

Research Article

Zhihui Zhang, Liu Yang, Yujiang Li, Yunfei Wu, Xiang Li, Xu Wu*

Four long noncoding RNAs act as biomarkers in lung adenocarcinoma

<https://doi.org/10.1515/med-2021-0276>

received November 17, 2020; accepted March 21, 2021

Abstract

Introduction – Lung adenocarcinoma (LUAD) is currently one of the most common malignant tumors worldwide. However, there is a lack of long noncoding RNA (lncRNA)-based effective markers for predicting the prognosis of LUAD patients. We identified four lncRNAs that can effectively predict the prognosis of LUAD patients.

Methods – We used data gene expression profile for 446 patients from The Cancer Genome Atlas database. The patients were randomly divided into a training set and a test set. Significant lncRNAs were identified by univariate regression. Then, multivariate regression was used to identify lncRNAs significantly associated with the survival rate. We constructed four-lncRNA risk formulas for LUAD patients and divided patients into high-risk and low-risk groups. Identified lncRNAs subsequently verified in the test set, and the clinical independence of the lncRNA model was evaluated by stratified analysis. Then mutated genes were identified in the high-risk and low-risk groups. Enrichment analysis was used to determine the relationships between lncRNAs and co-expressed genes. Finally, the accuracy of the model was verified using external database.

Results – A four-lncRNA signature (AC018629.1, AC122134.1, AC119424.1, and AL138789.1) has been verified in the training and test sets to be significantly associated with the overall survival of LUAD patients.

Conclusions – The present study demonstrated that identified four-lncRNA signature can be used as an independent prognostic biomarker for the prediction of survival of LUAD patients.

Keywords: lung adenocarcinoma, lncRNA, prognosis, bioinformatics, biomarker

1 Introduction

Lung cancer is one of the most common cancers worldwide [1]. Adenocarcinoma accounts for 40% of all lung cancers [2]. Worsening of environmental pollution is associated with an increase in the number of patients with the disease. Many new methods are currently used to diagnose and treat this disease [3]. Epidermal growth factor receptor (EGFR) and anaplastic lymphoma kinase are two key oncogenes that trigger lung adenocarcinoma (LUAD) [4–6]. Despite great achievements, overall survival (OS) time for LUAD is less than 5 years because of aggressive properties of the tumor and a lack of effective early diagnostic and prognostic biomarkers [7]. Currently, there are only a few prognostic markers for LUAD, and thus identification of valuable prognostic indicators for clinical treatment is very important.

Long noncoding RNAs (lncRNAs) are RNAs that are longer than 200 bp [8]. lncRNAs have been recently shown to regulate large-scale gene expression programs; lncRNAs influence gene expression by interactions across multiple chromosomes and at various locations on a given chromosome. The mechanisms by which some lncRNAs target a distal binding site are not completely clear, and hence, spatial organization analysis of the genome may help to determine the function of lncRNAs. Recent studies suggest that lncRNAs may have extensive effects on the regulation of genes related to tumor formation and metastasis than previously thought [9–11]. Moreover, many lncRNAs can be used as predictors of prognosis and survival for tumor patients, providing a reliable basis for cancer diagnosis and treatment. LOXL1-AS1 can be used as a biomarker to assess the prognosis of glioma [12]. AK001796 is an independent predictor of poor prognosis

* **Corresponding author: Xu Wu**, Department of Thoracic and Cardiovascular Surgery/Huiqiao Medical Center, Nanfang Hospital, Southern Medical University, Jingxi Street, Guangzhou, Guangdong 510515, China, e-mail: wuxu_southhospital@163.com
Zhihui Zhang, Yujiang Li, Yunfei Wu: Department of Thoracic and Cardiovascular Surgery/Huiqiao Medical Center, Nanfang Hospital, Southern Medical University, Jingxi Street, Guangzhou, Guangdong 510515, China

Liu Yang: Department of Radiology, Charité – Universitätsmedizin Berlin, Berlin, Germany

Xiang Li: Department of Emergency Surgery, Nanfang Hospital, Southern Medical University, Guangzhou, Guangdong, China

in esophageal squamous cell carcinoma [13]. Prognostic lncRNA signatures have been detected in many cancer types [14–17]; however, the studies of lncRNAs with the prognosis of LUAD patients are in the early stages and this subject requires long-term efforts.

High-throughput sequencing data and basic clinical information used in the present study were provided by The Cancer Genome Atlas (TCGA). We used high-throughput data and clinical information of 466 LUAD patients in the TCGA database to identify potential lncRNA markers that can effectively predict the survival of LUAD patients. Eventually, we identified four lncRNAs, which were shown to be independent of clinical factors according to stratified analysis, and these four RNAs have not been reported previously. The findings of the present study demonstrated the important role of these four lncRNAs in predicting the survival of LUAD patients.

2 Materials and methods

2.1 Basic information of the patients

The lncRNA sequencing profile and LUAD clinical information were obtained from the TCGA data set. This study included 466 LUAD patients, excluding the data that lacked complete survival information. Associated clinical information included OS, age, sex, AJCC tumor stage, etc. The patients were randomly divided into two groups, including 233 patients as a training set and 233 patients as a test set. Detailed information on the sets used in the present study is provided in Table 1.

2.2 lncRNA sequencing profile source

LUAD RNA sequencing data were downloaded from the TCGA database (<https://tcga-data.nci.nih.gov/tcga/>). The data were compared with the human genome, and the reads per thousand bases per million exon models (RPKM) determine the expression levels of lncRNAs and mRNAs. We selected lncRNAs in the TCGA database according to the following three criteria: (1) the transcript was not found in any protein coding regions; (2) the transcribed sequence was found in the GENCODE project [18]; and (3) it was expressed in at least more than half of the LUAD patients. The transcripts with an average RPKM > 0.1 were considered a part of the lncRNA profile of 466 LUAD samples. Finally, 3,881 lncRNAs were identified in the data set. Then, the “edgeR” package for R [19] was used for differential expression analysis based on adjusted $P < 0.001$ and $|\log_2(F_c)| \geq 2$ threshold to determine differentially expressed RNAs.

Table 1: Clinical characteristics of LUAD patients

Characteristics	Training set (<i>n</i> = 233)	Test set (<i>n</i> = 233)	Total set (<i>n</i> = 466)
Vital status, <i>n</i> (%)			
Alive	169 (72.5)	178 (76.4)	347 (74.5)
Dead	64 (27.5)	55 (23.6)	119 (25.5)
Age (years), <i>n</i> (%)			
≤60	81 (34.7)	66 (28.3)	147 (31.5)
>60	152 (65.3)	167 (71.7)	319 (68.5)
Sex, <i>n</i> (%)			
Female	127 (54.5)	126 (54.1)	253 (54.2)
Male	106 (45.5)	107 (45.9)	213 (45.8)
Stage, <i>n</i> (%)			
I	124 (53.2)	130 (55.8)	254 (54.5)
II	57 (24.5)	53 (22.7)	110 (23.6)
III	37 (15.9)	40 (17.2)	77 (16.5)
IV	15 (6.4)	10 (4.3)	25 (5.4)
Mutation status, <i>n</i> (%)			
EGFR	31 (13.3)	28 (12.0)	59 (12.6)
ALK	9 (3.8)	13 (5.5)	21 (4.5)

Externally validated sequencing data, clinical information, and platform annotation information were obtained from the GSE11969 data set of the Gene Expression Omnibus (GEO) database (<http://www.ncbi.nlm.nih.gov/geo>). A total of 148 patients were included in this part of the study.

2.3 Establishment of the risk formula

Univariate Cox regression was used to calculate the associations between the expression level of each lncRNA and the survival of patients in the training set. lncRNAs with P values less than 0.001 were considered to be significant. Then, random forest and multivariate Cox regression models were used to analyze the lncRNAs that had the best independence from the prognosis of the patients. The regression coefficient obtained using multivariate Cox regression model was multiplied by the linear combination of the expression levels of corresponding lncRNAs define the prognostic risk formula. Risk formula = $\sum_{i=1}^N (Exp_i \times Coe_i)$, where N represents the total number of prognostic lncRNAs, Exp_i the expression of a certain lncRNA and Coe_i the regression coefficient obtained from the multivariate Cox regression analysis for a certain lncRNA numbered as i .

2.3.1 Statistical analysis

Kaplan–Meier survival analysis was used to determine the differences in the OS and progression-free survival (PFS) between the high-risk and low-risk groups. Multivariate

Cox regression and stratified analysis were used to determine whether the prognosis prediction based on identified lncRNA is independent of clinical variables. Receiver operating characteristic (ROC) curves were used to compare the sensitivity and specificity of survival prediction based on the risk score. The R package “maftools” was used to analyze mutated genes in LUAD. All analyses were performed using R software (version 3.6.3).

3 Results

3.1 Differentially expressed lncRNAs

A total of 381 differentially expressed lncRNAs were identified according to the criteria of $|\log_2(F_C)| \geq 2$ and $P_{\text{adj}} < 0.001$. These lncRNAs included 291 upregulated and 90 downregulated lncRNAs (Figure 1). Subsequent prognostic analysis used lncRNAs with expression levels significantly different between the LUAD and control groups.

3.2 Identification of four lncRNAs that predicted prognosis in the training set

A total of 466 LUAD patients were randomly divided into a training set ($n = 223$) and a test set ($n = 223$). Univariate

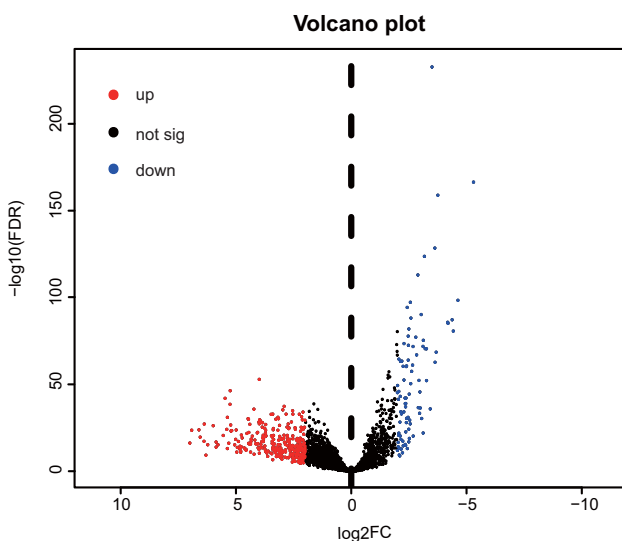


Figure 1: Identification of 381 differentially expressed lncRNAs. Red dots represent 291 upregulated lncRNAs and blue dots represent 90 downregulated lncRNAs, and the intermediate black dots represent lncRNAs without significant differences. lncRNA, long non-coding RNA; FC, fold change; sig, significance; up, upregulated; down, downregulated; FDR, false discovery rate.

Cox regression analysis based on lncRNAs in the training set was used to identify a total of 43 genes ($P < 0.05$) for subsequent analysis. Multivariate Cox regression analysis based on age, sex, and AJCC stage identified four lncRNAs as independent prognostic biomarkers for LUAD patients. These four lncRNAs included AC018629.1, AC122134.1, AC119424.1, and AL138789.1. Three of these lncRNAs (AC018629.1, AC122134.1, and AL138789.1) were positively associated with survival, indicating that higher expression levels of these lncRNA are associated with shorter survival of the patients. The negative coefficient estimated for the remaining lncRNA (AC119424.1) suggested that high expression level was associated with longer survival. Detailed information about these four lncRNAs is presented in Table 2.

3.3 Evaluation of the prognostic model in the training set

The expression levels of these four lncRNAs and their prognostic values obtained by multivariate Cox regression analysis were used to design a risk score formula to predict the survival of patients with LUAD. The risk score formula was as follows: risk score = $(0.22 \times \text{the expression level of AC018629.1}) + (0.11 \times \text{the expression level of AC122134.1}) + (-0.22 \times \text{the expression level of AC119424.1}) + (0.14 \times \text{the expression level of AL138789.1})$. The risk score was calculated for each patient in the training set based on the lncRNA signature, then the median risk score was used as a threshold to divide LUAD patients into high-risk ($n = 117$) and low-risk groups ($n = 116$). Additionally, Kaplan–Meier analysis was used to assess the effects of these prognostic indicators on the survival of patients in the training group. The Kaplan–Meier analysis showed that patients in the high-risk group had worse outcome compared with that in the low-risk group ($P = 6.91 \times 10^{-5}$; Figure 2a). The 3-year OS rate after diagnosis in the low-risk group was 87%, and the 5-year OS rate after diagnosis in the low-risk group was 77%; the 3-year and 5-year survival rates in the high-risk group were only 71 and 67%, respectively. Time-dependent ROC curves were used to evaluate the prognostic significance of four lncRNAs. The area under the curve (AUC) of the four-lncRNA signatures was 0.766, indicating good prognostic value for survival prediction of the patients (Figure 2b). Univariate and multivariable Cox regression analyses showed that risk score based on four lncRNAs was significantly associated with patient survival (Table 3). The basic information about the risk score, survival rate,

Table 2: Four lncRNAs significantly associated with the OS

Gene symbol	Ensemble ID	Coefficient	Hazard ratio	95% CI of HR	P value
AC018629.1	ENSG00000279806	0.22	1.25	1.10–1.42	5.97×10^{-4}
AC122134.1	ENSG00000261298	0.11	1.12	1.01–1.24	2.83×10^{-2}
AC119424.1	ENSG00000243384	-0.22	0.81	0.70–0.93	3.00×10^{-3}
AL138789.1	ENSG00000233589	0.14	1.16	1.01–1.33	3.97×10^{-2}

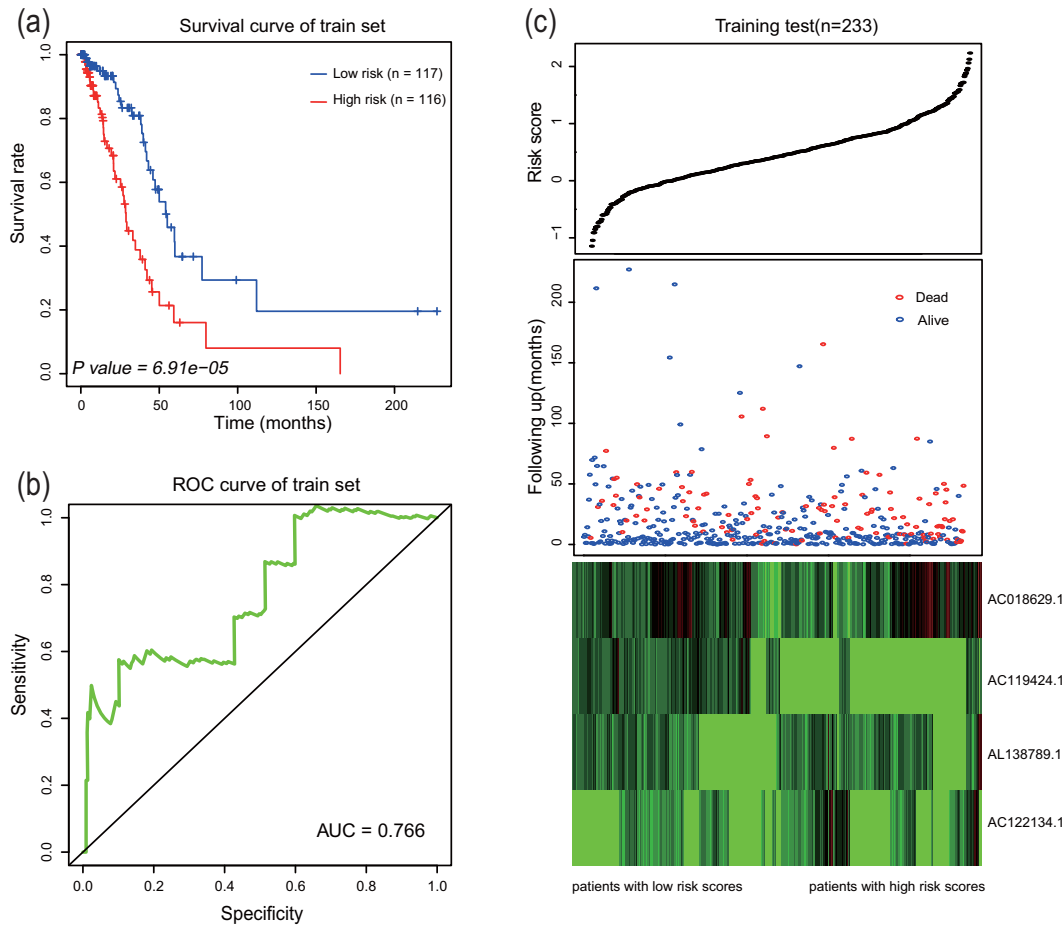


Figure 2: The four-lncRNA associated risk score model predicted the OS of LUAD patients in the training set. (a) Kaplan–Meier assessment of the survival of patients with LUAD in the high-risk and low-risk groups. *P* value represents the difference between the two curves according to the results of two-sided rank test. (b) The ROC curves were used to calculate the AUC and the sensitivity and specificity to evaluate the performance of the score. (c) Distribution of the four-lncRNA based risk score distribution, survival status of the patients, and heatmap expression profiles of four lncRNAs.

and expression profiles of four lncRNAs in the samples of the training data set are shown in Figure 2c. The distributions of the risk score, survival rates, and expression profiles of four lncRNAs in the samples of the training set were ranked according to the risk scores. Patients with a high risk

score had shorter survival compared with that of patients with a low risk score. In patients with high risk scores, the expression of three lncRNAs (AC018629.1, AC122134.1, and AL138789.1) was significantly upregulated, and the expression of one lncRNA (AC119424.1) was downregulated.

Table 3: Univariable and multivariable Cox regression analyses in each data set

Variables	Univariable model ^a			Multivariable model		
	HR	95% CI of HR	<i>P</i> value	HR	95% CI of HR	<i>P</i> value
Training set (<i>n</i> = 223)						
Four-lncRNA risk score	1.551	1.382–1.741	<0.001	1.614	1.410–1.846	<0.001
Gender	0.919	0.561–1.503	0.736	1.132	0.687–1.865	0.627
Age	1.007	0.982–1.032	0.56	1.046	1.019–1.075	<0.001
AJCC stage						
(II vs I)	2.593	1.163–4.166	<0.001	3.122	1.549–6.292	<0.001
(III vs I)	3.379	2.120–5.386	<0.001	3.442	1.739–6.811	<0.001
(IV vs I)	3.012	1.524–5.594	<0.001	4.987	2.091–7.139	<0.001
Testing set (<i>n</i> = 223)						
Four-lncRNA risk score	1.094	0.971–1.233	0.014	1.112	0.984–1.258	0.009
Gender	1.489	0.855–2.591	0.159	1.799	0.996–3.248	0.051
Age	1.022	0.994–1.502	0.124	1.018	0.989–1.048	0.226
AJCC stage						
(II vs I)	0.462	0.157–1.361	0.016	1.85	0.911–3.758	0.086
(III vs I)	0.781	0.251–2.424	0.066	4.913	2.476–9.751	<0.001
(IV vs I)	2.012	0.659–6.413	0.021	2.411	0.785–7.401	0.124
Entire set (<i>n</i> = 446)						
Four-lncRNA risk score	1.216	1.135–1.303	<0.001	1.247	1.156–1.345	<0.001
Gender	1.152	0.800–1.657	0.447	1.301	0.890–1.902	0.174
Age	1.014	0.996–1.033	0.13	1.026	1.008–1.045	0.005
AJCC stage						
(II vs I)	4.678	2.376–9.211	<0.001	2.844	1.772–4.564	<0.001
(III vs I)	4.434	2.231–8.816	<0.001	4.641	2.888–7.458	<0.001
(IV vs I)	5.15	2.060–9.874	<0.001	3.749	1.873–7.501	<0.001

^a In univariable and multivariable Cox regression analyses, risk score, gender, and age were evaluated as continuous variables. AJCC stages were evaluated as category variables. $P < 0.05$ was considered statistically significant in all analyses.

3.4 Prognostic value of four-lncRNA signatures and analysis of PFS

The test set and the entire set were used to validate the prognostic model based on four lncRNAs. The risk score for each patient in the test set was calculated according to risk score formula. LUAD patients in the test set were divided into high-risk ($n = 111$) and low-risk groups ($n = 122$) using the threshold similar to that for the division of the training set. Kaplan–Meier analysis was performed in the test set. The results indicated that the survival time of patients in the high-risk group was significantly shorter than that in the low-risk group ($P = 0.0436$). The AUC obtained by ROC curve analysis in the test set was 0.778 (Figure 3a). We analyzed the entire data set in a similar manner. A high risk score based on the four-lncRNA signature was associated with poor OS of LUAD patients ($P = 2.65 \times 10^{-5}$). The AUC obtained by time-dependent ROC curve analysis was 0.769 (Figure 3b). Then we analyzed the PFS rate in the training set, test set, and the entire set (Figure 4a). The data indicated that

the PFS of patients in the high-risk group was significantly lower than that in the low-risk group, and the areas under the ROC curves (Figure 4b) were 0.664, 0.652, and 0.657 in the training set, test set, and the entire set, respectively. The results showed that the predictive value of the signature based on four lncRNAs had great potential for predicting prognosis.

3.5 Independence of the four-lncRNA-based biomarkers and clinical factors

The predictive ability of the four-lncRNA signature in combination with clinical factors was used to determine whether the signature is independent from other factors associated with survival. These factors included lncRNA risk score, sex, age, and AJCC tumor stages (Table 3). The data showed that the four-lncRNA risk score remained closely associated with the survival rate. Then, entire data set was assessed by stratified analysis. The patients

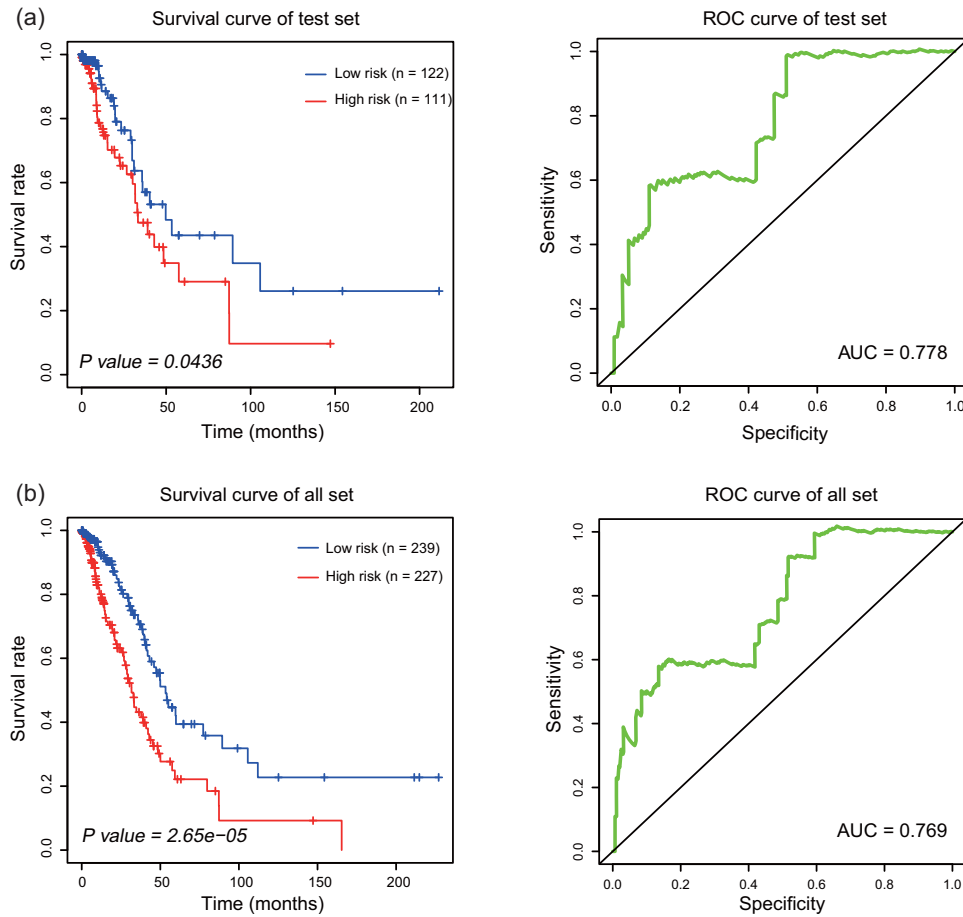


Figure 3: Prognostic value of four-lncRNA signature in the test set and entire set. (a) Kaplan–Meier curve and ROC curve analysis in the test set ($n = 233$). (b) Kaplan–Meier curve and ROC curve analysis in the entire set ($n = 466$).

were divided based on the threshold age of 60 years into the younger ($n = 147$) and older groups ($n = 319$) according to a threshold age of 60. LUAD patients in each group were divided according to the risk score into the high-risk and low-risk subgroups. The results suggested that the four-lncRNA signature can predict the prognosis of patients regardless of age (Figure 5a–c). We also analyzed the AJCC staging of the tumors. Stages I and II were classified as early stage ($n = 364$), and stages III and IV were classified as late stage ($n = 102$). The results suggest that the four-lncRNA risk score can adequately predict the survival of patients with LUAD (Figure 5d–f).

3.6 Analyses of somatic mutation and function prediction of four prognostic lncRNAs

Tumor-specific mutations are one of the important factors affecting the prognosis of the patients, and thus we

analyzed the gene mutations in the high-risk group and low-risk groups by using a waterfall diagram (Figure 6). TP63 and TTN were the genes with a high mutation frequency. EGFR accounted for 12 and 14% of mutations in the high-risk and low-risk groups, respectively, and ALK accounted for 5 and 4% of the mutations. The functions of four lncRNAs were predicted using coexpression network to investigate the mechanism of action of these lncRNAs in LUAD in detail. Spearman coefficients were calculated based on the expression of lncRNAs and protein-coding genes, and genes with a correlation coefficients >0.4 were selected. A total of 232 genes were used for pathway enrichment analysis. The web-based tool Metascape (<http://metascape.org/>) was used to analyze the biological function of these coexpressed genes. Most of these genes were enriched in 20 pathways (Figure 7), including interleukin-7-mediated signaling, nucleosome positioning, and regulation of glycogen biosynthetic process. These results suggested that four prognostic lncRNAs may be involved in glycogen metabolism and tumor immunity.

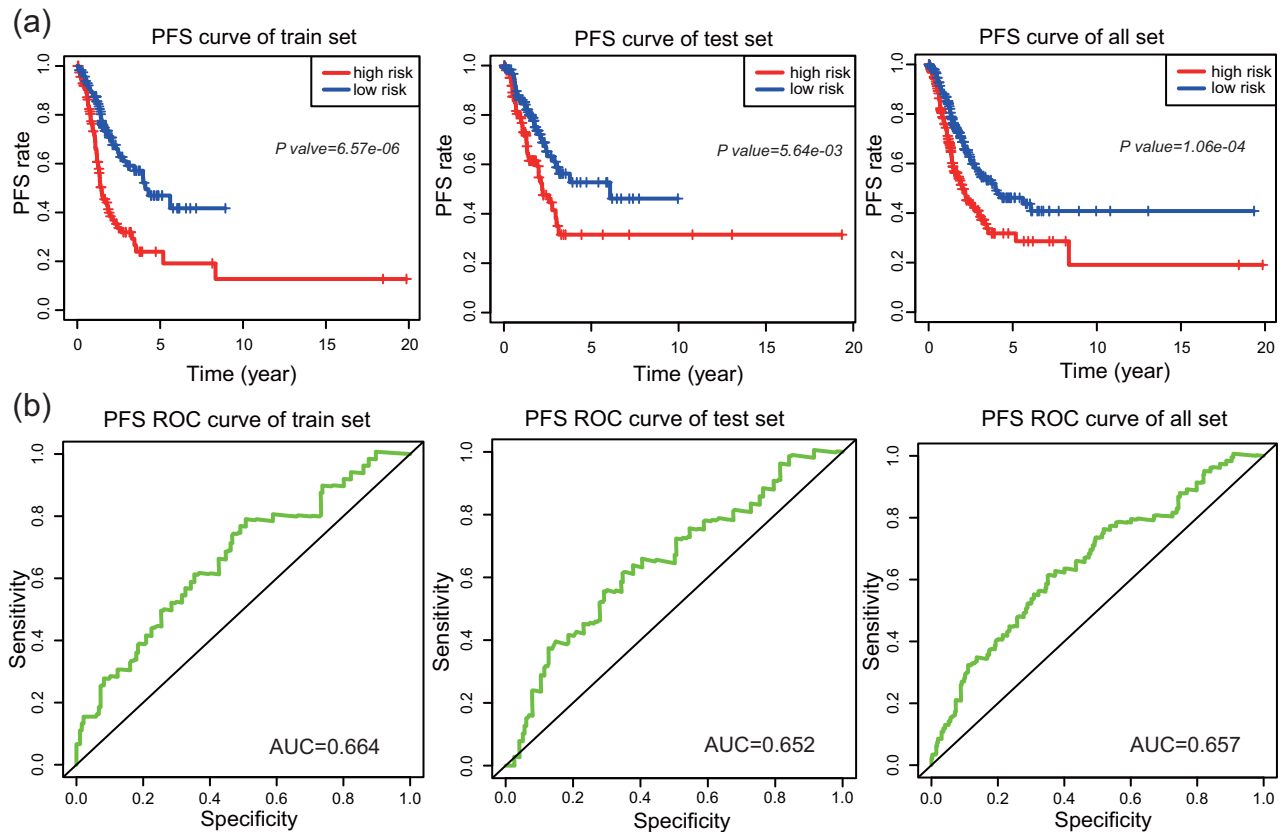


Figure 4: PFS analysis of the four-lncRNA signature. (a) The Kaplan–Meier curve of PFS. (b) The ROC curve of PFS in the training set, test set, and entire set.

3.7 Validation of external databases

The data of the GEO database were used to validate the four-lncRNA signature. Patients with LUAD in the GSE11969 data set were analyzed. Patients in the high-risk group had lower survival rate than those in the low-risk group, and the AUC was 0.705 (Figure 8). Therefore, four-lncRNA signature had predictive significance for the prognosis of patients with LUAD.

4 Discussion

LUAD has become one of the common killers of cancer patients, and its incidence has been recently increasing [20]. The main causes of lung cancer are smoking, air pollution, and gene variation, which have important effects on the development and clinical outcomes of LUAD. Effective therapies have been currently developed to target EGFR, KRAS, and MET; however, the prognosis of patients with advanced lung cancer remains unsatisfactory [21]. The lack of effective and reliable prognostic

biomarkers or models is the major problem in enhancing the clinical prognosis of LUAD patients. Recent studies demonstrated that lncRNAs play an important role in the development of tumors. Relevant studies have detected abnormal lncRNA expression in many cancer types, which is closely related to cancer progression [22–24]. Therefore, specific lncRNAs can be used as independent diagnostic and prognostic markers in cancer. lncRNAs were rarely reported to predict the prognosis of patients with LUAD, and hence, the present study aimed to provide a more effective prognostic marker for LUAD.

In this study, we analyzed the expression of lncRNAs in LUAD patients and identified four genes as prognostic markers. Subsequent analysis demonstrated that the four-lncRNA signature was an independent predictor of prognosis in patients. The performance of four-lncRNA signatures was evaluated by ROC analysis of the training, test, and entire sets, indicating a good prognostic value of the signature. Additionally, stratified analysis showed that the 4-lncRNA signature was more accurate than other factors in predicting the prognosis of LUAD patients. These four lncRNAs have not been reported previously. Therefore, the enrichment analysis of these lncRNAs was used to

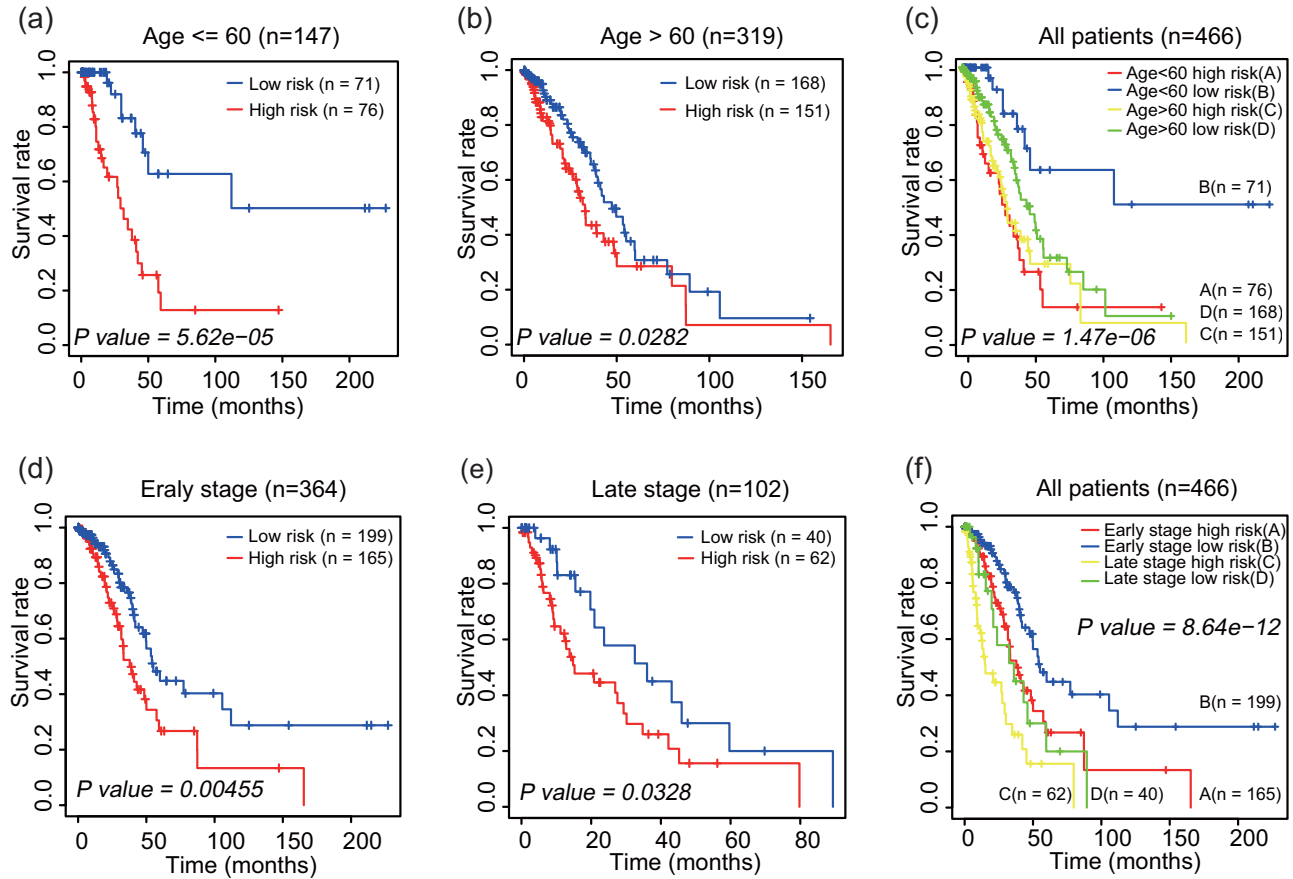


Figure 5: Analysis of associations of four-lncRNA in the groups with various age and tumor stages. (a) The Kaplan–Meier curve of young LUAD patients. (b) Kaplan–Meier curve of elderly LUAD patients. (c) The Kaplan–Meier curve of the entire LUAD patients. (d) The Kaplan–Meier curve of LUAD patients in early stage. (e) Kaplan–Meier curve of LUAD patients in late stage. (f) The Kaplan–Meier curve of the entire LUAD patients. The P value represents the difference between the two curves according to two-sided log rank test results.

understand the details of the regulatory mechanism of lncRNAs.

A large number of references demonstrated that microRNA can be used as diagnostic and prognostic markers [25,26]. For example, abnormally high expression of miR-21-5p is helpful for the diagnosis of non-small cell lung cancer [27]. Technological development resulted in the low costs of microchips and sequencing, and thus, aberrant lncRNA expression in tumors has received increasing attention and has been confirmed to influence the occurrence, development, and metastasis of the tumors. Moreover, lncRNAs have been reported to be more specific biomarkers than microRNAs in some cancer types [21,28,29]. Thus, lncRNAs have shown great potential as new molecular diagnostic biomarkers. For example, upregulation of LINC00963 can inhibit ubiquitination of PGK1 to activate the AKT/mTOR pathway, thus promoting the metastasis of lung cancer [30], and DLX6-AS1 promotes tumor proliferation by

regulating the JAK/STAT pathway [31]. The present study demonstrated that multiple lncRNAs predicted the prognosis of LUAD. Kaplan–Meier analysis showed that the risk score model has good prognostic ability. Analysis of the test set and entire set demonstrated the reliability of these biomarkers.

Analysis of the gene mutations in patients with LUAD identified common mutated genes and the genes with the highest mutation frequency. EGFR and ALK are not the most mutated; however, patients with mutations in the EGFR and ALK genes have poor prognosis [32]. These mutations play a very crucial role in tumor surgery, chemotherapy, and precision treatments [33–35]. The functions of four lncRNAs identified in the present study are not completely clear; hence, we used enrichment analysis to search for possible biological functions of these lncRNAs. Most of the enriched functions were concentrated in the processes related to nucleosomes, regulation of glycogen synthesis, and interleukin-7 (IL-7)-mediated signaling.

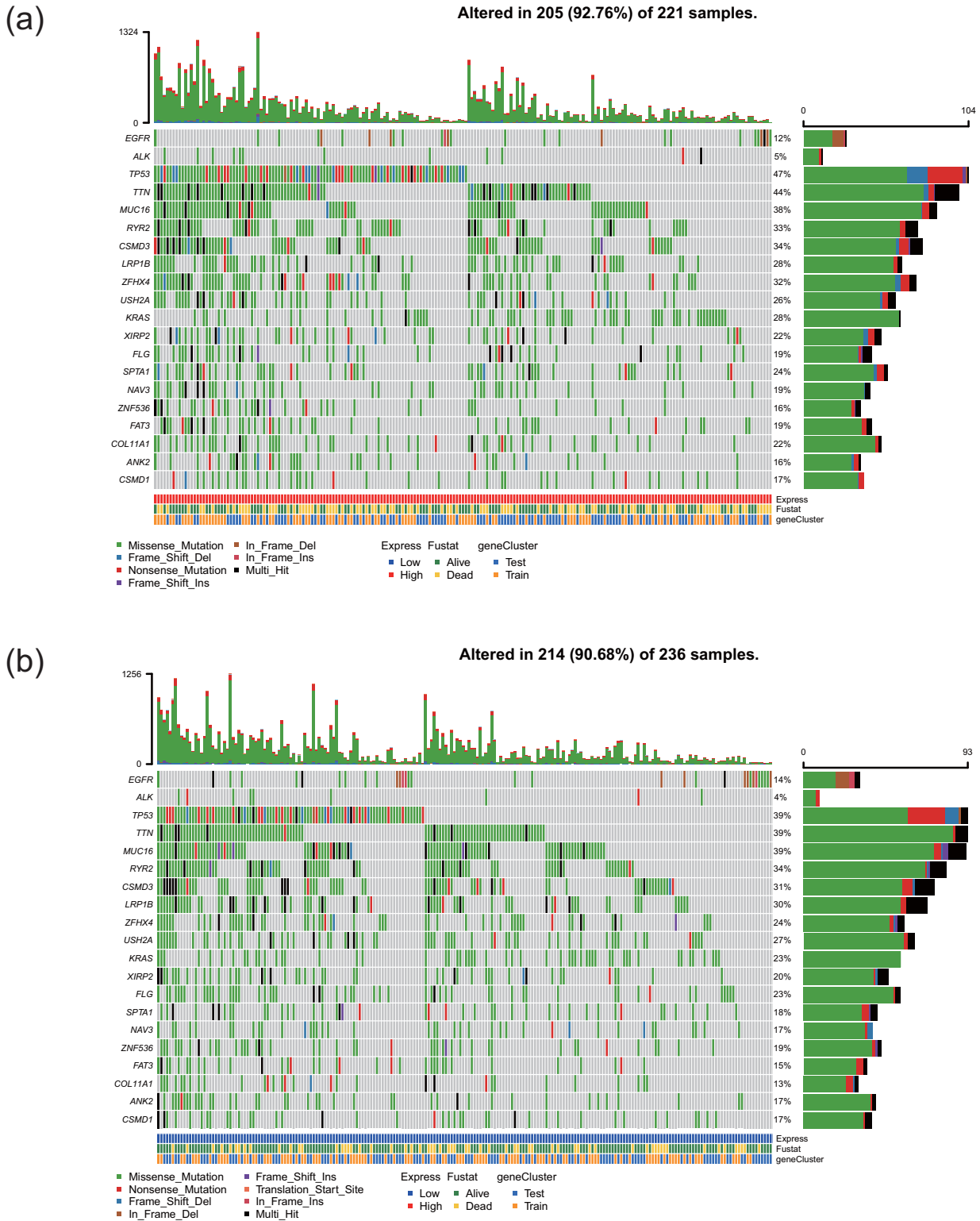


Figure 6: Analysis of somatic mutations in high-risk and low-risk groups. The waterfall plots of the high- (a) and low-risk (b) groups showing the distribution of mutations in the genes with the highest frequency of mutations. The central panel shows the type of mutation in each sample. The upper panel shows the mutation frequency for each sample. The bar chart on the right shows the frequency and type of mutations. The bottom panel is a legend describing the mutation types.

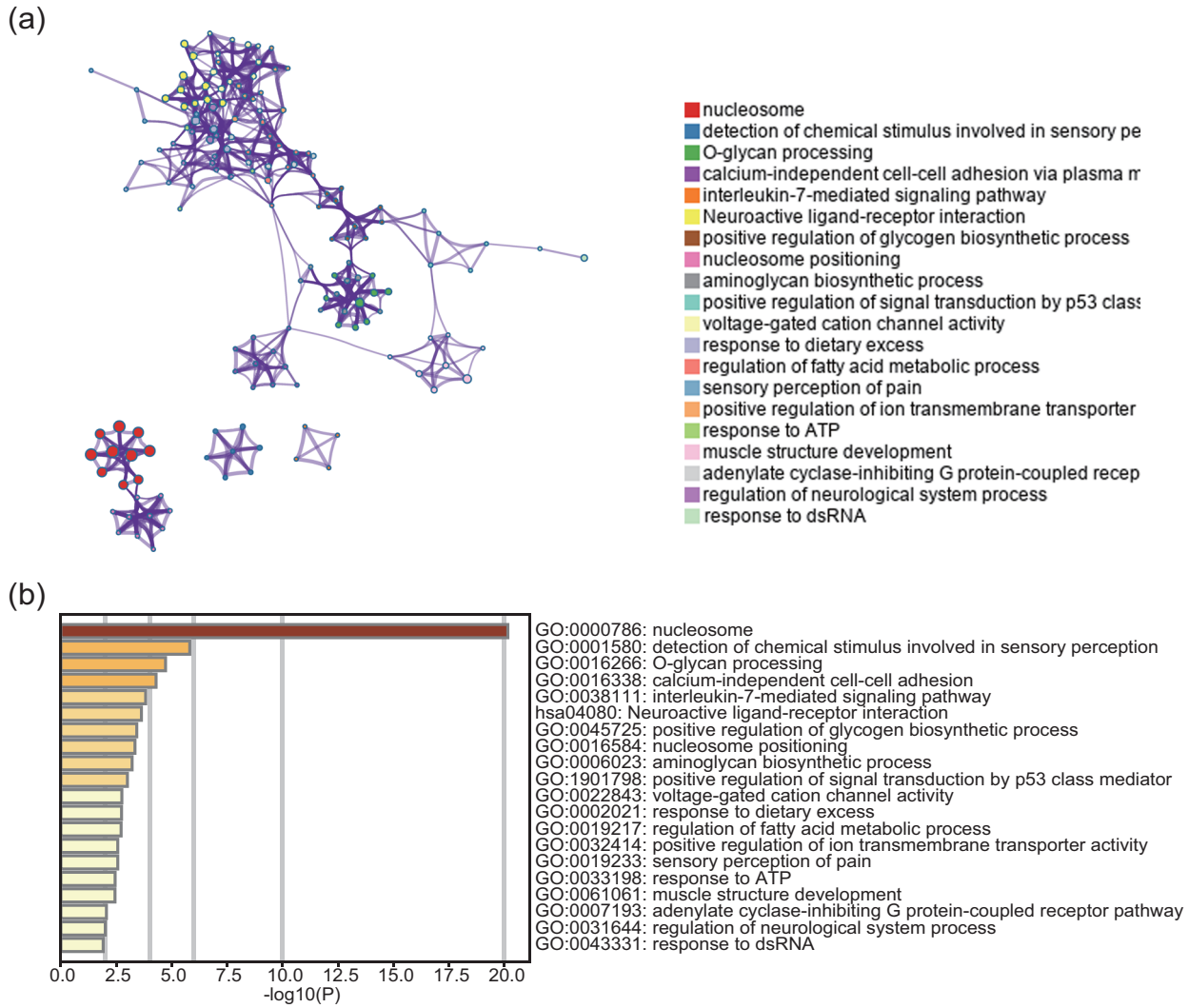


Figure 7: Functional enrichment results of the co-expressed protein-coding genes with four lncRNAs. (a) Significantly enriched pathways of the 232 correlated genes. (b) Functional enrichment diagram of the pathway. Each node represents a GO item. Node size represents the number of genes in the pathway.

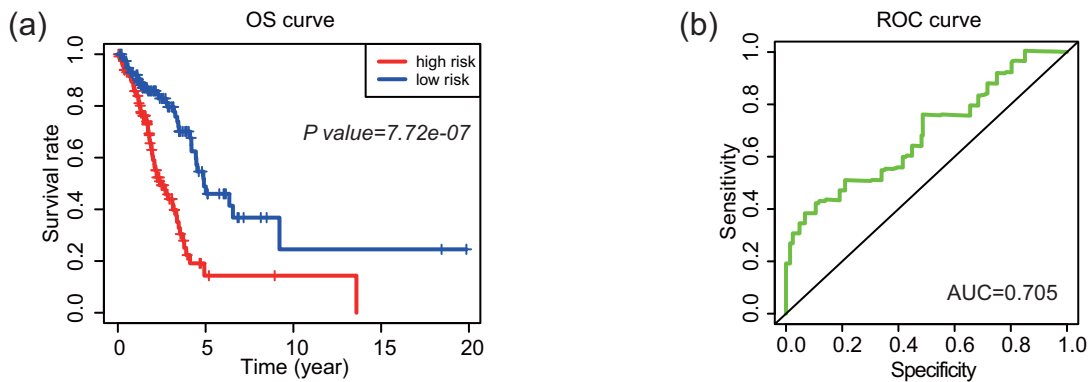


Figure 8: Verification of four-lncRNA signature in the GEO database. Kaplan–Meier survival curves for (a) verifying and ROC curve (b) for calculating the AUCs.

Farman *et al.* demonstrated that the nucleosome position around the transcriptional start site of a tumor suppressor gene in breast cancer is different from that in normal individuals [36]. High levels of glycogen have been reported in most types of cancer cells. The genes involved in glycogen synthesis enable cancer cells to use glucose for anaerobic glycolysis to provide glycogen required during brief energy deficits [37–40]. IL-7 regulates tumor proliferation, apoptosis, and tumor lymphangiogenesis [41–45]. Therefore, we hypothesized that these four lncRNAs may play an important role in tumor immune regulation and glycogen metabolism in the tumor microenvironment of LUAD.

The present study has several limitations. First, the lncRNA sequencing data were from the TCGA database, and these samples were mainly from African Americans and Caucasians. Therefore, additional studies are needed to determine whether the lncRNA signature is effective in Asian populations. Second, we analyzed and verified four lncRNAs based only on prognostic ability, and hence, the diagnostic and therapeutic capabilities of the signature require more additional investigations. Third, lncRNAs identified in the present study have not been reported in the literature, and further experiments are needed to analyze their functions and mechanisms.

On the whole, based on a large volume of expression profile data, a four-lncRNA signature was shown to have a good prognostic value in LUAD patients. Our results indicate that these four lncRNAs may play an important role in tumor immune regulation and in tumor microenvironment and constitute a reliable biomarker for the prediction of prognosis in LUAD patients.

Funding information: This research was funded by the National Natural Science Foundation (81671905).

Conflict of interest: The authors declare there are no competing interests.

Data availability statement: The data sets generated and analyzed in the present study are available in the TCGA data set LUAD project (<https://tcga-data.nci.nih.gov/tcga/>) and as a GEO data set.

References

- [1] Mao Y, Yang D, He J, Krasna MJ. Epidemiology of lung cancer. *Surg Oncol Clin N Am.* 2016;25(3):439–45.

- [2] Denisenko TV, Budkevich IN, Zhivotovsky B. Cell death-based treatment of lung adenocarcinoma. *Cell Death Dis.* 2018;9(2):117–31.
- [3] Ruiz-Cordero R, Devine WP. Targeted therapy and checkpoint immunotherapy in lung cancer. *Surg Pathol Clin.* 2020;13(1):17–33.
- [4] Lu GS, Li M, Xu CX, Wang D. APE1 stimulates EGFR-TKI resistance by activating Akt signaling through a redox-dependent mechanism in lung adenocarcinoma. *Cell Death Dis.* 2018;9(11):1111.
- [5] Macheleidt IF, Dalvi PS, Lim SY, Meemboor S, Meder L, Kasgen O, *et al.* Preclinical studies reveal that LSD1 inhibition results in tumor growth arrest in lung adenocarcinoma independently of driver mutations. *Mol Oncol.* 2018;12(11):1965–79.
- [6] Mizutani K, Guo X, Shioya A, Zhang J, Zheng J, Kurose N, *et al.* The impact of PRDX4 and the EGFR mutation status on cellular proliferation in lung adenocarcinoma. *Int J Med Sci.* 2019;16(9):1199–206.
- [7] Alidousty C, Baar T, Heydt C, Wagener-Rydzek S, Kron A, Wolf J, *et al.* Advance of theragnosis biomarkers in lung cancer: from clinical to molecular pathology and biology. *J Thorac Dis.* 2019;11(1):S3–8.
- [8] Kopp F, Mendell JT. Functional classification and experimental dissection of long noncoding RNAs. *Cell.* 2018;172(3):393–407.
- [9] Jandura A, Krause HM. The new RNA world: growing evidence for long noncoding RNA functionality. *Trends Genet.* 2017;33(10):665–76.
- [10] Vance KW, Ponting CP. Transcriptional regulatory functions of nuclear long noncoding RNAs. *Trends Genet.* 2014;30(8):348–55.
- [11] Yu X, Zhang Y. Identification of a long non-coding RNA signature for predicting prognosis and biomarkers in lung adenocarcinoma. *Oncol Lett.* 2020;19(4):2793–800.
- [12] Li M, Long S, Hu J, Wang Z, Geng C, Ou S. Systematic identification of lncRNA-based prognostic biomarkers for glioblastoma. *Aging (Albany NY).* 2019;11(21):9405–23.
- [13] Zong MZ, Shao Q, An XS. Expression and prognostic significance of long noncoding RNA AK001796 in esophageal squamous cell carcinoma. *Eur Rev Med Pharmacol Sci.* 2019;23(1):181–6.
- [14] Baassiri A, Nassar FA-O, Mukherji DA-O, Shamseddine AA-O, Nasr RA-O, Temraz S. Exosomal non coding RNA in liquid biopsies as a promising biomarker for colorectal cancer. *Int J Mol Sci.* 2020;21(4):E1398.
- [15] Chaudhary R, Wang X, Cao B, De La Iglesia J, Masannat J, Song F, *et al.* Long noncoding RNA, LINC00460, as a prognostic biomarker in head and neck squamous cell carcinoma (HNSCC). *Am J Transl Res.* 2020;15(2):684–96.
- [16] Wu ZY, Wang SG, Li Q, Zhao QS, Shao MM. Identification of 10 differently expressed lncRNAs as prognostic biomarkers for prostate adenocarcinoma. *Math Biosci Eng.* 2019;17(3):2037–47.
- [17] Zhang P, Shi L, Song L, Long Y, Yuan K, Ding W, *et al.* lncRNA CRNDE and lncRNA SNHG7 are promising biomarkers for prognosis in synchronous colorectal liver metastasis following hepatectomy. *Cancer Manag Res.* 2020;12(3):1681–92.

- [18] Derrien T, Johnson R, Bussotti G, Tanzer A, Tanzer A, Djebali S, et al. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res.* 2012;22(9):1775–89.
- [19] Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics.* 2010;26(1):139–40.
- [20] Siegel RLA-O, Miller KDA-O, Jemal A. *Cancer statistics, 2020.* *CA Cancer J Clin.* 2020;70(1):7–30.
- [21] Jiang L, Li Z, Wang R. Long noncoding RNAs in lung cancer: Regulation patterns, biologic function and diagnosis implications (Review). *Int J Oncol.* 2019;55(3):585–96.
- [22] Liu W, Yang YJ, An Q. LINC00963 promotes ovarian cancer proliferation, migration and EMT via the miR-378g/CHI3L1 axis. *Cancer Manag Res.* 2020;21(12):463–73.
- [23] Yang G, He F, Duan H, Shen J, Dong Q. lncRNA FLVCR-AS1 promotes osteosarcoma growth by targeting miR381-3p/CCND1. *Onco Targets Ther.* 2020;9(13):163–72.
- [24] Zhao L, Zhou Y, Zhao Y, Li Q, Zhou J, Mao Y. Long non-coding RNA TUSC8 inhibits breast cancer growth and metastasis via miR-190b-5p/MYLIP axis. *Aging (Albany NY).* 2020;12(3):2974–91.
- [25] Kuroda H, Yoshida T, Arimura T, Mizuno T, Sakakura N, Yatabe Y, et al. Contribution of smoking habit to the prognosis of stage I KRAS-mutated non-small cell lung cancer. *Cancer Biomark.* 2018;23(3):419–26.
- [26] Yang Y, Ding L, Hu Q, Xia J, Sun J, Wang X, et al. MicroRNA-218 functions as a tumor suppressor in lung cancer by targeting IL-6/STAT3 and negatively correlates with poor prognosis. *Mol Cancer.* 2017;16(1):141.
- [27] Yang Y, Hu Z, Zhou Y, Zhao G, Lei Y, Li G, et al. The clinical use of circulating microRNAs as non-invasive diagnostic biomarkers for lung cancers. *Oncotarget.* 2017;8(52):90197–214.
- [28] Prensner JR, Iyer MK, Balbin OA, Dhanasekaran SM, Cao Q, Brenner JC, et al. Transcriptome sequencing across a prostate cancer cohort identifies PCAT-1, an unannotated lincRNA implicated in disease progression. *Nat Biotechnol.* 2011;29(8):742–9.
- [29] Wang YL, Shao J, Wu X, Li T, Xu M, Shi D. A long non-coding RNA signature for predicting survival in patients with colorectal cancer. *Oncotarget.* 2018;9(31):21687–95.
- [30] Yu T, Zhao Y, Hu Z, Li J, Chu D, Zhang J, et al. MetaLnc9 facilitates lung cancer metastasis via a PGK1-activated AKT/mTOR pathway. *Cancer Res.* 2017;77(21):5782–94.
- [31] Li J, Li P, Zhao W, Yang R, Chen S, Bai Y, et al. Expression of long non-coding RNA DLX6-AS1 in lung adenocarcinoma. *Cancer Cell Int.* 2015;15(2):48.
- [32] Zhuang X, Zhao CA-O, Li J, Su C, Chen X, Ren SA-O, et al. Clinical features and therapeutic options in non-small cell lung cancer patients with concomitant mutations of EGFR, ALK, ROS1, KRAS or BRAF. *Cancer Med.* 2019;8(6):2858–66.
- [33] Arbour KC, Riely GJ. Systemic therapy for locally advanced and metastatic non-small cell lung cancer: a review. *JAMA.* 2019;322(8):764–74.
- [34] Li P, Gao Q, Jiang X, Zhan Z, Yan Q, Li Z, et al. Comparison of clinicopathological features and prognosis between ALK rearrangements and EGFR mutations in surgically resected early-stage lung adenocarcinoma. *J Cancer.* 2019;10(1):61–71.
- [35] Skoulidis FA-O, Heymach JV. Co-occurring genomic alterations in non-small-cell lung cancer biology and therapy. *Nat Rev Cancer.* 2019;19(9):495–509.
- [36] Farman FU, Iqbal M, Azam M, Saeed M. Nucleosomes positioning around transcriptional start site of tumor suppressor (Rbl2/p130) gene in breast cancer. *Mol Biol Rep.* 2018;45(2):185–94.
- [37] Pelletier J, Bellot G, Gounon P, Lacas-Gervais S, Pouyssegur J, Mazure NM. Glycogen synthesis is induced in hypoxia by the hypoxia-inducible factor and promotes cancer cell survival. *Front Oncol.* 2012;2(2):18.
- [38] Ritterson Lew C, Guin S, Theodorescu D. Targeting glycogen metabolism in bladder cancer. *Nat Rev Urol.* 2015;12(7):383–91.
- [39] Zhou Y, Jin Z, Wang C. Glycogen phosphorylase B promotes ovarian cancer progression via Wnt/beta-catenin signaling and is regulated by miR-133a-3p. *Biomed Pharmacother.* 2019;120(3):109.
- [40] Zois CE, Harris AL. Glycogen metabolism has a key role in the cancer microenvironment and provides new targets for cancer therapy. *J Mol Med (Berl).* 2016;94(2):137–54.
- [41] Jian M, Yunjia Z, Zhiying D, Yanduo J, Guocheng J. Interleukin 7 receptor activates PI3K/Akt/mTOR signaling pathway via downregulation of Beclin-1 in lung cancer. *Mol Carcinog.* 2019;58(3):358–65.
- [42] Li S, Wang Z, Zhang G, Fu J, Zhang X. Interleukin-7 promotes lung-resident CD14(+) monocytes activity in patients with lung squamous carcinoma. *Int Immunopharmacol.* 2019;67(2):202–10.
- [43] Lin J, Zhu Z, Xiao H, Wakefield MR, Ding VA, Bai Q, et al. The role of IL-7 in immunity and cancer. *Anticancer Res.* 2017;37(3):963–7.
- [44] Xu X, Sun Q, Mei Y, Liu Y, Zhao L. Newcastle disease virus co-expressing interleukin 7 and interleukin 15 modified tumor cells as a vaccine for cancer immunotherapy. *Cancer Sci.* 2018;109(2):279–88.
- [45] Zhang C, Su P, Chen W, Li Q, Dai R, Cheng Y, et al. Genetic polymorphisms in IL-7 and IL-7R are correlated with lung cancer risk in the Chinese Han population. *Cancer Manag Res.* 2019;11:5393–401.