enriched in the high-risk group, suggesting significant correlation of high-risk scores with EMT and cancer progression.

Correlation of the prognostic signature and immune cell infiltration

To determine the relationship between the risk score and immune cell infiltration, the relative quantities of 28 types of immune cells were compared between the two risk groups. There were significant differences in many types of immune cells between the two groups, and the relative quantity of immune cells in the high-risk group was significantly lower than that of the low-risk group. Specifically, the relative quantity of CD8 T cells was significantly lower in the high-risk group than that in the low-risk group, suggesting a weaker ability to kill cancer cells in the former group (**Figure 9A**). A correlation analysis between immune cell infiltration and each lncRNA showed that expression levels of AP000695.6, AGAP2-AS1, TRG-AS1, and LINC00892 were significantly and positively related to the infiltration level of most immune cell types. However, the expression levels of LINC01128, CRNDE, RP11-342K6.1, and RP11-116O18.1 were significantly and negatively associated with the infiltration level of most immune cells (**Figure 9B**). These results showed that the signature was significantly correlated with infiltration by immune cells.

Construction of a ceRNA network and functional enrichment analysis

To explore the interactions of lncRNAs, mRNAs and miRNAs, a ceRNA network based on 5 lncRNAs, 8 miRNAs, and 217 target mRNAs was constructed (**Figure 10A**). Additionally, the top 10 hub genes were acquired from the PPI (**Figure 10B**). GO analysis showed that the top GO terms included positive regulation of protein serine/threonine kinase activity, lung de-

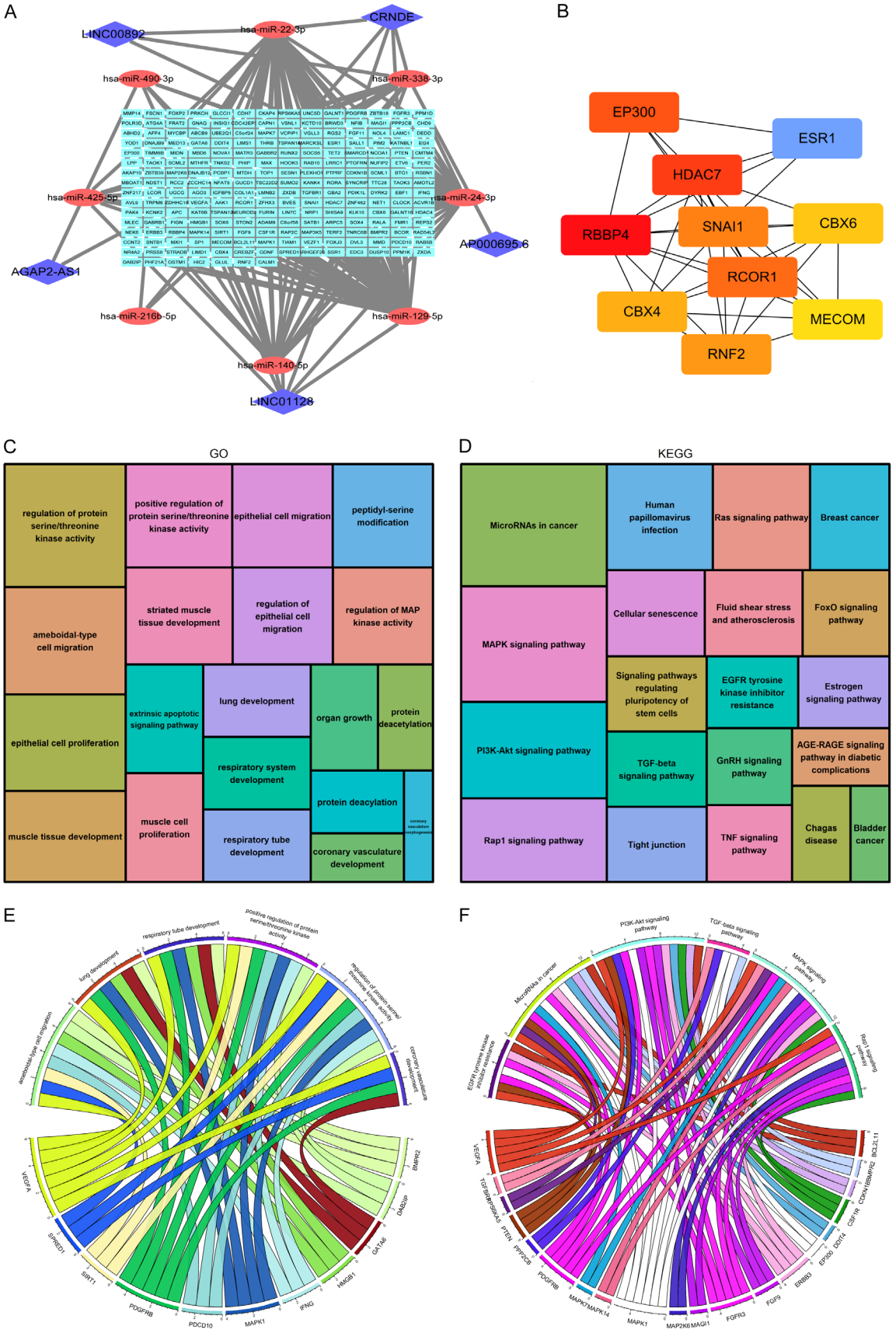
A
LINC00892
hsa-miR-22-3p
CRNDE
hsa-miR-490-3p
hsa-miR-338-3p
hsa-miR-425-5p
hsa-miR-24-3p
AP000695.6
AGAP2-AS1
hsa-miR-216b-5p
hsa-miR-129-5p
hsa-miR-140-5p
LINC01128
B
EP300
ESR1
HDAC7
RBBP4
SNAI1
CBX6
RCOR1
CBX4
MECOM
RNF2
C
GO
regulation of protein serine/threonine kinase activity
positive regulation of protein serine/threonine kinase activity
epithelial cell migration
peptidyl-serine modification
striated muscle tissue development
regulation of epithelial cell migration
regulation of MAP kinase activity
ameboidal-type cell migration
extrinsic apoptotic signaling pathway
lung development
organ growth
protein deacetylation
epithelial cell proliferation
respiratory system development
muscle cell proliferation
protein deacylation
muscle tissue development
respiratory tube development
coronary vasculature development
D
KEGG
MicroRNAs in cancer
Human papillomavirus infection
Ras signaling pathway
Breast cancer
MAPK signaling pathway
Cellular senescence
Fluid shear stress and atherosclerosis
FoxO signaling pathway
Signaling pathways regulating pluripotency of stem cells
EGFR tyrosine kinase inhibitor resistance
Estrogen signaling pathway
PI3K-Akt signaling pathway
TGF-beta signaling pathway
GnRH signaling pathway
AGE-RAGE signaling pathway in diabetic complications
Rap1 signaling pathway
Tight junction
TNF signaling pathway
Chagas disease
Bladder cancer
E
F

Figure 10. The ceRNA network of five EMT-related lncRNAs (blue) and their target miRNAs (red) and mRNAs (green) (A). Network diagram of top 10 hub genes (B). Functional analysis of 217 target mRNAs: the dotplot based on top 20 Gene ontology (GO) biological process analysis (C) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis (D). Network diagram between GO biologic processes (E), KEGG pathways (F), and mRNA.

velopment, amaroidal-type cell migration, extrinsic apoptotic signaling pathway, and peptidyl-serine modification (**Figure 10C**). The top KEGG signaling pathways were MAPK, Rap1, TGF-beta, PI3K-Akt signaling pathway, and EGFR tyrosine kinase inhibitor resistance (**Figure 10D**). We also identified genes in significantly related functions and pathways. Several important genes such as VEGFA, PTEN, and MAPK1, were highlighted (**Figure 10E**, **10F**).

Discussion

Lung adenocarcinoma (LUAD) is the most frequent histologic subtype of lung cancer, and most patients at the time of diagnosis are at an advanced stage with metastasis and poor prognosis. Thus, it is essential to define new biomarkers to improve the management of LUAD [13]. EMT can play a critical role in the biologic processes, including carcinogenesis, migration, metastasis, and drug resistance [14]. lncRNAs are a diverse class of RNAs involved in numerous biologic processes, including chromatin modification, imprinting, transcription, posttranslational processing, and regulation of gene expression [8]. The lncRNAs that act as decoys, scaffolds, and enhancer RNAs have been evaluated as probable diagnostic and prognostic biomarkers in cancers and provide targets for therapeutic application [15]. While most studies have mainly focused on the function of specific EMT and lncRNAs in the occurrence and progression of cancers, the clinical value of EMT-related lncRNAs have not been fully explored in cancers, particularly in LUAD [16-19].

Herein, we established a novel prognostic risk signature with eight EMT-related lncRNAs for LUAD, and validated that the prognostic signature possesses reliable predictive capability. Univariate and multivariate Cox regression analysis illustrated that the risk score based on the signature is an independent prognostic predictor in LUAD.

In the signature, AP000695.6, AGAP2-AS1, and RP11-342K6.1 were unfavorable lncRNAs, whereas, LINC00892, LINC01128, TRG-AS1, CRNDE, and RP11-116O18.1 were protective lncRNA (Figure S2). The roles, functions, and working mechanisms, of most lncRNAs in the signature have been investigated in many types of cancers. AP000695.6 has been identified as a prognostic risk signature in gastric carcinoma [20]. LINC00892 has been identified as a prognostic risk signature in bladder cancer [21, 22]. LINC01128 has been reported to facilitate osteosarcoma by functioning as a sponge of miR-299-3p [23]. In addition, AGAP2-AS1 [24, 25], TRG-AS1 [26, 27] and CRNDE [28, 29] have been reported to act as competitive endogenous RNA to promote proliferation and metastasis by sponging different miRNAs in hepatocellular carcinoma, colorectal cancer and lung cancer, respectively. The other two lncRNAs, RP11-116O18.1 and RP11-342K6.1, have not been reported on yet.

The molecular mechanism remains to be further explored despite the excellent prognostic capacity of the eight EMT-related lncRNAs in LUAD. Therefore, we also performed GO, KEGG, and GSEA analysis to explore biological pathways and processes in lncRNAs. Results displayed that EMT and malignancy-related pathways were significantly enriched in the high-risk group, suggesting a high correlation of the risk score with tumor metastasis and progression.

Immune cell infiltration in the tumor microenvironment (TME), particularly T cells, is a key mediator of tumor destruction and plays an important role in immunotherapy [30]. EMT-related lncRNAs had been reported to affect the prognosis of patients through immune mechanisms in bladder cancer [31], colorectal cancer [32] and melanoma [33]. Nevertheless, the function and prognostic value of EMT-related lncRNAs in mediating immune infiltration and prognosis of LUAD remain to be characterized. In this study, a correlation analysis of risk scores and immune-cell infiltration demonstrated the number of immune cells in the low-risk group was significantly higher than that of the high-risk group, and this contributed to a prolonged survival outcome.

Finally, a ceRNA network was built to study possible downstream targets and mechanisms of EMT-related lncRNAs in LUAD. The results showed some important genes such as VEGFA, PTEN and MAPK1 that are included in MAPK, Rap1, TGF-beta, PI3K-Akt signaling pathways may mediate the function of EMT-related lncRNAs in LUAD. Precise molecular mechanisms need to be further elucidated.

However, there are some limitations that should be noted in this study. First, the study is based on a public dataset. Second, the sample size from the LUAD cohort from TCGA database is still limited. The signature has not been validated in other databases and requires independent external validation to assess its clinical relevance. Finally, further biochemical experiments including quantitative polymerase chain reaction (qPCR), western blot analysis, and clinical data analyses are needed to further confirm the signature.

Thus, we identified EMT-related lncRNAs significantly associated with survival outcome and constructed a prognostic risk signature with robust predictive potential in LUAD. The findings offer valuable insights for future study into prognostic biomarkers and therapeutic targets to enable individualized treatments for LUAD patients in different risk categories.

Acknowledgements

We are grateful to the contributors to the public databases used in this study.

Disclosure of conflict of interest

None.

Address correspondence to: Dr. Xiaoju Zhang, Department of Respiratory and Critical Care Medicine, Zhengzhou University People's Hospital, China; Henan Provincial People's Hospital, Weiwu Road No. 7, Zhengzhou 450003, Henan, China. E-mail: 15837101166@163.com

References

[1] Siegel RL, Miller KD, Fuchs HE and Jemal A. Cancer statistics, 2021. CA Cancer J Clin 2021; 71: 7-33.
[2] Balzer BWR, Loo C, Lewis CR, Trahair TN and Anazodo AC. Adenocarcinoma of the lung in childhood and adolescence: a systematic review. J Thorac Oncol 2018; 13: 1832-1841.
[3] Devarakonda S, Morgensztern D and Govindan R. Genomic alterations in lung adenocarcinoma. Lancet Oncol 2015; 16: e342-351.
[4] Lamouille S, Xu J and Derynck R. Molecular mechanisms of epithelial-mesenchymal transition. Nat Rev Mol Cell Biol 2014; 15: 178-196.
[5] Dongre A and Weinberg RA. New insights into the mechanisms of epithelial-mesenchymal transition and implications for cancer. Nat Rev Mol Cell Biol 2019; 20: 69-84.
[6] Mittal V. Epithelial mesenchymal transition in tumor metastasis. Annu Rev Pathol 2018; 13: 395-412.
[7] Goodall GJ and Wickramasinghe VO. RNA in cancer. Nat Rev Cancer 2021; 21: 22-36.
[8] Quinn JJ and Chang HY. Unique features of long non-coding RNA biogenesis and function. Nat Rev Genet 2016; 17: 47-62.
[9] Sanchez Calle A, Kawamura Y, Yamamoto Y, Takeshita F and Ochiya T. Emerging roles of long non-coding RNA in cancer. Cancer Sci 2018; 109: 2093-2100.
[10] O'Brien SJ, Bishop C, Hallion J, Fiechter C, Scheurlen K, Paas M, Burton J and Galandiuk S. Long non-coding RNA (lncRNA) and epithelial-mesenchymal transition (EMT) in colorectal cancer: a systematic review. Cancer Biol Ther 2020; 21: 769-781.
[11] Li N, Li J, Mi Q, Xie Y, Li P, Wang L, Binang H, Wang Q, Wang Y, Chen Y, Wang Y, Mao H, Du L and Wang C. Long non-coding RNA ADAMTS9-AS1 suppresses colorectal cancer by inhibiting the Wnt/β-catenin signalling pathway and is a potential diagnostic biomarker. J Cell Mol Med 2020; 24: 11318-11329.
[12] Liang H, Yu T, Han Y, Jiang H, Wang C, You T, Zhao X, Shan H, Yang R, Yang L, Shan H and Gu Y. Correction to: LncRNA PTAR promotes EMT and invasionmetastasis in serous ovarian cancer by competitively binding miR-101-3p to regulate ZEB1 expression. Mol Cancer 2021; 20: 64.
[13] Bade BC and Dela Cruz CS. Lung Cancer 2020: epidemiology, etiology, and prevention. Clin Chest Med 2020; 41: 1-24.
[14] Shibue T and Weinberg RA. EMT, CSCs, and drug resistance: the mechanistic link and clinical implications. Nat Rev Clin Oncol 2017; 14: 611-629.
[15] Ransohoff JD, Wei Y and Khavari PA. The functions and unique features of long intergenic non-coding RNA. Nat Rev Mol Cell Biol 2018; 19: 143-157.
[16] Li J, Li Z, Zheng W, Li X, Wang Z, Cui Y and Jiang X. LncRNA-ATB: an indispensable cancer-related long noncoding RNA. Cell Prolif 2017; 50: e12381.
[17] Lei H, Gao Y and Xu X. LncRNA TUG1 influences papillary thyroid cancer cell proliferation, migration and EMT formation through target-

ing miR-145. Acta Biochim Biophys Sin (Shanghai) 2017; 49: 588-597.
[18] Gong J, Wang Y and Shu C. LncRNA CHRF promotes cell invasion and migration via EMT in gastric cancer. Eur Rev Med Pharmacol Sci 2020; 24: 1168-1176.
[19] Wang X, Lai Q, He J, Li Q, Ding J, Lan Z, Gu C, Yan Q, Fang Y, Zhao X and Liu S. LncRNA SNHG6 promotes proliferation, invasion and migration in colorectal cancer cells by activating TGF-β/Smad signaling pathway via targeting UPF1 and inducing EMT via regulation of ZEB1. Int J Med Sci 2019; 16: 51-59.
[20] Cheng P. A prognostic 3-long noncoding RNA signature for patients with gastric cancer. J Cell Biochem 2018; 119: 9261-9269.
[21] Lin G, Guo B, Wei Y, Lan T, Wen S and Li G. Impact of long non-coding RNAs associated with microenvironment on survival for bladder cancer patients. Front Genet 2020; 11: 567200.
[22] Wu Y, Zhang L, He S, Guan B, He A, Yang K, Gong Y, Li X and Zhou L. Identification of immune-related LncRNA for predicting prognosis and immunotherapeutic response in bladder cancer. Aging (Albany NY) 2020; 12: 23306-23325.
[23] Yao Q and Chen T. LINC01128 regulates the development of osteosarcoma by sponging miR-299-3p to mediate MMP2 expression and activating Wnt/beta-catenin signalling pathway. J Cell Mol Med 2020; 24: 14293-14305.
[24] Hong S, Yan Z, Song Y, Bi M and Li S. LncRNA AGAP2-AS1 augments cell viability and mobility, and confers gemcitabine resistance by inhibiting miR-497 in colorectal cancer. Aging (Albany NY) 2020; 12: 5183-5194.
[25] Liu Z, Wang Y, Wang L, Yao B, Sun L, Liu R, Chen T, Niu Y, Tu K and Liu Q. Long non-coding RNA AGAP2-AS1, functioning as a competitive endogenous RNA, upregulates ANXA11 expression by sponging miR-16-5p and promotes proliferation and metastasis in hepatocellular carcinoma. J Exp Clin Cancer Res 2019; 38: 194.
[26] Sun X, Qian Y, Wang X, Cao R, Zhang J, Chen W and Fang M. LncRNA TRG-AS1 stimulates hepatocellular carcinoma progression by sponging miR-4500 to modulate BACH1. Cancer Cell Int 2020; 20: 367.
[27] Zhang M, Zhu W, Haeryfar M, Jiang S, Jiang X, Chen W and Li J. Long Non-coding RNA TRG-AS1 promoted proliferation and invasion of lung cancer cells through the miR-224-5p/SMAD4 axis. Onco Targets Ther 2021; 14: 4415-4426.
[28] Han P, Li JW, Zhang BM, Lv JC, Li YM, Gu XY, Yu ZW, Jia YH, Bai XF, Li L, Liu YL and Cui BB. The lncRNA CRNDE promotes colorectal cancer cell proliferation and chemoresistance via miR-181a-5p-mediated regulation of Wnt/beta-catenin signaling. Mol Cancer 2017; 16: 9.
[29] Lu Y, Sha H, Sun X, Zhang Y, Wu Y, Zhang J, Zhang H, Wu J and Feng J. CRNDE: an oncogenic long non-coding RNA in cancers. Cancer Cell Int 2020; 20: 162.
[30] Lei X, Lei Y, Li JK, Du WX, Li RG, Yang J, Li J, Li F and Tan HB. Immune cells within the tumor microenvironment: biological functions and roles in cancer immunotherapy. Cancer Lett 2020; 470: 126-133.
[31] Du Y, Wang B, Jiang X, Cao J, Yu J, Wang Y, Wang X and Liu H. Identification and validation of a stromal EMT related LncRNA signature as a potential marker to predict bladder cancer outcome. Front Oncol 2021; 11: 620674.
[32] Liu C, Hu C, Li J, Jiang L and Zhao C. Identification of epithelial-mesenchymal transition-related lncRNAs that associated with the prognosis and immune microenvironment in colorectal cancer. Front Mol Biosci 2021; 8: 633951.
[33] Xiao B, Liu L, Chen Z, Li A, Wang P, Xiang C, Zeng Y, Li H and Xiao T. Identification of epithelial-mesenchymal transition-related prognostic lncRNAs biomarkers associated with melanoma microenvironment. Front Cell Dev Biol 2021; 9: 679133.

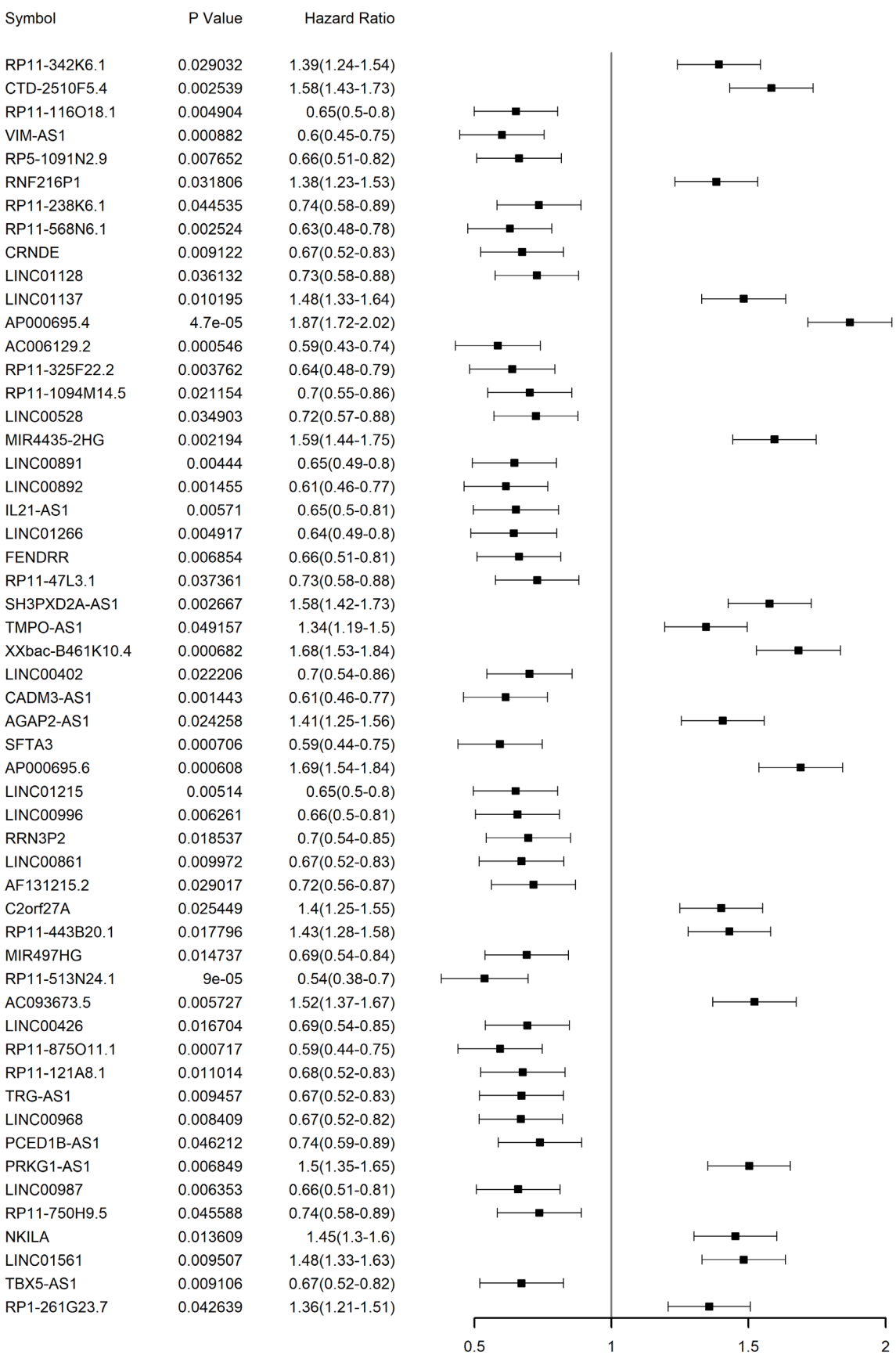

Figure S1. 54 EMT-related lncRNAs with significant prognostic values determined by univariate Cox regression analysis.

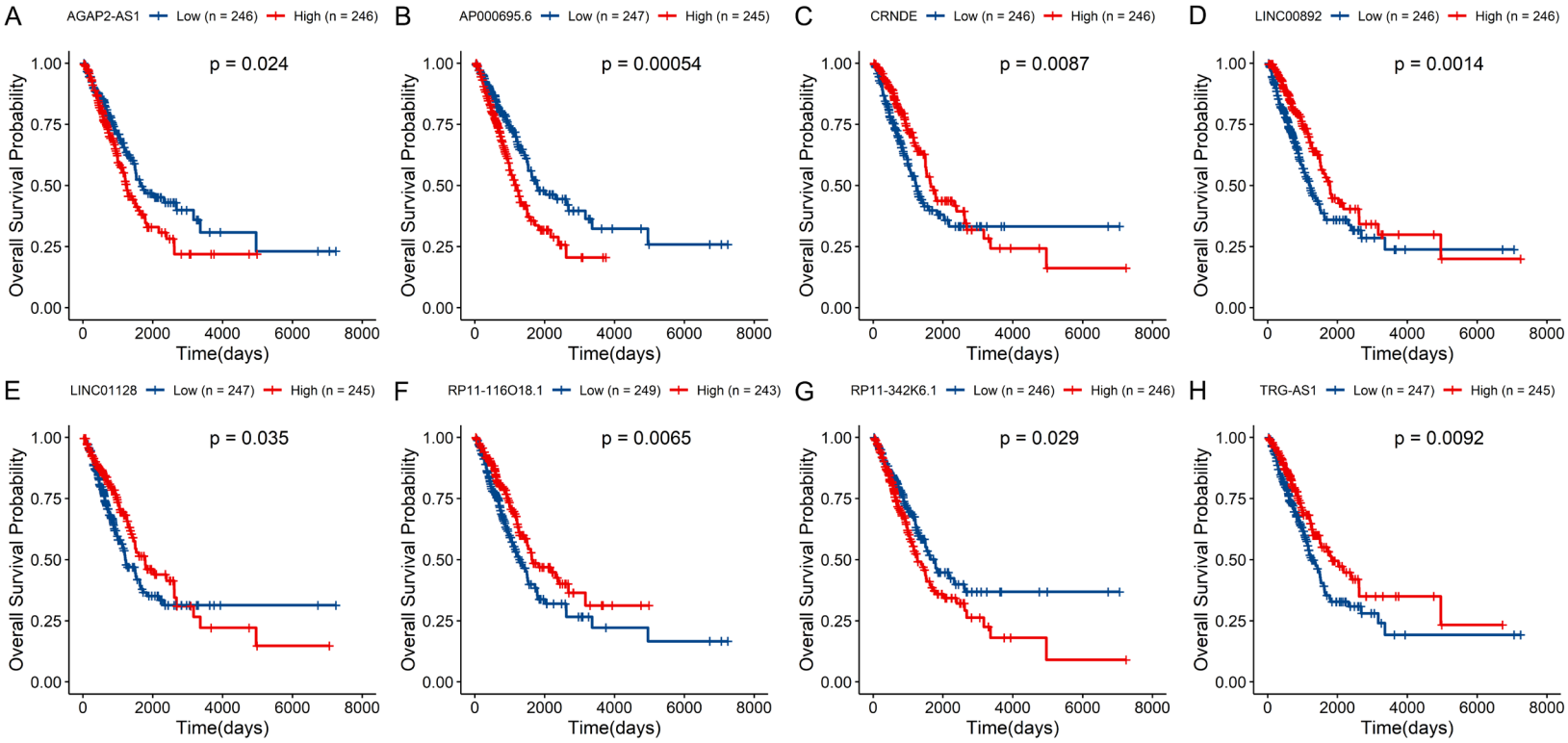

Figure S2. Survival analysis in the two subgroups of patients classified by the expression level of each gene in the signature.