



A six-long noncoding RNA model predicts prognosis in lung adenocarcinoma

Yuquan Bai¹, Senyi Deng^{1,2}

¹Department of Thoracic Surgery research laboratory, West China Hospital, Sichuan University, Chengdu, China; ²Western China Collaborative Innovation Center for Early Diagnosis and Multidisciplinary Therapy of Lung Cancer, Sichuan University, Chengdu, China

Contributions: (I) Conception and design: All authors; (II) Administrative support: S Deng; (III) Provision of study materials or patients: Y Bai; (IV) Collection and assembly of data: Y Bai; (V) Data analysis and interpretation: All authors; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Senyi Deng, PhD. Department of Thoracic Surgery, West China Hospital, Sichuan University, Chengdu 610041, China. Email: senyi_deng@scu.edu.cn.

Background: The incidence and mortality of lung cancer rank first among various malignant tumors. The lack of clear molecular classification and effective individualized treatment greatly limits the treatment benefits of patients. Long non-coding RNAs (lncRNAs) have been demonstrated widely involve in tumor progressing, and been proved easy to detect for occupying majority in transcriptome. However, less work focuses on studying the potency of lncRNAs as molecular typing and prognostic indicator in lung cancer.

Methods: Based on the 448 lung adenocarcinoma (LUAD) samples and the expression of 14,127 lncRNAs from the Cancer Genome Atlas (TCGA) database, we constructed a co-expression network using weighted gene co-expression network analysis. Then based on the feature module and the overall survival of patients, we constructed a risk score model through Cox proportional hazards regression and verified it with a validation cohort. Finally, according to the median of risk score, the function of this model was enriched.

Results: We identified a module containing 123 lncRNAs that is related with the prognosis of LUAD. Using univariate and multivariate Cox proportional hazards regression with lasso regression, six lncRNAs were identified to construct a risk score model. The calculation formula shown as follows: risk score = $(-0.3057 \times \text{EXP}_{\text{VIM-AS1}}) + (0.9678 \times \text{EXP}_{\text{AC092811.1}}) + (1.0829 \times \text{EXP}_{\text{NF1A-AS1}}) + (-0.3505 \times \text{EXP}_{\text{AL035701.1}}) + (3.9378 \times \text{EXP}_{\text{AC079336.4}}) + (-0.2810 \times \text{EXP}_{\text{AL121790.2}})$. Six-lncRNA model can be used as an independent prognostic indicator in LUAD ($P < 0.001$) and the area under the 5-year receiver operating characteristic (ROC) curve is 0.715.

Conclusions: We developed a six-lncRNA model, which could be used for predicting prognosis and guiding medical treatment in LUAD patients.

Keywords: Long non-coding RNAs (lncRNAs); model, lung adenocarcinoma (LUAD); prognosis

Submitted Jun 29, 2020. Accepted for publication Oct 21, 2020.

doi: 10.21037/tcr-20-2436

View this article at: <http://dx.doi.org/10.21037/tcr-20-2436>

Introduction

Lung cancer is one of the most malignant tumors that poses great threat to the population health (1). According to surgery improving (2), molecular targeted drug developing and the application of immunotherapy, the overall survival (OS) of lung cancer has been significantly improved. Molecular targeted therapy has the characteristics of strong specificity and small side effects (3). The currently marketed molecular targeted drugs for non-small cell lung cancer (NSCLC) mainly include oncogene molecular targeted drugs, anti-angiogenesis drugs, immune targeted therapy drugs, and multi-target inhibitors. With the advancement of genetic testing technology and the wide application of small molecule tyrosine kinase inhibitors (TKIs), most patients with advanced NSCLC have achieved good therapeutic effects in the treatment of TKIs (4). However, some difficult issues still need to be resolved. Epidemiological studies show that lung adenocarcinoma (LUAD) has replaced lung squamous cell carcinoma (LUSC) as the main pathological type in lung cancers, but its pathogenesis and progressive mechanism remain unclear (5-7). Two-thirds of patients with LUAD are diagnosed at advanced stage, and they prone to earn poor prognosis for lacking effective individual therapy (8). Therefore, studying the molecular mechanism of LUAD to identify precise molecular typing markers is urgently needed.

It is well known that about 70% human genome would transcript into RNAs, of which protein-coding sequences account for less than 2%, the rest thousands of transcripts are non-coding RNAs (9,10). Long non-coding RNAs (lncRNAs) are defined more than 200 nucleotides, which occupied majority in non-coding RNAs (11). lncRNAs have been demonstrated involve in cell cycling controlling, cell differentiation mediating, epigenetic regulation and so on (12). Compared with protein-coding genes (PCGs), lncRNAs are composed of fewer exons to existing higher evolutionary conservation (13). Moreover, lncRNAs show more stable feature against degradation as often forming secondary structure. These characters make lncRNAs easy to be detected in body fluids including blood and urine (14).

In recent years, the role of lncRNAs have also been widespread reported in tumorigenesis (15-18). lncRNA UCA1 is associated with poor prognosis in patients with gastric cancer (19). LINC00963 could promote tumorigenesis and radiation resistance of breast cancer by interacting with miR-324-3p (20). lncRNA PCNAP1

could enhance the replication of hepatitis B virus (HBV) and the occurrence of liver cancer (21). In ovarian cancer, lncRNA HOTTIP could indirectly up-regulate the expression of PD-L1, inhibit the activity of T cells and eventually accelerate the immune escape (22). And in lung cancer, the LCAT1-miR-4715-5p-RAC1/PAK1 axis plays an essential role during tumor progression and could be a potential therapeutic target (23). These findings suggest that the variation of lncRNAs are closely related to tumor prognosis. Up to now, most studies only focus on single lncRNA (24), there is still less work aim to analysis the correlation between lncRNAs and tumor prognosis systematically.

In this study, based on the Cancer Genome Atlas (TCGA) database, we analyzed the expression of all lncRNAs by Weighted Correlation Network analysis (WGCNA) to determine prognosis related module. Then, a six-lncRNA model with reliable prognostic value in LUAD was constructed by Cox proportional hazards regression analysis. In addition, we have conducted in-depth studies on the biological functions of this six-lncRNA model. Our results confirmed that this six-lncRNA model could be used to predict OS in LUAD independently, which could couple with traditional clinical prognostic factors to promote LUAD survival. We present the following article in accordance with the TRIPOD reporting checklist (available at <http://dx.doi.org/10.21037/tcr-20-2436>).

Methods

Patients and data pre-processing

A total of 479 LUAD samples containing clinical information were collected from TCGA database (<https://portal.gdc.cancer.gov>). Thirty-one patients with less than 30 days survival time were deleted, leaving 448 patients. We extracted 14,127 lncRNAs from the expression profile, and then screened 25% lncRNAs (n=3,532) with the largest variance differences for subsequent analysis. All data were filtered to reduce outliers. The flow chart of data collection and analysis were shown in *Figure 1*. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

WGCNA

According to abline =650, 21 outlier samples were excluded, and 427 samples were remained. By choosing $\beta = 5$ as the

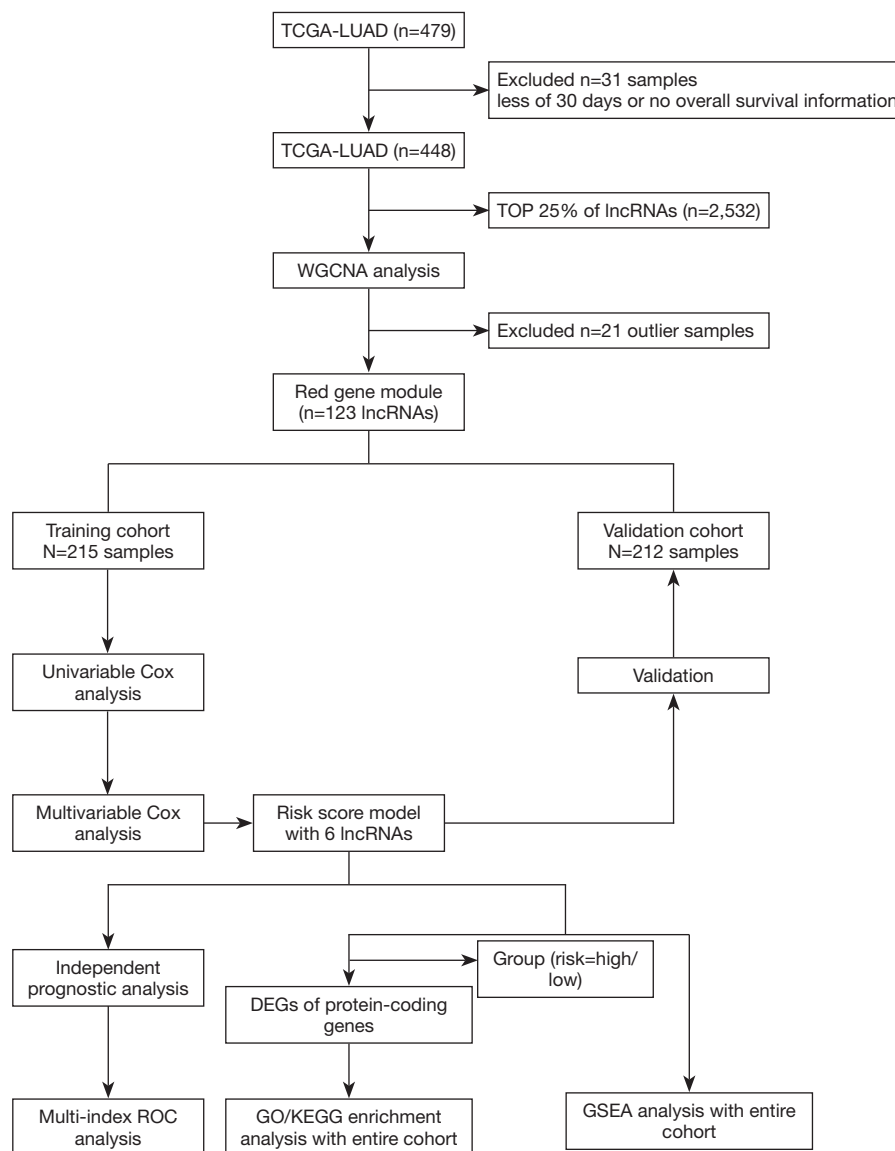


Figure 1 Flow chart of data collection and analysis.

soft threshold, a WGCNA R package (25) was used to construct a weighted gene co-expression network. The topological overlap measurement (TOM) and the dynamic hybrid cutting method were used to identify co-expressed gene modules (26,27). Finally, removing the grey module, we got 9 modules, and the minimum number of lncRNAs in each clustering was set 30. According to the heatmap of module-trait relationships, we found the red module was significantly correlated with the prognosis. The 123 lncRNAs in the red module were used to construct a risk score model.

Construction of risk score model

The 427 samples were randomly divided into a training cohort (n=215) and a validation cohort (n=212) (Table S1). In order to verify the importance of lncRNAs in the red module, 123 lncRNAs in the training cohort were screened by univariate Cox proportional hazards regression, and 9 lncRNAs with $P < 0.05$ were obtained. Then, lncRNAs with high correlation were removed by lasso regression. Finally, multivariate Cox proportional hazards regression was performed to determine the risk score model, the coefficients and hazard ratio (HR) values were obtained by

Akaike information criterion (AIC). The calculation formula is as follows: risk score = $\Sigma(C \times \text{EXP}_{\text{lncRNA}})$. In our formula, $\text{EXP}_{\text{lncRNA}}$ represents the expression of six lncRNAs, and C represents the corresponding coefficient of multivariate Cox proportional hazards regression.

Validation and analysis of risk score model

Based on the median of risk score, the samples were divided into high-risk and low-risk groups. The receiver operating characteristic (ROC) curve was used to judge the prediction accuracy of risk score model. Kaplan-Meier curves were used to calculate OS and the statistical differences were determined by a log-rank test (28-30). Univariate and multivariate independent prognostic analysis were performed to determine whether the risk score could be distinguished from other clinical variables. Statistically, $P < 0.05$ was set as significant differences.

Functional enrichment analysis

The entire LUAD samples were divided into two groups according to the median risk score. PCGs were screened by the cutoff criterion $\log_{2}\text{FC} > 1.5$, $P < 0.01$. Finally, 689 differential genes were obtained. Functional enrichment analysis of Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway of 689 differential genes were performed using the DAVID Bioinformatics Tool (<https://david.ncifcrf.gov/>, version 6.8) (31). Use the ClueGO package (<http://apps.cytoscape.org/apps/cluego>) in Cytoscape software to draw a network diagram of the biological process (BP). $P < 0.01$ was used as the cutoff criterion for functional annotation of GO terms and KEGG pathways.

Gene set enrichment analysis (GSEA) was performed to explore potential pathways between high-risk and low-risk groups (32). Set the false discovery rate (FDR) < 0.05 as the cutoff criterion.

Statistical analysis

All statistical tests were two-sided, and $P < 0.05$ was considered statistically significant. Statistical analyses were conducted

using R software (version 3.6.1, www.r-project.org).

Results

Data pre-processing of lncRNA profiles in LUAD

A total of 479 LUAD samples containing clinical information were downloaded from the TCGA database. According to the survival time, 31 patients were excluded, and 448 patients were remained. And the top 25% of lncRNAs ($n=3,532$) with the largest variance differences were screened for WGCNA analysis.

Identification of module related to the survival status of LUAD

To determine the expression characteristics of lncRNAs in LUAD, we constructed a co-expression network using WGCNA. After removing outliers, an adjacency matrix was constructed using 427 samples (Figure 2A). We choose $\beta = 5$ as the soft power threshold to ensure that the correlation coefficient was close to 0.9 (Figure S1A,B). Then, 9 different-color co-expression modules were determined (Figure 2B). Finally, we found that this red module was significantly correlated with the survival status ($\text{cor} = 0.77$, $P < 0.01$) by analyzing the relationship between the modules and the clinical variables of LUAD (Figure 2C,D).

Identification of Cox proportional hazards regression model

The above 427 LUAD samples were randomly divided into two groups: training cohort ($n=215$) and validation cohort ($n=212$) (Table S1). In the training cohort, we performed univariate Cox and lasso regression analysis to obtain 9 lncRNAs based on 123 lncRNAs in the red module (Figure 3A). Finally, we constructed a risk score model through six lncRNAs by performing multivariate Cox analysis (Figure 3B, Table 1). Based on six lncRNAs, the risk score is calculated using the following formula: risk score = $(-0.3057 \times \text{EXP}_{\text{VIM-AS1}}) + (0.9678 \times \text{EXP}_{\text{AC092811.1}}) + (1.0829 \times \text{EXP}_{\text{NF1A-AS1}}) + (-0.3505 \times \text{EXP}_{\text{AL035701.1}}) + (3.9378 \times \text{EXP}_{\text{AC079336.4}}) + (-0.2810 \times \text{EXP}_{\text{AL121790.2}})$. At the same time, based on the overall data of TCGA-LUAD,

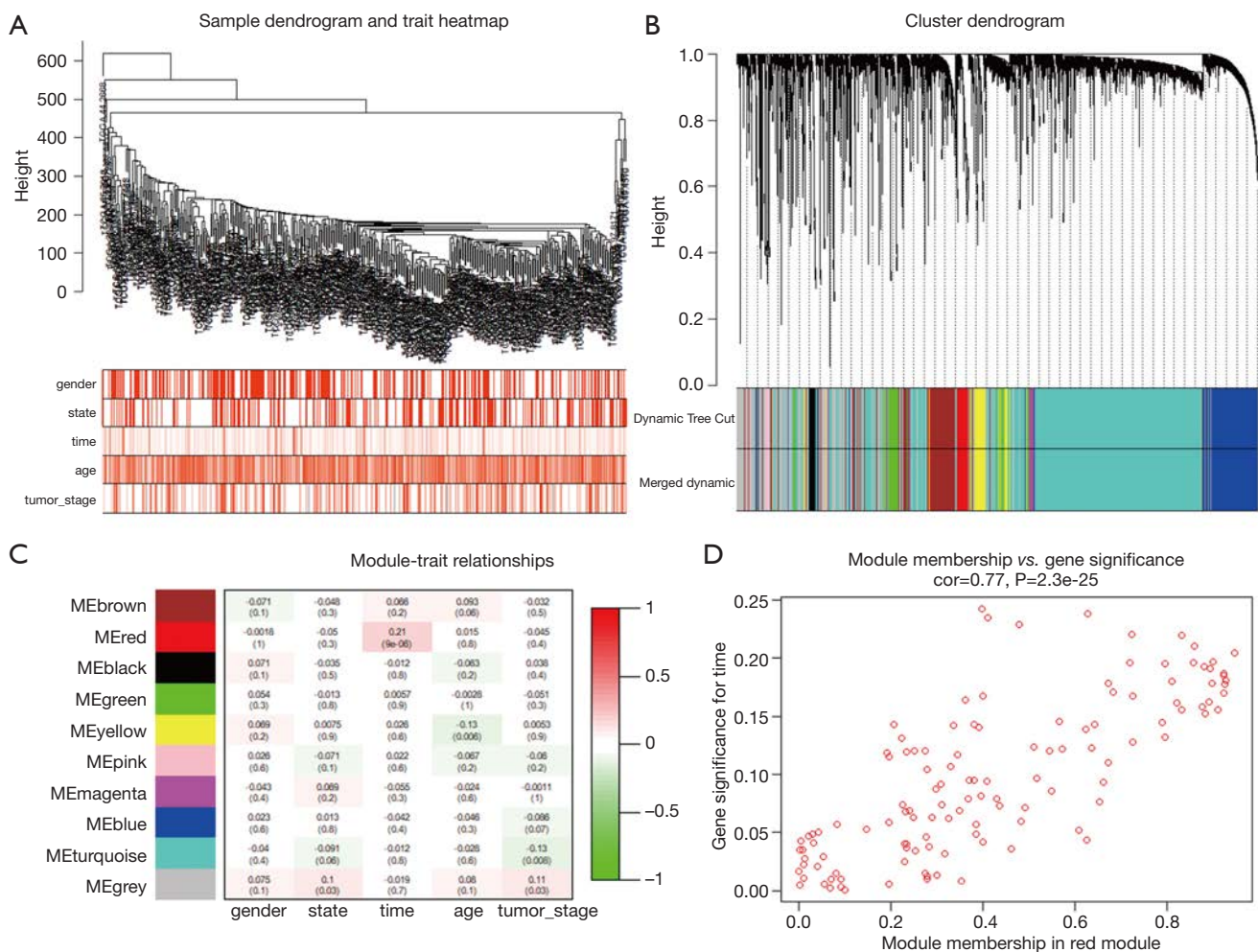


Figure 2 Construction of a weighted co-expression gene network and its relationship with clinical traits. (A) TCGA-LUAD cluster tree diagram and its clinical traits; (B) cluster dendrogram of top 25% lncRNAs based on dissimilarity measure (1-TOM); (C) heatmap of the correlation between module traits lncRNAs and clinical information of LUAD; (D) scatter plot of lncRNAs in red module. TCGA, the Cancer Genome Atlas; LUAD, lung adenocarcinoma; lncRNAs, long non-coding RNAs; TOM, topological overlap measurement.

we plotted the Kaplan-Meier curve of six lncRNAs (Figure 3C,D,E,F,G,H).

Prognostic efficiency and validation of six-lncRNA model

Based on the median risk score, we divided the training cohort into high-risk and low-risk groups. It could be seen from the scatter plot of Figure 4A, the mortality of high-risk patients was significantly higher than low-risk patients. From the heatmap of Figure 4A, as the risk score increases, three lncRNAs expression (VIM-AS1, AL035701.1 and AL121790.2) were gradually decreased,

and three lncRNAs expression (AC092811.1, NFIA-AS1 and AC079336.4) were gradually increased. Using the same classification method, we divided the validation cohort into high-risk and low-risk groups. We also found a significant increase of mortality in high-risk patients (Figure 4B). In addition, our results indicated that the 5-year survival areas under the ROC curve were 0.715, 0.735 and 0.721 in the training cohort, validation cohort and the entire TCGA data, which imply that this six-lncRNA model has good predictive value (Figure 4C). Finally, we found that the 5-year survival rates were significantly worse in the high-risk group compared with

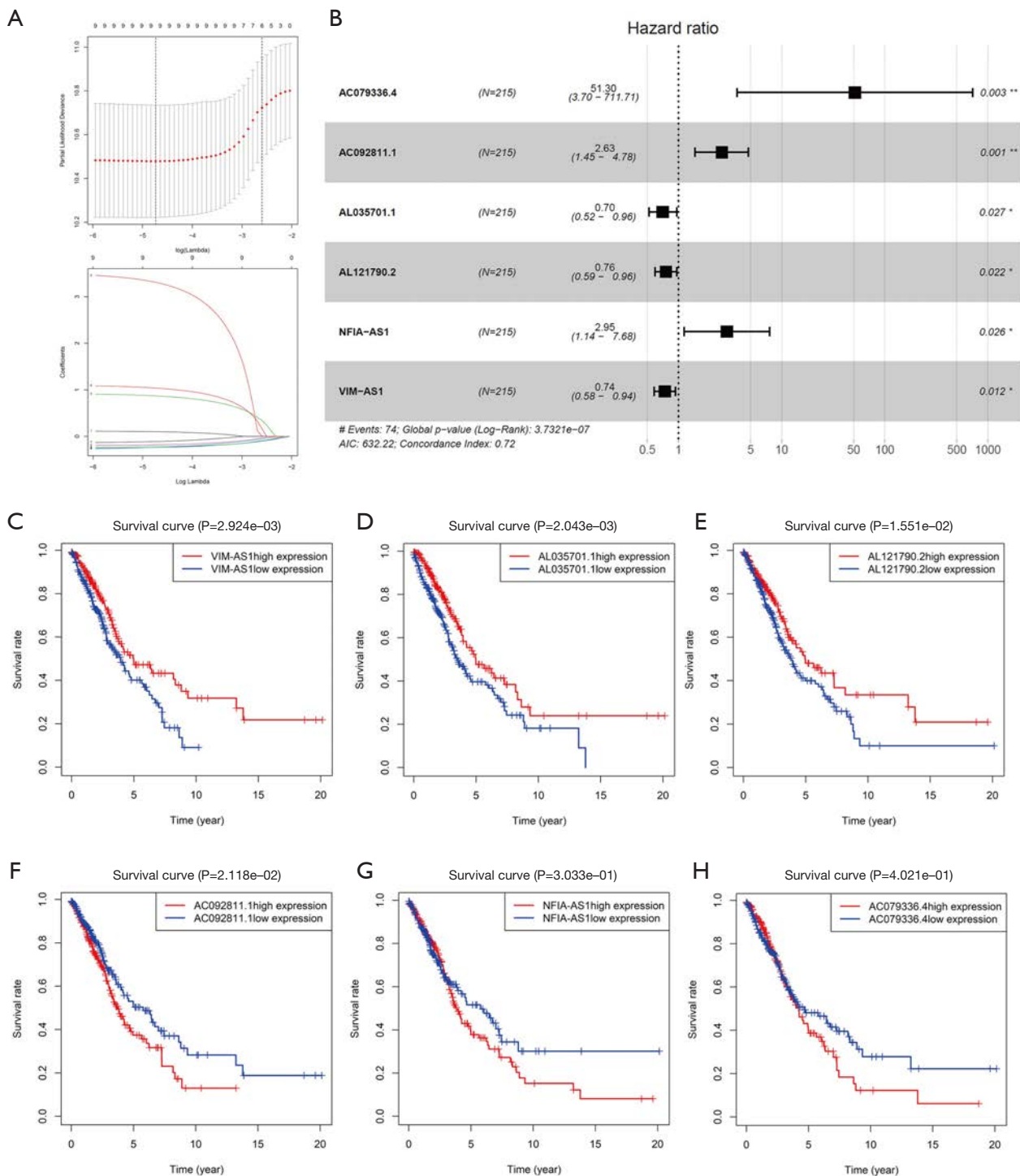


Figure 3 Identification of Cox proportional hazards regression model. (A) Lasso regression removes highly correlated genes; (B) six lncRNAs were significantly related with survival time to construct risk score model by multivariate Cox analysis; (C,D,E,F,G,H) based on the overall data of TCGA-LUAD, the individual survival curves of six lncRNAs are plotted. lncRNAs, long non-coding RNAs; TCGA, the Cancer Genome Atlas; LUAD, lung adenocarcinoma. *, $P < 0.05$; **, $P < 0.01$.

Table 1 6 lncRNAs significantly correlated with overall survival in the training cohort

Symbol	Univariate analysis			Multivariate analysis			
	HR	95% CI of HR	P value	HR	95% CI of HR	P value	Coefficient
VIM-AS1	0.714	0.573–0.889	0.003	0.737	0.580–0.936	0.012	–0.30569451
AC092811.1	2.559	1.388–4.719	0.003	2.632	1.450–4.777	0.001	0.96776812
NFIA-AS1	3.488	1.468–8.287	0.005	2.953	1.136–7.680	0.026	1.08294756
AL035701.1	0.663	0.480–0.916	0.013	0.704	0.516–0.962	0.027	–0.35051832
AC079336.4	28.828	1.397–594.845	0.030	51.304	3.698–711.708	0.003	3.93777712
AL121790.2	0.778	0.612–0.988	0.039	0.755	0.594–0.960	0.022	–0.28098658

lncRNAs, long non-coding RNAs; HR, hazard ratio; CI, confidence interval.

the low-risk group (all $P < 0.001$) (Figure 4D).

Independent prognostic ability and prognostic value of six-lncRNA model

Based on the training cohort, univariate and multivariate independent prognostic analysis were used to analyze the correlation of gender, age, TNM stage, T stage, N stage, M stage, risk score and prognosis. Through univariate independent prognostic analysis, we found that risk score was a powerful variable related to prognosis ($P < 0.001$) (Figure 5A, Table 2). By adding other clinical variables for multivariate independent prognostic analysis, we found that risk score could also be used as an independent prognostic variable ($P < 0.001$) (Figure 5B, Table 2).

Throughout the TCGA database, we verified the independent prognostic capabilities of risk score. According to TNM staging, patients were divided into early stage (stage I & II) and advanced stage (stage III & IV) for analysis. We found that the risk score can successfully predict the survival outcome in two subgroups (all $P < 0.01$) (Figure 5C). According to the T stage, the patients were divided into the highly differentiated group (T1 & T2) and the poorly differentiated group (T3 & T4) for analysis. We found that both groups were significantly different ($P < 0.01$, $P = 0.01$) (Figure 5D). According to the analysis of lymph node metastasis (N0, N1 & N2 & N3) and distant metastasis (M0, M1), we found that in the N0, N1 & N2 & N3 or M0 stage, the risk score could successfully predict the survival outcome (all $P < 0.01$) (Figure 5E,F). However, we have not found significant difference in the M1 stage, which may be related with the few samples. But we found that the high-risk group in the M1 stage had a lower survival rate than the low-risk group (Figure 5F). Similarly, subgroup

analysis in age (< 66 , ≥ 66) and gender (male, female) showed that the risk score could predict the survival outcome (Figure S2A,B). These results indicated that the predictive ability of six-lncRNA model was not affected by gender, age, TNM stage, T stage, N stage, and M stage. In addition, based on the training cohort, we found that the 3-year and 5-year areas of six-lncRNA model under the ROC curve were higher (AUC = 0.798, AUC = 0.771) than the other clinical variables (Figure 5G,H).

Functional enrichment analysis

The biological function of lncRNAs is still unknown. Therefore, in order to accurately evaluate the biological function of this six-lncRNA model, we analyzed the function of 689 differential genes according to the high-risk and low-risk groups of risk score (Figure S3A). As shown in Figure 6A, BPs were mainly involved in the cell cycle process and DNA metabolic process. The cellular components (CCs) were mainly enriched in nuclear chromosome part and chromatin (Figure 6B). Enriched molecular functions (MFs) were mainly enriched in chromatin binding and cell adhesion molecule binding (Figure 6C). KEGG functional analysis found that the main enrichment was spliceosome, cell cycle and DNA replication (Figure 6D).

The functional GSEA showed that the high-risk group were highly enriched in proteasome and protein export (Figure S3B), and the low-risk group were highly enriched in cell adhesion molecules and T/B cell receptor signaling pathway (Figure 6E).

Discussion

In the past, most researchers concentrated on studying

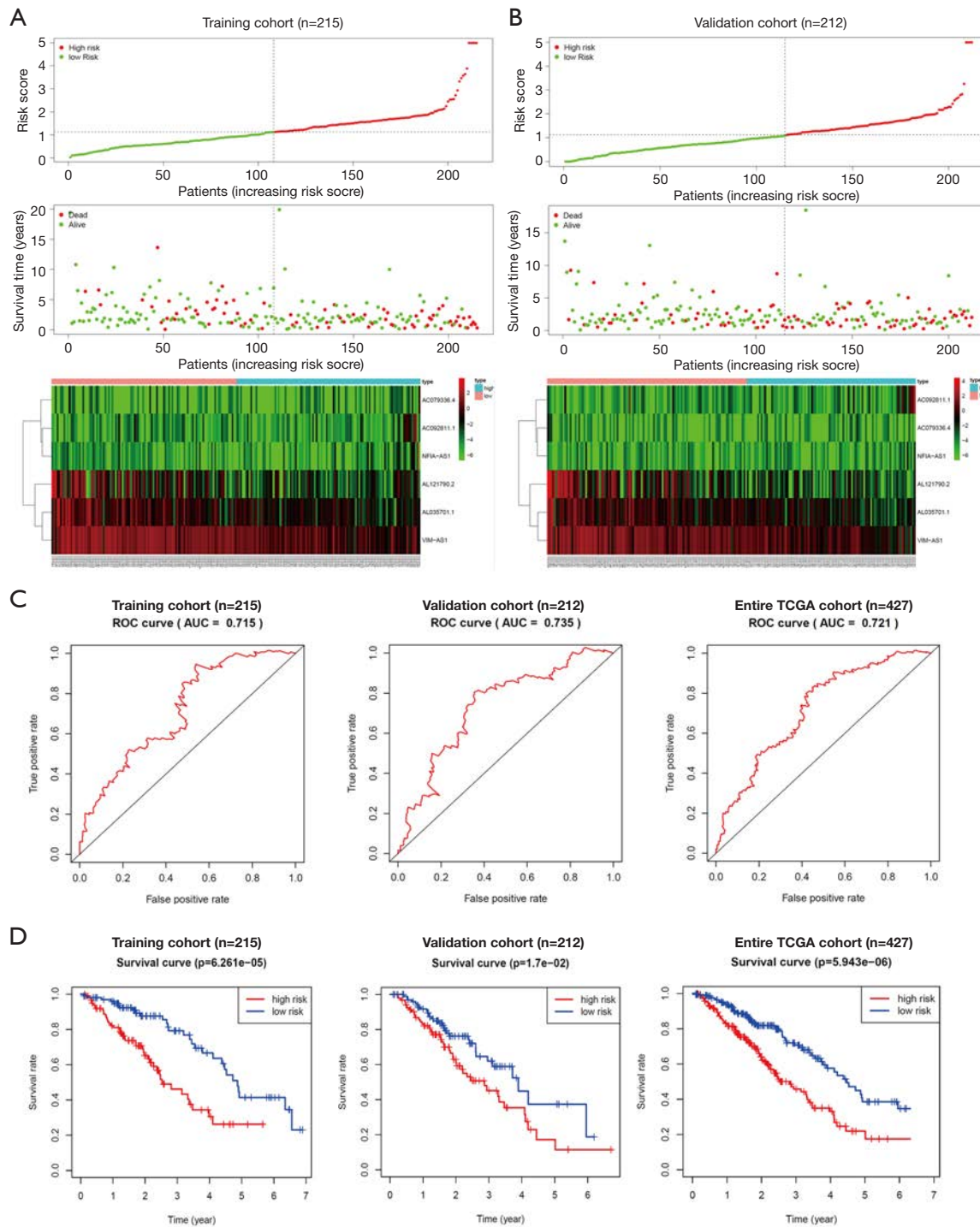


Figure 4 Prognostic efficiency and validation of six-lncRNA model. The risk score distribution, the vital status of patients and the heatmap based on the six-lncRNA model in the training cohort (A) and the validation cohort (B). (C) The 5-year area under the ROC curve has a good predictive value in the training cohort, validation cohort and the entire TCGA cohort. (D) Kaplan-Meier curve analysis of survival in high-risk and low-risk groups. Compared with the low-risk group, the prognosis of high-risk group was poor in the training cohort, validation cohort and the entire TCGA cohort ($P < 0.01$). lncRNAs, long non-coding RNAs; ROC, receiver operating characteristic; TCGA, the Cancer Genome Atlas.

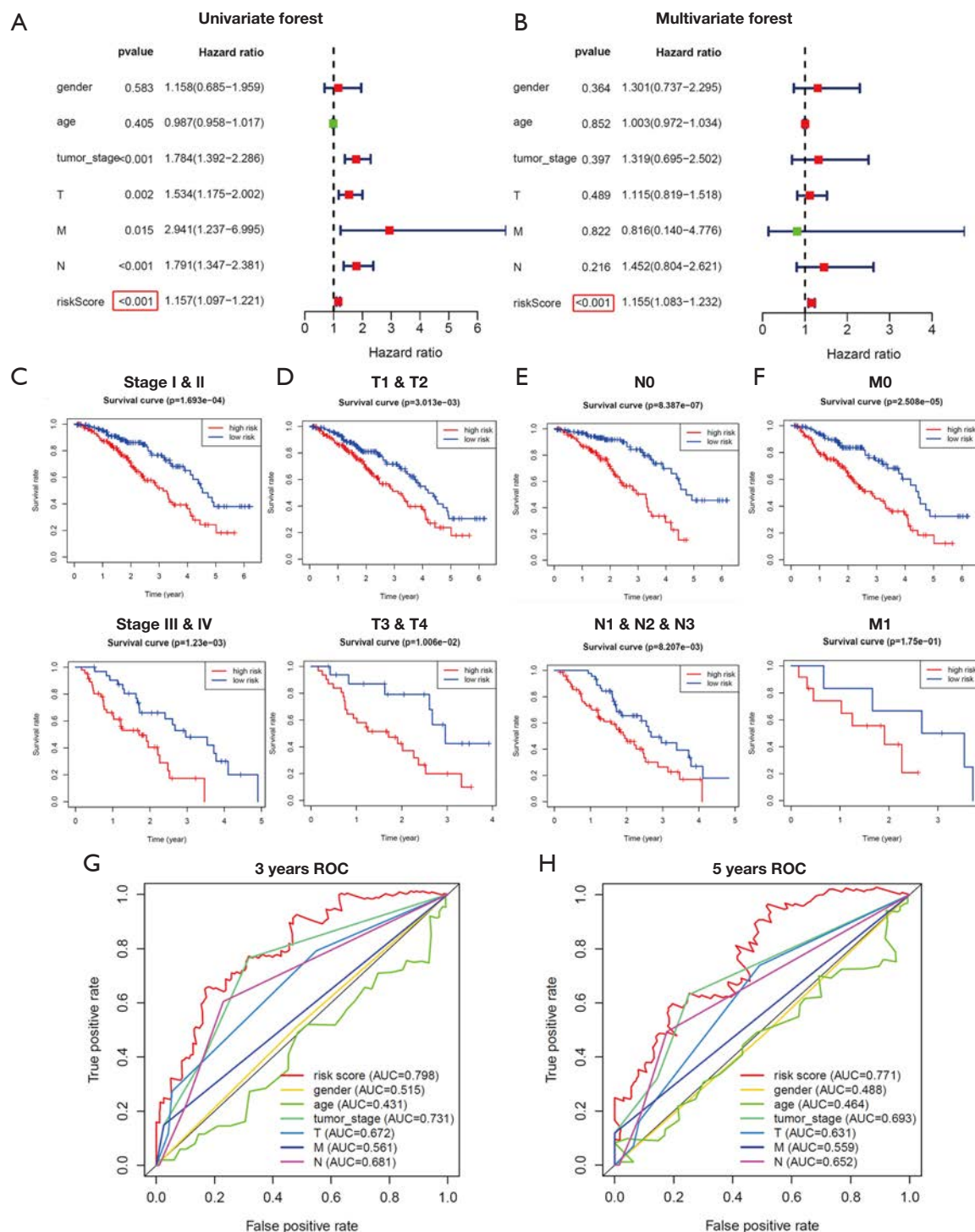


Figure 5 Analysis of independent prognostic ability and prognostic value of six-lncRNA model. Univariate (A) and multivariate (B) independent prognostic analysis based on the gender, age, TNM stage, T stage, N stage, M stage and risk score in the training cohort. Based on the TNM stage (C), T stage (D), N stage (E) and M stage(F), we verified the independent prognostic capabilities of risk score in the entire TCGA cohort. Compared with the individual clinical variables, the 3-year (G) and 5-year (H) areas of risk score under the ROC curve were analyzed.

Table 2 The prognostic effect of different clinical characteristics

Variables (n=215)	Univariate analysis			Multivariate analysis		
	HR	95% CI of HR	P value	HR	95% CI of HR	P value
Gender	1.158	0.685–1.959	0.583	1.301	0.737–2.295	0.364
Age	0.987	0.958–1.017	0.405	1.003	0.972–1.034	0.852
Tumor stage	1.784	1.392–2.286	4.70E–06	1.319	0.695–2.502	0.397
T	1.534	1.175–2.002	0.002	1.115	0.819–1.518	0.489
M	2.941	1.237–6.995	0.01466	0.816	0.140–4.776	0.822
N	1.791	1.347–2.381	6.05E–05	1.452	0.804–2.621	0.216
Risk score	1.157	1.097–1.221	9.53E–08	1.155	1.083–1.232	1.14E–05

HR, hazard ratio; CI, confidence interval.

the role of PCGs in cellular behavior regulation (33–35). However, some cellular behaviors alterations have been found not mediated by PCGs, either linked to “gene deserts” regions. It is reported that only 2% of the human genome encodes proteins, which further supports the potential regulation role of “gene deserts” regions in cellular behaviors (36,37). As products of “gene deserts” regions, lncRNA dysfunctions have been presented in various tumors and are closely related with progressing (38,39). But there is still rare work try to analysis the correlation between lncRNAs and tumor prognosis systematically (24). In this work, we employed the LUAD database from TCGA to construct a six-lncRNA prognostic model and investigated its prognostic evaluation efficiency by the following steps: a prognosis module was screened through WGCNA, and then a six-lncRNA model was identified by constructing multivariate Cox proportional hazards regression and verified by validation cohort. Finally, we have conducted in-depth studies on the biological functions of this six-lncRNA model.

This six-lncRNA model is consisted by VIM-AS1, AL035701.1, AL121790.2, AC092811.1, NFIA-AS1 and AC079336.4. Our study found that among these lncRNAs, the low expression of 3 lncRNAs (VIM-AS1, AL035701.1 and AL121790.2) and the high expression of 3 lncRNAs (AC092811.1, NFIA-AS1 and AC079336.4) are related with the poor prognosis in LUAD (Figure 3C,D,E,F,G,H). It is reported that VIM-AS1 could promote the progression and metastasis of colorectal cancer by inducing EMT (40). In addition, the abnormal expression of VIM-AS1 in cumulus cells during embryonic development is crucial for oocyte growth (41). Although the other

five lncRNAs have not been reported in LUAD-related research, more works are needed to verify this finding in the future. In addition, based on the TNM stage, T stage, N stage and M stage, we have verified that the independent prognostic capabilities of six-lncRNA model is statistically significant (all $P < 0.001$, Figure 5C,D,E,F). And we also found that the 5-year areas of the six-lncRNA model (AUC =0.771) under the ROC curve was higher than other clinical variables, such as gender (AUC =0.488), age (AUC =0.464), tumor stage (AUC =0.693), T stage (AUC =0.631), N stage (AUC =0.652) and M stage (AUC =0.559) (Figure 5H). These results indicate that the six-lncRNA model shows a higher-risk detection efficiency as contrasted to other clinical variables.

To determine the biological function of the six-lncRNA model, we analyzed the function of 689 differential genes obtained by high-risk and low-risk groups. We found that these differential genes were almost related with cell cycle process, cell adhesion molecule binding and the cell cycle and DNA replication pathway. Among them, cell adhesion regulation is a key factor for tumor invasion and occurrence (42), cell cycle processing and DNA replication also been shown related to tumor development and occurrence (29,30). GSEA analysis found that the low-risk group was mainly enriched in cell adhesion molecules and T/B cell receptor signaling pathway, which imply that the immune system is involve in suppressing the malignant processes of tumor.

Conclusions

Epidemiological statistics show that the incidence of various types of lung cancer has changed significantly compared

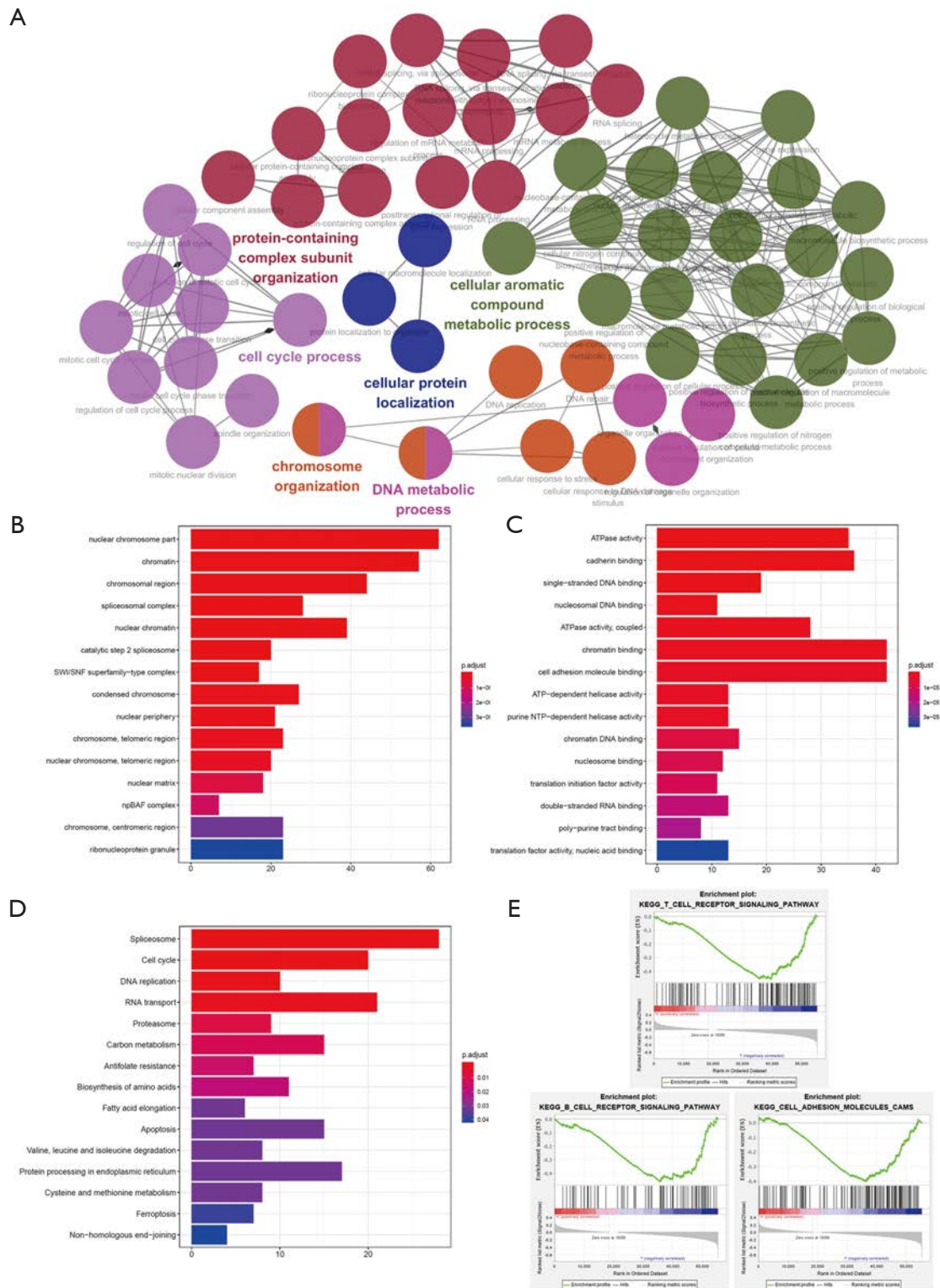


Figure 6 Functional enrichment analysis. (A) Using the ClueGO package to draw the network diagram of the biological process based on the prognostic differential genes. Histogram of (B) cellular component, (C) molecular function and (D) enrichment of KEGG pathway analysis of prognostic differential genes. (E) Based on the low-risk group in the entire TCGA cohort, gene sets were analyzed by GSEA.

with the past. LUAD has replaced LUSC as the main type of lung cancers (7). It shown that even LUAD patients exhibit same clinical stage and nearly pathological subtype, they may bear different prognosis and recurrence risk after surgery. In this study, we constructed a Cox proportional hazards regression model consisting of six-lncRNA, which could help physicians to précising subtype LUAD patients and give them more individual treatment.

Acknowledgments

The authors are grateful to all study participants.

Funding: The research was supported by the National Natural Science Foundation of China (NSFC) (No. 81402240 to Senyi Deng), the Science and Technology Plan Project (grant number 2019YFS0335), Sichuan Province, China and 1.3.5 Project for Disciplines of Excellence, West China Hospital, Sichuan University (ZYJC18009 to Jiandong Mei).

Footnote

Reporting Checklist: The authors have completed the TRIPOD reporting checklist. Available at <http://dx.doi.org/10.21037/tcr-20-2436>

Peer Review File: Available at <http://dx.doi.org/10.21037/tcr-20-2436>

Conflicts of Interest: Both authors have completed the ICMJE uniform disclosure form (available at <http://dx.doi.org/10.21037/tcr-20-2436>). Both authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the

formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

1. Yoda S, Dagogo-Jack I, Hata AN. Targeting oncogenic drivers in lung cancer: Recent progress, current challenges and future opportunities. *Pharmacol Ther* 2019;193:20-30.
2. Petrella F, Spaggiari L. Therapeutic options following pneumonectomy in non-small cell lung cancer. *Expert Rev Respir Med* 2016;10:919-25.
3. Reungwetwattana T, Dy GK. Targeted therapies in development for non-small cell lung cancer. *J Carcinog* 2013;12:22.
4. Sa H, Ma K, Gao Y, et al. Predictive Value of Tumor Mutation Burden in Immunotherapy for Lung Cancer. *Zhongguo Fei Ai Za Zhi* 2019;22:380-4.
5. Inamura K. Clinicopathological Characteristics and Mutations Driving Development of Early Lung Adenocarcinoma: Tumor Initiation and Progression. *Int J Mol Sci* 2018;19:1259.
6. Matěj R, Rohan Z, Němejcová K, et al. Molecular pathology of lung cancer in routine diagnostic practice: 2017 update. *Cesk Patol* 2017;53:159-66.
7. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. *CA Cancer J Clin* 2020;70:7-30.
8. Peng W, Wang J, Shan B, et al. Diagnostic and Prognostic Potential of Circulating Long Non-Coding RNAs in Non Small Cell Lung Cancer. *Cell Physiol Biochem* 2018;49:816-27.
9. Derrien T, Johnson R, Bussotti G, et al. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res* 2012;22:1775-89.
10. Djebali S, Davis CA, Merkel A, et al. Landscape of transcription in human cells. *Nature* 2012;489:101-8.
11. Batista PJ, Chang HY. Long noncoding RNAs: cellular address codes in development and disease. *Cell* 2013;152:1298-307.
12. Chen JH, Zhou LY, Xu S, et al. Overexpression of lncRNA HOXA11-AS promotes cell epithelial-mesenchymal transition by repressing miR-200b in non-small cell lung cancer. *Cancer Cell Int* 2017;17:64.
13. Cabili MN, Trapnell C, Goff L, et al. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev* 2011;25:1915-27.

14. Su M, Xiao Y, Tang J, et al. Role of lncRNA and EZH2 Interaction/Regulatory Network in Lung Cancer. *J Cancer* 2018;9:4156-65.
15. Yao G, Chen K, Qin Y, et al. Long Non-coding RNA JHDM1D-AS1 Interacts with DHX15 Protein to Enhance Non-Small-Cell Lung Cancer Growth and Metastasis. *Mol Ther Nucleic Acids* 2019;18:831-40.
16. Wu W, Zhao Y, Gao E, et al. LncRNA DLEU2 accelerates the tumorigenesis and invasion of non-small cell lung cancer by sponging miR-30a-5p. *J Cell Mol Med* 2020;24:441-50.
17. Huang N, Guo W, Ren K, et al. LncRNA AFAP1-AS1 Suppresses miR-139-5p and Promotes Cell Proliferation and Chemotherapy Resistance of Non-small Cell Lung Cancer by Competitively Upregulating RRM2. *Front Oncol* 2019;9:1103.
18. Zhao W, Zhang LN, Wang XL, et al. Long noncoding RNA NSCLCAT1 increases non-small cell lung cancer cell invasion and migration through the Hippo signaling pathway by interacting with CDH1. *FASEB J* 2019;33:1151-66.
19. He X, Wang J, Chen J, et al. lncRNA UCA1 Predicts a Poor Prognosis and Regulates Cell Proliferation and Migration by Repressing p21 and SPRY1 Expression in GC. *Mol Ther Nucleic Acids* 2019;18:605-16.
20. Zhang N, Zeng X, Sun C, et al. LncRNA LINC00963 Promotes Tumorigenesis and Radioresistance in Breast Cancer by Sponging miR-324-3p and Inducing ACK1 Expression. *Mol Ther Nucleic Acids* 2019;18:871-81.
21. Feng J, Yang G, Liu Y, et al. LncRNA PCNAP1 modulates hepatitis B virus replication and enhances tumor growth of liver cancer. *Theranostics* 2019;9:5227-45.
22. Shang A, Wang W, Gu C, et al. Long non-coding RNA HOTTIP enhances IL-6 expression to potentiate immune escape of ovarian cancer cells by upregulating the expression of PD-L1 in neutrophils. *J Exp Clin Cancer Res* 2019;38:411.
23. Yang J, Qiu Q, Qian X, et al. Long noncoding RNA LCAT1 functions as a ceRNA to regulate RAC1 function by sponging miR-4715-5p in lung cancer. *Mol Cancer* 2019;18:171.
24. Lin T, Fu Y, Zhang X, et al. A seven-long noncoding RNA signature predicts overall survival for patients with early stage non-small cell lung cancer. *Aging (Albany NY)* 2018;10:2356-66.
25. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 2008;9:559.
26. Li A, Horvath S. Network module detection: Affinity search technique with the multi-node topological overlap measure. *BMC Res Notes* 2009;2:142.
27. Langfelder P, Zhang B, Horvath S. Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. *Bioinformatics* 2008;24:719-20.
28. Ramos M, Schiffer L, Re A, et al. Software for the Integration of Multiomics Experiments in Bioconductor. *Cancer Res* 2017;77:e39-42.
29. Singhal S, Vachani A, Antin-Ozerkis D, et al. Prognostic implications of cell cycle, apoptosis, and angiogenesis biomarkers in non-small cell lung cancer: a review. *Clin Cancer Res* 2005;11:3974-86.
30. Ho ST, Lin CC, Tung YT, et al. Molecular Mechanisms Underlying Yatein-Induced Cell-Cycle Arrest and Microtubule Destabilization in Human Lung Adenocarcinoma Cells. *Cancers (Basel)* 2019;11:1384.
31. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 2009;4:44-57.
32. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 2005;102:15545-50.
33. Seidl C, Panzitt K, Bertsch A, et al. MicroRNA-182-5p regulates hedgehog signaling pathway and chemosensitivity of cisplatin-resistant lung adenocarcinoma cells via targeting GLI2. *Cancer Lett* 2020;469:266-76.
34. Bai Y, Xiong L, Zhu M, et al. Co-expression network analysis identified KIF2C in association with progression and prognosis in lung adenocarcinoma. *Cancer Biomark* 2019;24:371-82.
35. Zeng F, Wang Q, Wang S, et al. Linc00173 promotes chemoresistance and progression of small cell lung cancer by sponging miR-218 to regulate Etk expression. *Oncogene* 2020;39:293-307.
36. Beroukhi R, Mermel CH, Porter D, et al. The landscape of somatic copy-number alteration across human cancers. *Nature* 2010;463:899-905.
37. Zack TI, Schumacher SE, Carter SL, et al. Pan-cancer patterns of somatic copy number alteration. *Nat Genet* 2013;45:1134-40.
38. 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;16:98.
39. Xie Y, Zhang Y, Du L, et al. Circulating long noncoding RNA act as potential novel biomarkers for diagnosis

- and prognosis of non-small cell lung cancer. *Mol Oncol* 2018;12:648-58.
40. Rezanejad Bardaji H, Asadi MH, Yaghoobi MM. Long noncoding RNA VIM-AS1 promotes colorectal cancer progression and metastasis by inducing EMT. *Eur J Cell Biol* 2018;97:279-88.
 41. Bouckenheimer J, Fauque P, Lecellier CH, et al.

Cite this article as: Bai Y, Deng S. A six-long noncoding RNA model predicts prognosis in lung adenocarcinoma. *Transl Cancer Res* 2020;9(12):7505-7518. doi: /10.21037/tcr-20-2436

- Differential long non-coding RNA expression profiles in human oocytes and cumulus cells. *Sci Rep* 2018;8:2202.
42. Olajuyin AM, Olajuyin AK, Wang Z, et al. CD146 T cells in lung cancer: its function, detection, and clinical implications as a biomarker and therapeutic target. *Cancer Cell Int* 2019;19:247.

Supplementary

Table S1 Display of randomly grouped samples

Train	Validation
TCGA-05-4384	TCGA-05-4249
TCGA-05-4389	TCGA-05-4250
TCGA-05-4398	TCGA-05-4382
TCGA-05-4403	TCGA-05-4390
TCGA-05-4405	TCGA-05-4396
TCGA-05-4417	TCGA-05-4397
TCGA-05-4425	TCGA-05-4402
TCGA-05-4426	TCGA-05-4415
TCGA-05-4427	TCGA-05-4418
TCGA-05-4434	TCGA-05-4420
TCGA-05-5420	TCGA-05-4422
TCGA-05-5425	TCGA-05-4424
TCGA-05-5428	TCGA-05-4430
TCGA-05-5715	TCGA-05-4432
TCGA-35-4122	TCGA-05-4433
TCGA-35-4123	TCGA-05-5423
TCGA-38-4625	TCGA-05-5429
TCGA-38-4626	TCGA-35-5375
TCGA-38-4627	TCGA-38-4628
TCGA-38-4629	TCGA-38-4630
TCGA-38-6178	TCGA-38-4631
TCGA-44-2659	TCGA-38-4632
TCGA-44-2668	TCGA-38-7271
TCGA-44-3919	TCGA-44-2655
TCGA-44-6144	TCGA-44-2657
TCGA-44-6147	TCGA-44-2661
TCGA-44-6774	TCGA-44-2665
TCGA-44-6777	TCGA-44-3396
TCGA-44-6778	TCGA-44-3398
TCGA-44-6779	TCGA-44-6145
TCGA-44-7660	TCGA-44-6776
TCGA-44-7662	TCGA-44-7659
TCGA-44-8117	TCGA-44-7661
TCGA-44-8120	TCGA-44-7667
TCGA-44-A479	TCGA-44-7669

Table S1 (continued)

Table S1 (continued)

Train	Validation
TCGA-44-A47A	TCGA-44-7670
TCGA-44-A47G	TCGA-44-7671
TCGA-49-4486	TCGA-44-7672
TCGA-49-4505	TCGA-44-A47B
TCGA-49-4510	TCGA-44-A4SS
TCGA-49-4512	TCGA-44-A4SU
TCGA-49-4514	TCGA-49-4487
TCGA-49-6743	TCGA-49-4488
TCGA-49-6744	TCGA-49-4494
TCGA-49-6761	TCGA-49-4501
TCGA-49-6767	TCGA-49-4507
TCGA-49-AAQV	TCGA-49-6745
TCGA-49-AAR3	TCGA-49-AAR0
TCGA-49-AAR9	TCGA-49-AAR4
TCGA-49-AARE	TCGA-49-AARN
TCGA-49-AARO	TCGA-49-AARQ
TCGA-4B-A93V	TCGA-49-AARR
TCGA-50-5044	TCGA-50-5045
TCGA-50-5066	TCGA-50-5049
TCGA-50-5068	TCGA-50-5051
TCGA-50-5072	TCGA-50-5055
TCGA-50-5930	TCGA-50-5931
TCGA-50-5932	TCGA-50-5935
TCGA-50-5933	TCGA-50-5939
TCGA-50-5941	TCGA-50-5942
TCGA-50-5946	TCGA-50-6593
TCGA-50-6591	TCGA-50-6595
TCGA-50-6592	TCGA-50-7109
TCGA-50-6594	TCGA-50-8457
TCGA-50-6597	TCGA-53-7813
TCGA-50-8460	TCGA-55-1594
TCGA-53-7626	TCGA-55-5899
TCGA-55-1596	TCGA-55-6642
TCGA-55-6543	TCGA-55-6969
TCGA-55-6968	TCGA-55-6970

Table S1 (continued)

Table S1 (continued)

Train	Validation
TCGA-55-6972	TCGA-55-6971
TCGA-55-6979	TCGA-55-6986
TCGA-55-6980	TCGA-55-7227
TCGA-55-6981	TCGA-55-7283
TCGA-55-6985	TCGA-55-7284
TCGA-55-6987	TCGA-55-7570
TCGA-55-7281	TCGA-55-7576
TCGA-55-7573	TCGA-55-7724
TCGA-55-7574	TCGA-55-7725
TCGA-55-7726	TCGA-55-7815
TCGA-55-7727	TCGA-55-7910
TCGA-55-7816	TCGA-55-7913
TCGA-55-7903	TCGA-55-7914
TCGA-55-7907	TCGA-55-7994
TCGA-55-7911	TCGA-55-7995
TCGA-55-8087	TCGA-55-8085
TCGA-55-8089	TCGA-55-8090
TCGA-55-8091	TCGA-55-8092
TCGA-55-8096	TCGA-55-8205
TCGA-55-8097	TCGA-55-8206
TCGA-55-8203	TCGA-55-8299
TCGA-55-8204	TCGA-55-8301
TCGA-55-8208	TCGA-55-8505
TCGA-55-8302	TCGA-55-8507
TCGA-55-8510	TCGA-55-8511
TCGA-55-8614	TCGA-55-8512
TCGA-55-8615	TCGA-55-8514
TCGA-55-A48X	TCGA-55-8616
TCGA-55-A48Y	TCGA-55-8620
TCGA-55-A490	TCGA-55-8621
TCGA-55-A491	TCGA-55-A48Z
TCGA-55-A492	TCGA-62-8395
TCGA-55-A494	TCGA-62-8397
TCGA-55-A4DF	TCGA-62-8398
TCGA-55-A4DG	TCGA-62-8399

Table S1 (continued)

Table S1 (continued)

Train	Validation
TCGA-55-A57B	TCGA-62-A46P
TCGA-62-8394	TCGA-62-A46Y
TCGA-62-8402	TCGA-62-A472
TCGA-62-A46O	TCGA-64-1677
TCGA-62-A46R	TCGA-64-1680
TCGA-62-A46S	TCGA-64-5781
TCGA-62-A46V	TCGA-67-3771
TCGA-62-A471	TCGA-67-3774
TCGA-64-1676	TCGA-67-6217
TCGA-64-1679	TCGA-69-7760
TCGA-64-1681	TCGA-69-7764
TCGA-64-5778	TCGA-69-7973
TCGA-64-5815	TCGA-69-7978
TCGA-67-3770	TCGA-69-A59K
TCGA-67-3772	TCGA-71-6725
TCGA-67-3773	TCGA-71-8520
TCGA-67-4679	TCGA-73-4668
TCGA-67-6215	TCGA-73-4676
TCGA-67-6216	TCGA-73-7498
TCGA-69-7763	TCGA-73-7499
TCGA-69-7765	TCGA-75-7025
TCGA-69-7974	TCGA-75-7027
TCGA-69-7980	TCGA-78-7146
TCGA-69-8253	TCGA-78-7150
TCGA-69-8254	TCGA-78-7154
TCGA-69-8255	TCGA-78-7158
TCGA-73-4658	TCGA-78-7159
TCGA-73-4659	TCGA-78-7160
TCGA-73-4662	TCGA-78-7162
TCGA-73-4666	TCGA-78-7167
TCGA-73-4670	TCGA-78-7220
TCGA-73-4675	TCGA-78-7535
TCGA-73-A9RS	TCGA-78-7537
TCGA-75-5146	TCGA-78-7540
TCGA-75-5147	TCGA-78-7542

Table S1 (continued)

Table S1 (continued)

Train	Validation
TCGA-75-6206	TCGA-78-7633
TCGA-78-7143	TCGA-78-8648
TCGA-78-7145	TCGA-78-8662
TCGA-78-7147	TCGA-80-5611
TCGA-78-7148	TCGA-86-6851
TCGA-78-7149	TCGA-86-7711
TCGA-78-7152	TCGA-86-7713
TCGA-78-7153	TCGA-86-7953
TCGA-78-7156	TCGA-86-7954
TCGA-78-7161	TCGA-86-8056
TCGA-78-7163	TCGA-86-8075
TCGA-78-7166	TCGA-86-8279
TCGA-78-7539	TCGA-86-8358
TCGA-78-8640	TCGA-86-8671
TCGA-78-8655	TCGA-86-8673
TCGA-78-8660	TCGA-86-8674
TCGA-80-5608	TCGA-86-A4JF
TCGA-83-5908	TCGA-86-A4P8
TCGA-86-6562	TCGA-91-6828
TCGA-86-7701	TCGA-91-6831
TCGA-86-7714	TCGA-91-6835
TCGA-86-7955	TCGA-91-6848
TCGA-86-8055	TCGA-91-6849
TCGA-86-8073	TCGA-91-8496
TCGA-86-8076	TCGA-91-8497
TCGA-86-8280	TCGA-91-A4BD
TCGA-86-8359	TCGA-93-7348
TCGA-86-8585	TCGA-95-7039
TCGA-86-8668	TCGA-95-7562
TCGA-86-8669	TCGA-95-7567
TCGA-86-A456	TCGA-95-7944
TCGA-86-A4D0	TCGA-95-7947
TCGA-86-A4P7	TCGA-95-7948
TCGA-91-6829	TCGA-95-A4VK
TCGA-91-6830	TCGA-95-A4VN

Table S1 (continued)

Table S1 (continued)

Train	Validation
TCGA-91-6836	TCGA-97-7547
TCGA-91-6840	TCGA-97-7553
TCGA-91-7771	TCGA-97-7937
TCGA-91-8499	TCGA-97-8171
TCGA-91-A4BC	TCGA-97-8175
TCGA-93-7347	TCGA-97-8177
TCGA-93-8067	TCGA-97-8552
TCGA-93-A4JO	TCGA-97-A4LX
TCGA-93-A4JP	TCGA-97-A4M0
TCGA-95-7043	TCGA-97-A4M1
TCGA-95-8494	TCGA-97-A4M3
TCGA-97-7552	TCGA-97-A4M6
TCGA-97-7554	TCGA-99-8033
TCGA-97-8174	TCGA-99-AA5R
TCGA-97-A4M2	TCGA-J2-8192
TCGA-97-A4M5	TCGA-J2-A4AD
TCGA-97-A4M7	TCGA-J2-A4AG
TCGA-99-7458	TCGA-L4-A4E5
TCGA-99-8025	TCGA-L9-A443
TCGA-99-8028	TCGA-L9-A444
TCGA-99-8032	TCGA-L9-A8F4
TCGA-J2-8194	TCGA-MN-A4N1
TCGA-J2-A4AE	TCGA-MN-A4N4
TCGA-L4-A4E6	TCGA-MN-A4N5
TCGA-L9-A50W	TCGA-MP-A4SY
TCGA-L9-A5IP	TCGA-MP-A4T4
TCGA-L9-A743	TCGA-MP-A4T7
TCGA-L9-A7SV	TCGA-MP-A4T9
TCGA-MP-A4SV	TCGA-MP-A4TA
TCGA-MP-A4SW	TCGA-MP-A4TE
TCGA-MP-A4T6	TCGA-MP-A4TF
TCGA-MP-A4T8	TCGA-MP-A4TI
TCGA-MP-A4TC	TCGA-MP-A4TK
TCGA-MP-A4TD	TCGA-NJ-A4YG
TCGA-MP-A4TH	TCGA-NJ-A4YP

Table S1 (continued)

Table S1 (continued)

Train	Validation
TCGA-MP-A5C7	TCGA-NJ-A55R
TCGA-NJ-A4YF	TCGA-S2-AA1A
TCGA-NJ-A4YQ	
TCGA-NJ-A7XG	
TCGA-O1-A52J	

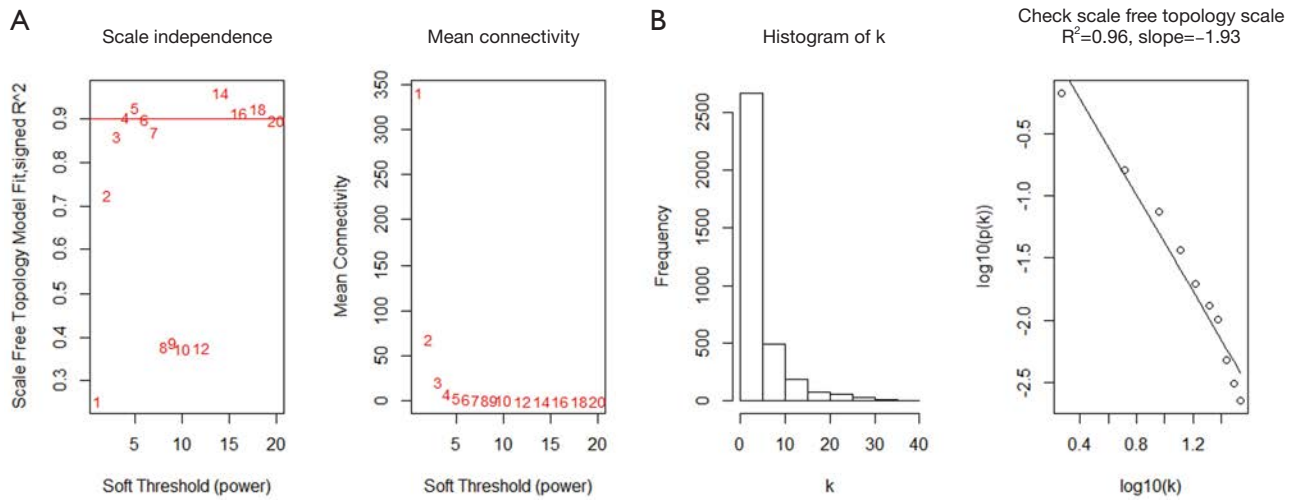


Figure S1 Determination of soft-threshold power in the weighted gene co-expression network analysis (WGCNA). (A) Analysis of the scale-free fit index and the mean connectivity for various soft-threshold powers. (B) Histogram of connectivity distribution and the scale free topology when $\beta = 5$.

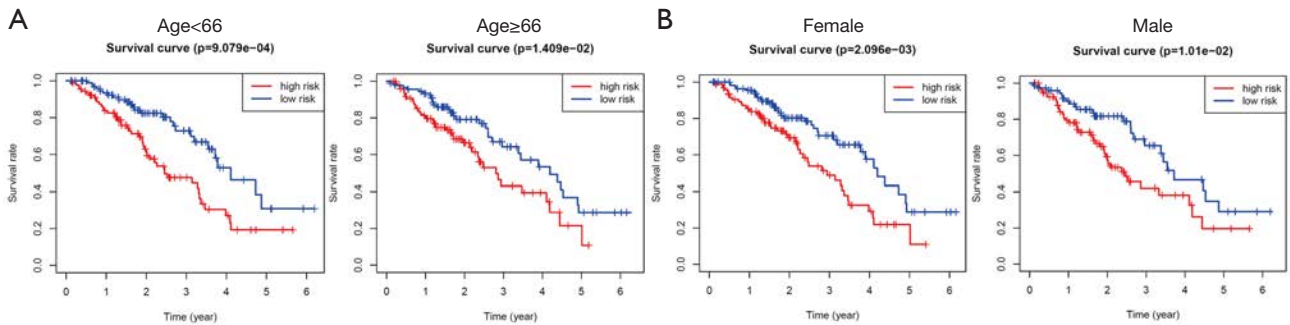


Figure S2 Analysis of independent prognostic ability of 6-lncRNA model. Based on the (A) age (<66, ≥66) and (B) gender (male, female), we verified the independent prognostic capabilities of 6-lncRNA model in the entire TCGA cohort.

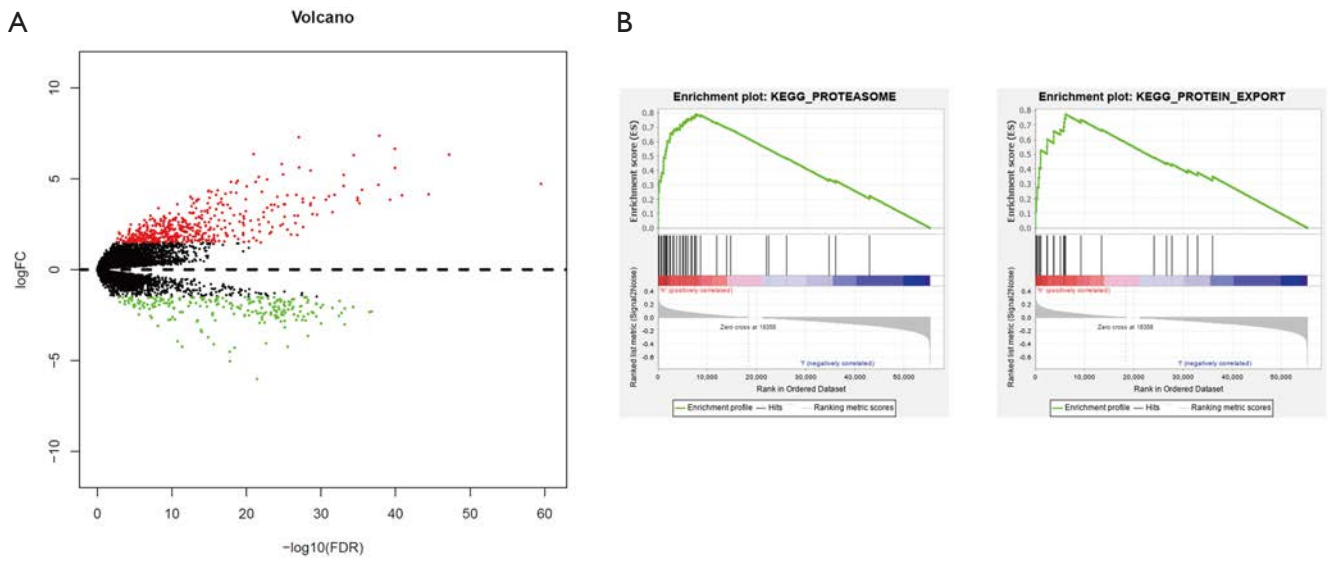


Figure S3 Functional enrichment analysis. (A) According to the median of risk score in the entire TCGA cohort, 689 differential genes were obtained. (B) Based on the high-risk group in the entire TCGA cohort, gene sets were analyzed by GSEA.