

Research Article

Development and Validation of a Seven-Gene Signature for Predicting the Prognosis of Lung Adenocarcinoma

Yingqing Zhang ¹, Xiaoping Zhang ², Xiaodong Lv ¹, Ming Zhang ¹, Xixi Gao ¹,
Jialiangu Liu ¹, Yufen Xu ³, Zhixian Fang ¹, and Wenyu Chen ¹

¹Department of Respiration, The First Hospital of Jiaxing, Affiliated Hospital of Jiaxing University, Jiaxing 314000, China

²Department of Science and Education, The First Hospital of Jiaxing, Affiliated Hospital of Jiaxing University, Jiaxing 314000, China

³Department of Oncology, The First Hospital of Jiaxing, Affiliated Hospital of Jiaxing University, Jiaxing 314000, China

Correspondence should be addressed to Yufen Xu; xyfxyf84842@163.com, Zhixian Fang; fang_zxzhixian@163.com, and Wenyu Chen; chwyjx@163.com

Yingqing Zhang, Xiaoping Zhang, and Xiaodong Lv contributed equally to this work.

Received 23 April 2020; Accepted 23 June 2020; Published 17 August 2020

Guest Editor: Tao Huang

Copyright © 2020 Yingqing Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Background. Prognosis is a main factor affecting the survival of patients with lung adenocarcinoma (LUAD), yet no robust prognostic model of high effectiveness has been developed. This study is aimed at constructing a stable and practicable gene signature-based model via bioinformatics methods for predicting the prognosis of LUAD sufferers. **Methods.** The mRNA expression data were accessed from the TCGA-LUAD dataset, and paired clinical information was collected from the GDC website. R package “edgeR” was employed to select the differentially expressed genes (DEGs), which were then used for the construction of a gene signature-based model via univariate COX, Lasso, and multivariate COX regression analyses. Kaplan-Meier and ROC survival analyses were conducted to comprehensively evaluate the performance of the model in predicting LUAD prognosis, and an independent dataset GSE26939 was accessed for further validation. **Results.** Totally, 1,655 DEGs were obtained, and a 7-gene signature-based risk score was developed and formulated as $\text{risk_score} = 0.000245 * \text{NTSR1} + (7.13E - 05) * \text{RHOV} + 0.000505 * \text{KLK8} + (7.01E - 05) * \text{TNS4} + 0.000288 * \text{C1QTNF6} + 0.00044 * \text{IVL} + 0.000161 * \text{B4GALNT2}$. Kaplan-Meier survival curves revealed that the survival rate of patients in the high-risk group was lower in both the TCGA-LUAD dataset and GSE26939 relative to that of patients in the low-risk group. The relationship between the risk score and clinical characteristics was further investigated, finding that the model was effective in prognosis prediction in the patients with different age (age > 65, age < 65) and TNM stage (N0&N1, T1&T2, and tumor stage I/II). In sum, our study provides a robust predictive model for LUAD prognosis, which boosts the clinical research on LUAD and helps to explore the mechanism underlying the occurrence and progression of LUAD.

1. Introduction

Lung cancer is a kind of malignant tumor with the morbidity (13% both in male and female) and mortality (24% in male and 23% in female), respectively, ranking second and top worldwide, according to the latest data released in *A Cancer Journal for Clinicians* [1]. Lung cancer can be classified into small-cell lung cancer (SCLC) and non-small-cell lung cancer (NSCLC), of which NSCLC sufferers are in the majority of the total lung cancer cases (around 80%). Lung adenocarcinoma

(LUAD), the main histological subtype of NSCLC, takes up over 40% among the overall lung cancer morbidity [2]. Given that around 80% of patients with lung cancer are diagnosed in middle and advanced stages, surgery is no more an available option, resulting in unfavorable outcomes with a 5-year overall survival (OS) rate of nearly 17% [3, 4]. While distant metastasis and relapse are main causes of poor cancer treatment and prognosis [5, 6], identification of cancer-associated genes and independent prognostic factors as well as investigation of their impact on tumor progression and prognosis is beneficial for

the implementation of precision medicine and helps to raise the cure rate and improve the prognosis. With the development of gene chip technology and RNA sequencing, gene expression profiles have been widely applied in the prediction of LUAD prognosis. For example, PHLPP2 has been reported as a novel biomarker in NSCLC metastasis and prognosis [7]. Thyroid transcription factor-1 is considered as a prognostic marker indicating the presence or absence of EGFR-sensitizing mutations in stage IV LUAD [8]. And the elevated CX3CL1 mRNA expression is found to be a positive factor involved in LUAD prognosis [9]. However, due to the variety of methods, experimental platforms, batch effects, or other factors, discrepancy appears in the genes screened for prognosis prediction. Besides, the prognostic models constructed might be only practicable in the current experimental samples, while the performance in other independent datasets is less pronounced. Therefore, it is urgent to find a model that is practicable in various datasets, making its value realized in different clinical researches.

In the present study, HTSeq-Counts data of LUAD comprising 522 tumor samples and 58 normal samples were accessed from the TCGA database. Based on the data, survival-associated genes were selected using univariate COX regression analysis, after which the Lasso regression model was constructed to rule out the genes of a relatively stronger correlation to prevent model overfitting. Afterwards, a series of multivariate COX regression models were established, and the optimal model was identified in line with the Akaike Information Criterion (AIC). To validate and evaluate the performance of the model in predicting LUAD prognosis, various aspects were taken into account, finding that the model was effective in the training set and testing set, and its performance in patients with different age and TNM stage was validated to be good as well. Furthermore, the model also exhibited a good ability in predicting the prognosis of LUAD patients in an independent dataset GSE26939. To sum, our study constructs a robust gene signature-based model available for predicting the prognosis of LUAD patients, which helps the clinical research on LUAD and lays a foundation for the future investigation on the molecular mechanism underlying LUAD occurrence and progression.

2. Methods and Materials

2.1. Data Collection and Preprocessing. HTSeq-Counts data of LUAD (including 522 tumor samples and 58 normal samples) were obtained from the TCGA database (<https://portal.gdc.cancer.gov/>) and then used for differential analysis with the aid of R package “edgeR” ($|\log_{2}FC| \geq 2$, adj. FDR < 0.05). The corresponding clinical information of TCGA-LUAD patients was collected in the GDC website (<https://portal.gdc.cancer.gov/>). Patients who were followed up less than 30 days were excluded in this study, and totally, 460 TCGA-LUAD patients were included eventually. Besides, to further verify the validity of the prognostic model, an independent dataset GSE26939 (including 115 patients with LUAD) and matched clinical information were accessed from the GEO database (<https://www.ncbi.nlm.nih.gov/geo/>).

2.2. Candidate Gene Selection. Differentially expressed genes (DEGs) screened out by “edgeR” were randomized into the training set and testing set (5:5) and then subjected to univariate COX regression analysis for identifying the genes associated with the survival of patients with LUAD. The Lasso regression model was employed to further analyze these survival-related genes to exclude the genes with a relatively higher correlation, contributing to the decrease in the complexity of the prognostic model [10] and helping to find the optimal signature genes.

2.3. Prognostic Model Construction. Candidate genes selected by Lasso regression analysis were used to construct multivariate COX models, and the Akaike Information Criterion (AIC) was referenced to find the optimal prognostic model.

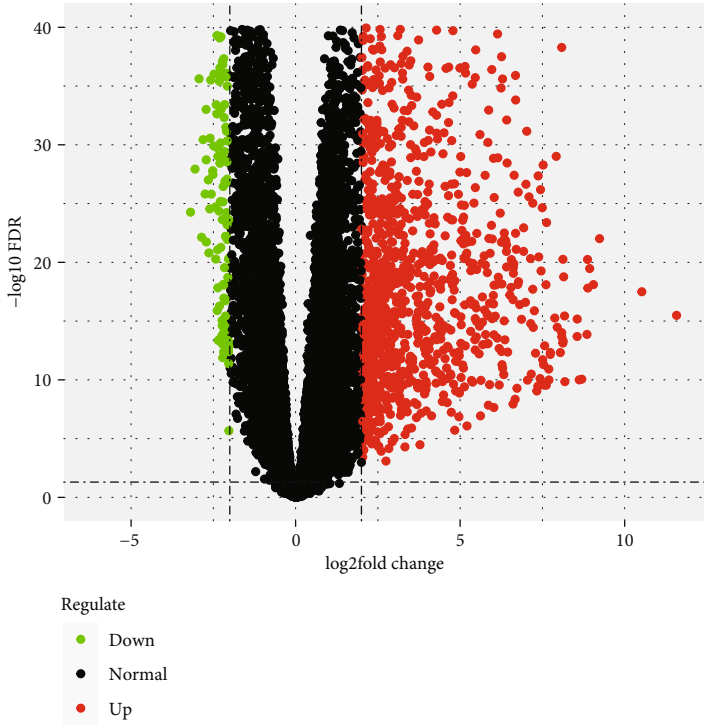
2.4. Stability and Validity Verification. Patients in the training set and testing set were conferred a risk score and grouped into the high-risk group and the low-risk group based on the median score. The Kaplan-Meier method was conducted to compare the survival of patients in two groups, and log-rank was performed to calculate the *p* value. Meanwhile, ROC analysis was carried out to analyze the performance of the model in predicting the prognosis of LUAD patients, and an independent dataset GSE26939 was applied for the verification of the model’s validity.

3. Results

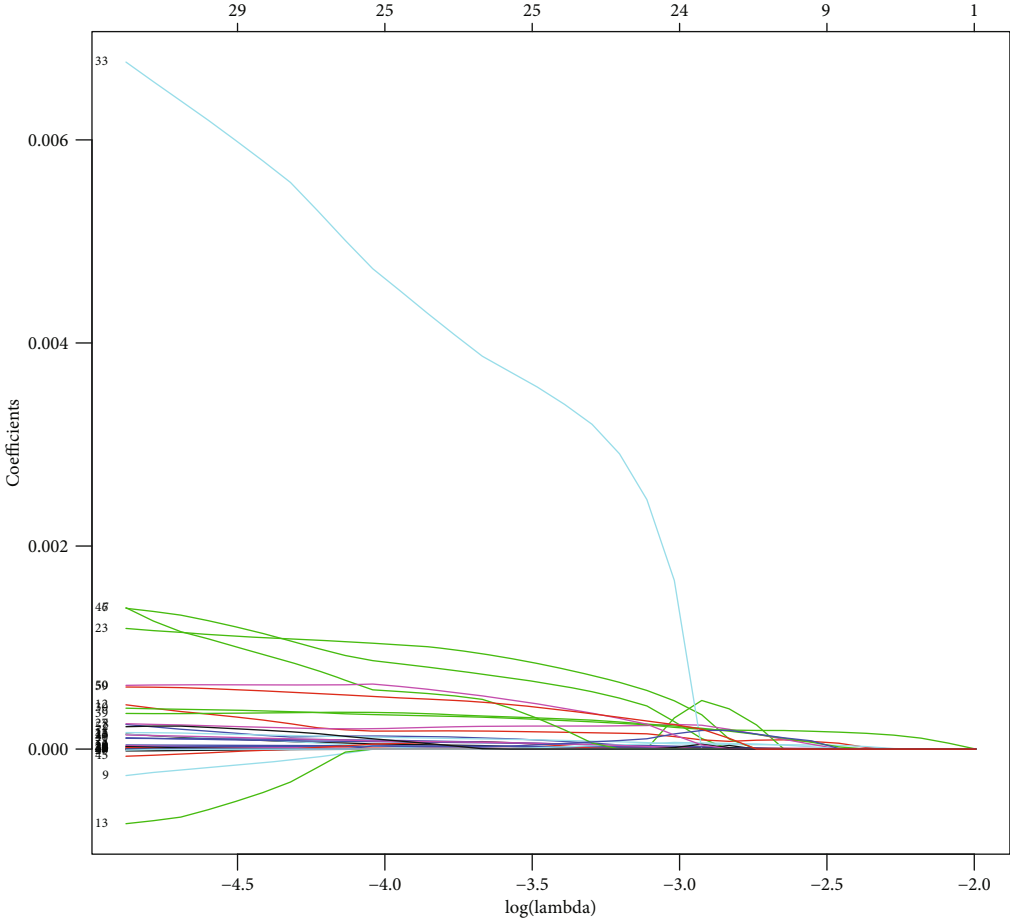
3.1. Identification of Candidate Genes. In total, 1,655 DEGs were obtained via differential analysis based on the TCGA-LUAD dataset (Figure 1(a)) and randomly assigned to the training set and testing set (5:5). Univariate COX analysis was performed to screen survival-related genes from the training set with the cut-off set as *p* value = 0.01, and initially, 60 genes were screened out as shown in Supplementary Table 1 (the top 20 genes associated with survival are listed in Table 1). Subsequently, these genes were analyzed in a Lasso regression model. Genes with a relatively higher correlation were removed to lower the complexity of the prognostic model, and finally, 9 candidate signature genes were identified, namely, NTSR1, RHOV, KLK8, TNS4, C1QTNF6, FAM83A, IVL, B4GALNT2, and CREG2 (Figures 1(b) and 1(c)).

3.2. Construction of a 7-Genes Signature-Based Prognostic Model for LUAD. A series of multivariate COX models were constructed based on the candidate genes, and the optimal model was then selected in line with AIC as shown in Table 2. A 7-gene signature-based risk score formula was established as Risk score = 0.000245 * NTSR1 + (7.13E - 05) * RHOV + 0.000505 * KLK8 + (7.01E - 05) * TNS4 + 0.000288 * C1QTNF6 + 0.00044 * IVL + 0.000161 * B4 GALNT2.

3.3. Evaluation of the 7-Genes Signature-Based Model in Predicting the Survival of LUAD Patients. Based on the formula, the 7-gene signature-based risk score of each patient in the training set and testing set was calculated, and patients were classified into the high-risk group and the low-risk



(a)



(b)

FIGURE 1: Continued.

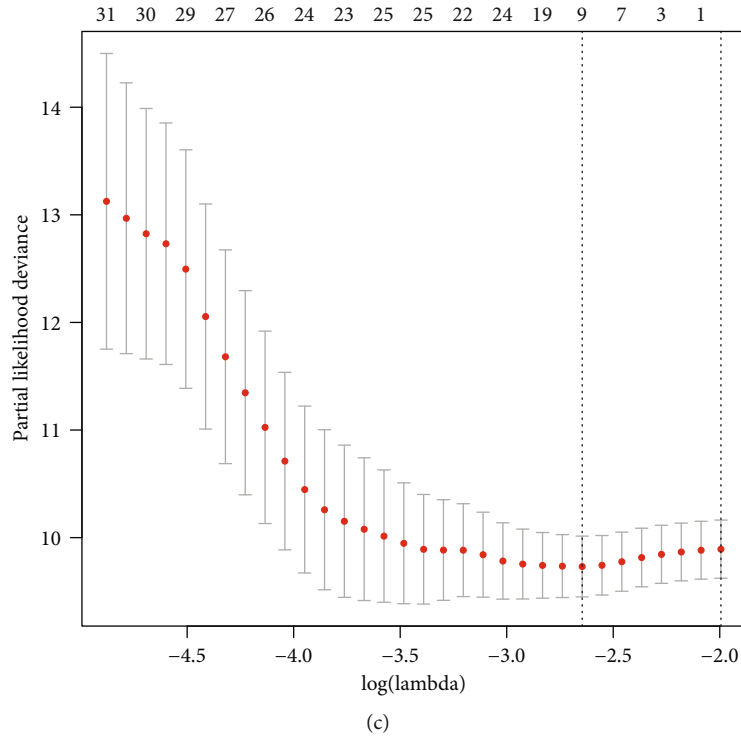


FIGURE 1: Candidate gene selection. (a) Volcano plot of the DEGs in the TCGA-LUAD dataset; (b) regression coefficients in Lasso regression analysis; (c) selection of lambda in the Lasso regression model through 10-fold crossvalidation method.

TABLE 1: Top 20 genes associated with survival.

ID	HR	HR.95L	HR.95H	<i>p</i> value
FAM83A	1.00004	1.000026	1.000053	$8.36E - 09$
RHOV	1.000144	1.000093	1.000196	$4.16E - 08$
TNS4	1.000136	1.000083	1.00019	$5.85E - 07$
C1QTNF6	1.000485	1.000288	1.000683	$1.51E - 06$
COL7A1	1.000145	1.000079	1.000211	$1.86E - 05$
TRPA1	1.002849	1.001522	1.004177	$2.53E - 05$
CREG2	1.000977	1.000508	1.001445	$4.34E - 05$
UNC5D	1.000503	1.000261	1.000746	$4.69E - 05$
INHA	1.000112	1.000056	1.000167	$8.79E - 05$
AHNAK2	1.000072	1.000035	1.000108	0.000107
OLFM4	1.000025	1.000012	1.000037	0.00011
KLK8	1.000654	1.000312	1.000996	0.000177
FAM83B	1.001885	1.000896	1.002875	0.000186
KRT6A	1.000015	1.000007	1.000022	0.000285
MUCL1	1.000166	1.000076	1.000255	0.000305
MFI2	1.000104	1.000047	1.000162	0.000381
ERO1L	1.000042	1.000019	1.000065	0.000388
PSG5	1.053914	1.023725	1.084994	0.000398
NTSR1	1.000394	1.000171	1.000616	0.00054
TMPRSS11E	1.000327	1.000141	1.000513	0.000562

TABLE 2: The 7 genes in the optimal multivariate COX regression model.

ID	Coefficient	HR	HR.95L	HR.95H	p value
NTSR1	0.000244865	1.000244895	0.999995	1.000495	0.05509
RHOV	$7.13E-05$	1.000071315	0.999989	1.000153	0.08795
KLK8	0.000505343	1.000505471	1.000028	1.000983	0.037875
TNS4	$7.01E-05$	1.000070108	0.999986	1.000154	0.103366
C1QTNF6	0.000287673	1.000287714	1.000022	1.000553	0.033585
IVL	0.000440486	1.000440583	1.000203	1.000678	0.000272
B4GALNT2	0.000161186	1.000161199	1.000032	1.00029	0.014511

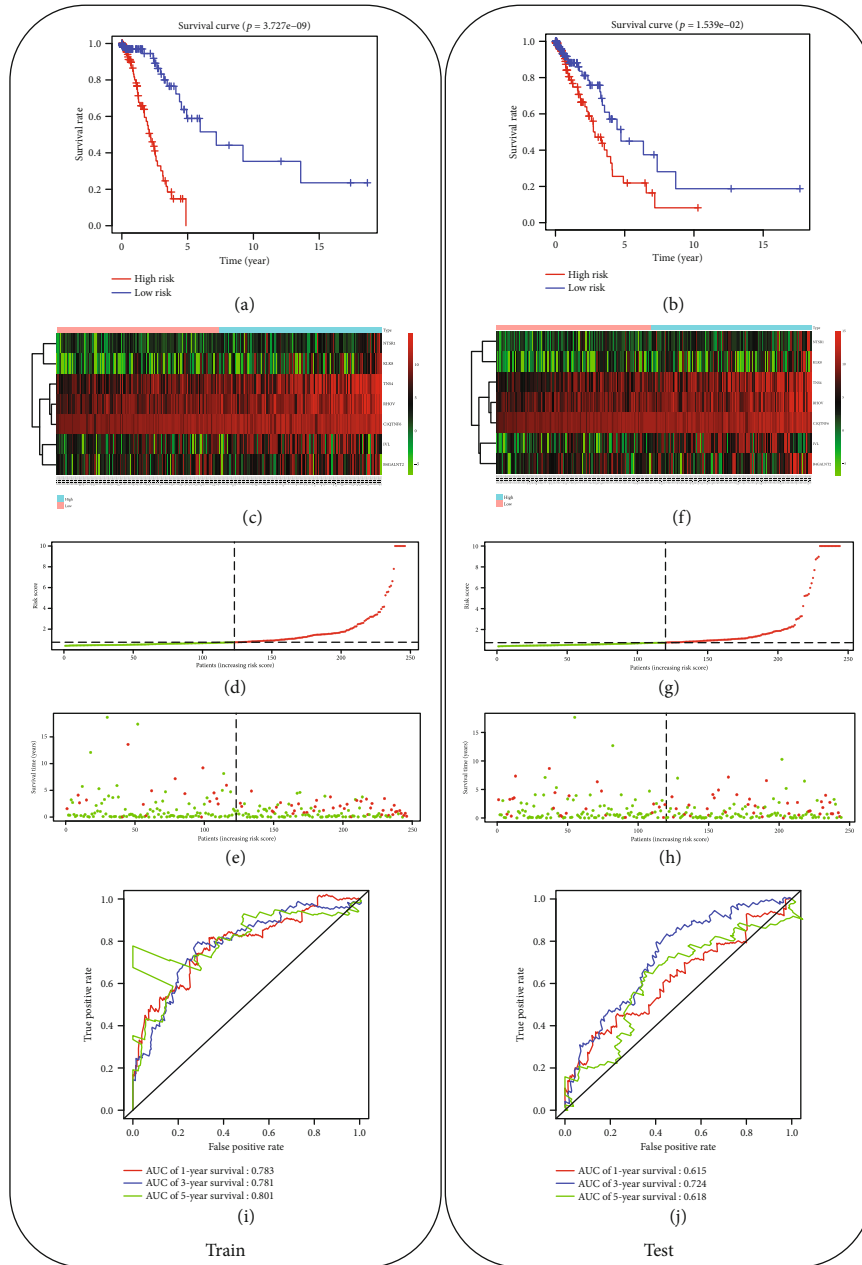


FIGURE 2: Evaluation of the 7-gene signature-based model in predicting the prognosis of LUAD patients. (a, b) The survival time between the patients of the high- and low-risk groups within the training set and testing set was analyzed by Kaplan-Meier analysis; (c–e) heat map for the 7 genes, as well as the risk score distribution and survival of patients in the training set; (f, g, h) heat map for the 7 genes, as well as the risk score distribution and survival of patients in the testing set; (i, j) ROC curves of the 7-gene signature-based model in two sets.

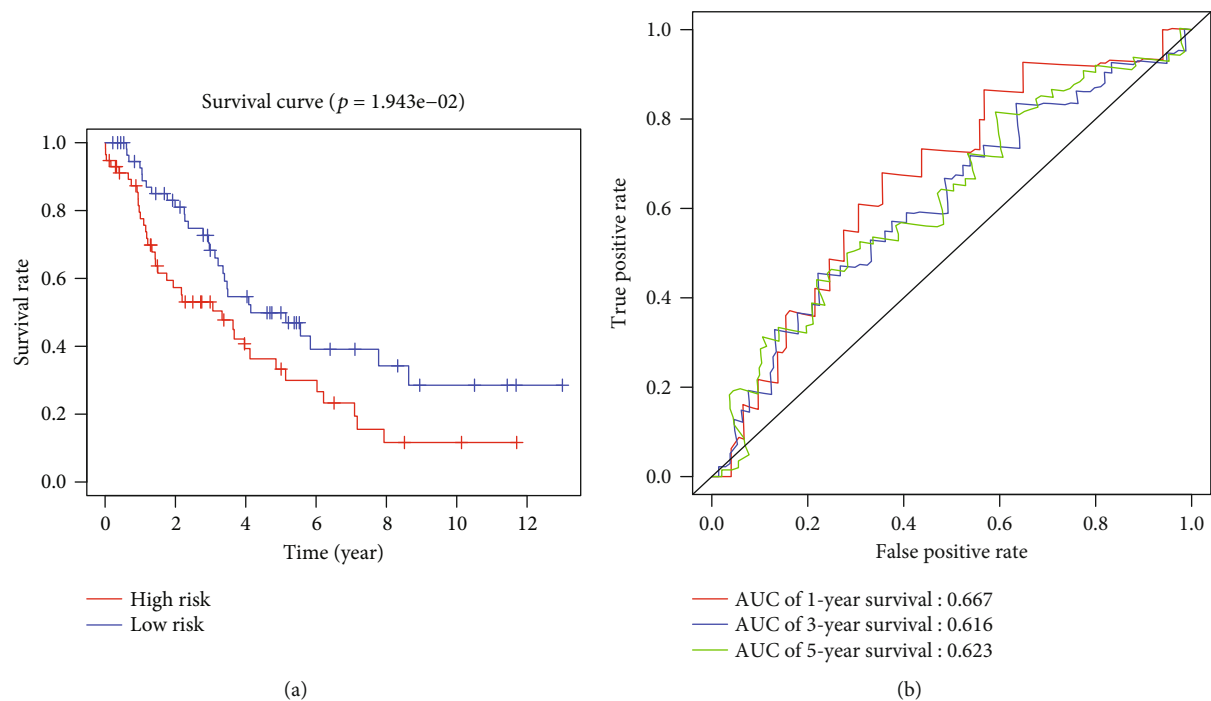


FIGURE 3: Verification of stability and validity of the prognostic model for LUAD with an independent dataset GSE26939. (a) Survival difference between the patients in the high- and low-risk groups was revealed by Kaplan-Meier analysis; (b) ROC curves of the 7-gene model using the independent dataset.

group according to the median score. Kaplan-Meier curves and log-rank test were used to compare the survival of the two groups in two independent sets, finding that patients in the high-risk group had poorer survival relative to those in the low-risk group in both sets ($p < 0.05$) (Figures 2(a) and 2(b)). To better know the expression level of the 7 genes, risk score distribution, and survival of the patients in two sets, data in the training set and testing set were obtained and plotted in Figures 2(c)–2(e) and 2(f)–2(h), respectively.

ROC analysis was conducted using the survivalROC package for the verification of the model performance in the training set and testing set. AUC values of 1-, 3-, and 5-year survival were calculated, with those in the training set as 0.783, 0.781, and 0.801 (Figure 2(i)) and in the testing set as 0.615, 0.724, and 0.618 (Figure 2(j)), respectively. Taken together, the 7-gene signature-based model was demonstrated to be capable of predicting the prognosis of LUAD patients.

3.4. Verification of Stability and Validity of the Prognostic Model for LUAD with an Independent Dataset GSE26939.

An independent dataset GSE26939 from the GEO database was applied to further verify the validity and stability of the 7-gene model. The same as the above procedures, patients were divided into the high-risk and low-risk groups based on the median risk score, and survival comparison was performed using Kaplan-Meier as shown in Figure 3(a), indicating the lower survival rate in the patients of the high-risk group ($p < 0.05$). Thereafter, ROC analysis was performed for further verification, with the AUC values of 1-, 3-, and 5-year survival of 0.667, 0.616, and 0.623

(Figure 3(b)), respectively. Collectively, this 7-gene model was practicable in other independent datasets.

3.5. Prognostic Impact of the Model on Clinical Characteristics.

To further discuss the correlation of the 7-gene signature-based risk score with the TNM (Tumor Node Metastasis) stage and overall survival (OS) of LUAD patients, matched clinical information of the training set and testing set was collected and is listed in Tables 3 and 4. The relationship between the risk score and TNM stage was explored, revealing that the risk score was significantly associated with pathologic T, N, and tumor stages of patients in both the training set (Figures 4(a)–4(c)) and testing set (Figures 4(d)–4(f)) ($p < 0.05$). Moreover, the performance of the model in predicting the prognosis of patients with different clinical characteristics in the two sets was investigated (Figure 4(g)), finding good performance on patients in different age and clinical stage (age > 65, age < 65, N0&N1, T1&T2, and tumor stage I/II). While in the independent dataset GSE26939, such correlation was less pronounced (Supplementary Table 2). Altogether, this 7-gene signature-based risk score model was a useful prognosis predictor in patients with different clinical characteristics and could be served as a novel biomarker in LUAD treatment.

4. Discussion

Lung cancer, with its mortality ranking top globally, often appears to be in middle and advanced stages when being initially diagnosed in most patients, and surgery is no more

TABLE 3: Clinical information of LUAD patients in the training set.

	Low risk (n = 116)	High risk (n = 117)	p value
<i>Age</i>			
<65	51 (44.0%)	59 (50.4%)	0.392
>65	65 (56.0%)	58 (49.6%)	
<i>Event</i>			
Yes	17 (14.7%)	40 (34.2%)	<0.001
No	99 (85.3%)	77 (65.8%)	
<i>Gender</i>			
Female	61 (52.6%)	65 (55.6%)	0.746
Male	55 (47.4%)	52 (44.4%)	
<i>T</i>			
T1	53 (45.7%)	33 (28.2%)	0.0246
T2	52 (44.8%)	71 (60.7%)	
T3	8 (6.9%)	12 (10.3%)	
T4	3 (2.6%)	1 (0.9%)	
<i>N</i>			
N0	85 (73.3%)	56 (47.9%)	<0.001
N1	16 (13.8%)	32 (27.4%)	
N2	15 (12.9%)	29 (24.8%)	
<i>Tumor stage</i>			
Stage I	75 (64.7%)	48 (41.0%)	0.0022
Stage II	22 (19.0%)	35 (29.9%)	
Stage III	15 (12.9%)	31 (26.5%)	
Stage IV	4 (3.4%)	3 (2.6%)	
<i>Smoking</i>			
≤40 pack years of smoke	71 (61.2%)	75 (64.1%)	0.748
>40 pack years of smoke	45 (38.8%)	42 (35.9%)	

TABLE 4: Clinical information of LUAD patients in the testing set.

	Low risk (n = 112)	High risk (n = 115)	p value
<i>Age</i>			
<65	41 (36.6%)	53 (46.1%)	0.189
>65	71 (63.4%)	62 (53.9%)	
<i>Event</i>			
Yes	22 (19.6%)	34 (29.6%)	0.114
No	90 (80.4%)	81 (70.4%)	
<i>Gender</i>			
Female	62 (55.4%)	59 (51.3%)	0.632
Male	50 (44.6%)	56 (48.7%)	
<i>T</i>			
T1	42 (37.5%)	30 (26.1%)	0.00148
T2	62 (55.4%)	57 (49.6%)	
T3	3 (2.7%)	20 (17.4%)	
T4	5 (4.5%)	8 (7.0%)	
<i>N</i>			
N0	94 (83.9%)	71 (61.7%)	<0.001
N1	13 (11.6%)	23 (20.0%)	
N2	5 (4.5%)	21 (18.3%)	
<i>Tumor stage</i>			
Stage I	77 (68.8%)	51 (44.3%)	0.00205
Stage II	20 (17.9%)	32 (27.8%)	
Stage III	9 (8.0%)	23 (20.0%)	
Stage IV	6 (5.4%)	9 (7.8%)	
<i>Smoking</i>			
≤40 pack years of smoke	75 (67.0%)	61 (53.0%)	0.045
>40 pack years of smoke	37 (33.0%)	54 (47.0%)	

useful. In addition, the treatment and prognosis of patients are mainly affected by distant metastasis and relapse. Thus, it is highly important to build a predictive model characterized by high stability and validity for better early diagnosis, medication guidance, and prognosis prediction. At present, many studies have focused on the construction of prognostic models for LUAD treatment. For instance, Li et al. suggested that clinical immune characteristics were a promising biomarker that could be used to evaluate OS of nonsquamous NSCLC patients (including early disease) [11]. Park et al. tried to construct a gene signature-based prognostic model for LUAD [12], and in 2016, Shukla et al. proposed the first RNA-seq-based prognostic signature through analyzing the RNA-seq and clinical data, making an attempt to develop a potent predictive tool for LUAD prognosis [13]. Despite the extensive research on signature genes used for LUAD prognosis, models with robust prediction capability have yet to be successfully constructed. Besides, with the development of high-throughput sequencing, more gene datasets of LUAD should be employed into new studies.

In our study, seven LUAD survival-related genes were identified, including NTSR1, RHOV, KLK8, TNS4, C1QTNF6, IVL, and B4GALNT2. These 7 signature genes

were obtained from the HTSeq-Counts in the TCGA-LUAD dataset using univariate COX, Lasso regression, and multivariate COX analyses. Sequentially, the risk score based on the 7-gene signature was established and formulated as $\text{risk_score} = 0.000245 * \text{NTSR1} + (7.13E - 05) * \text{RHOV} + 0.000505 * \text{KLK8} + (7.01E - 05) * \text{TNS4} + 0.000288 * \text{C1QTNF6} + 0.00044 * \text{IVL} + 0.000161 * \text{B4GALNT2}$. As reported, most of these 7 genes are closely related to cancer progression. For example, NTSR1 (Neurotensin Receptor 1) has been reported as a potential prognostic biomarker for surgically resected stage I LUAD [14] and prostate cancer [15]. RHOV (Ras Homolog Family Member V) has been verified to be highly expressed in NSCLC and can serve as a signature gene in LUAD prognosis [16]. KLK8 (Kallikrein Related Peptidase 8) has presented its research value in the prognosis of various cancers, such as lung cancer [17], ovarian cancer [18], breast cancer [19], colon cancer, and rectal cancer [20]. Moreover, TNS4 (Tensin 4) has been found to be upregulated in LUAD and able to predict poor prognosis, and it has been observed to be mediated by miR-150-3p [21]. Meanwhile, another study indicated that the aberrant methylation of TNS4 is significantly associated with the OS of LUAD patients [22]. C1QTNF6 (C1q/tumor necrosis

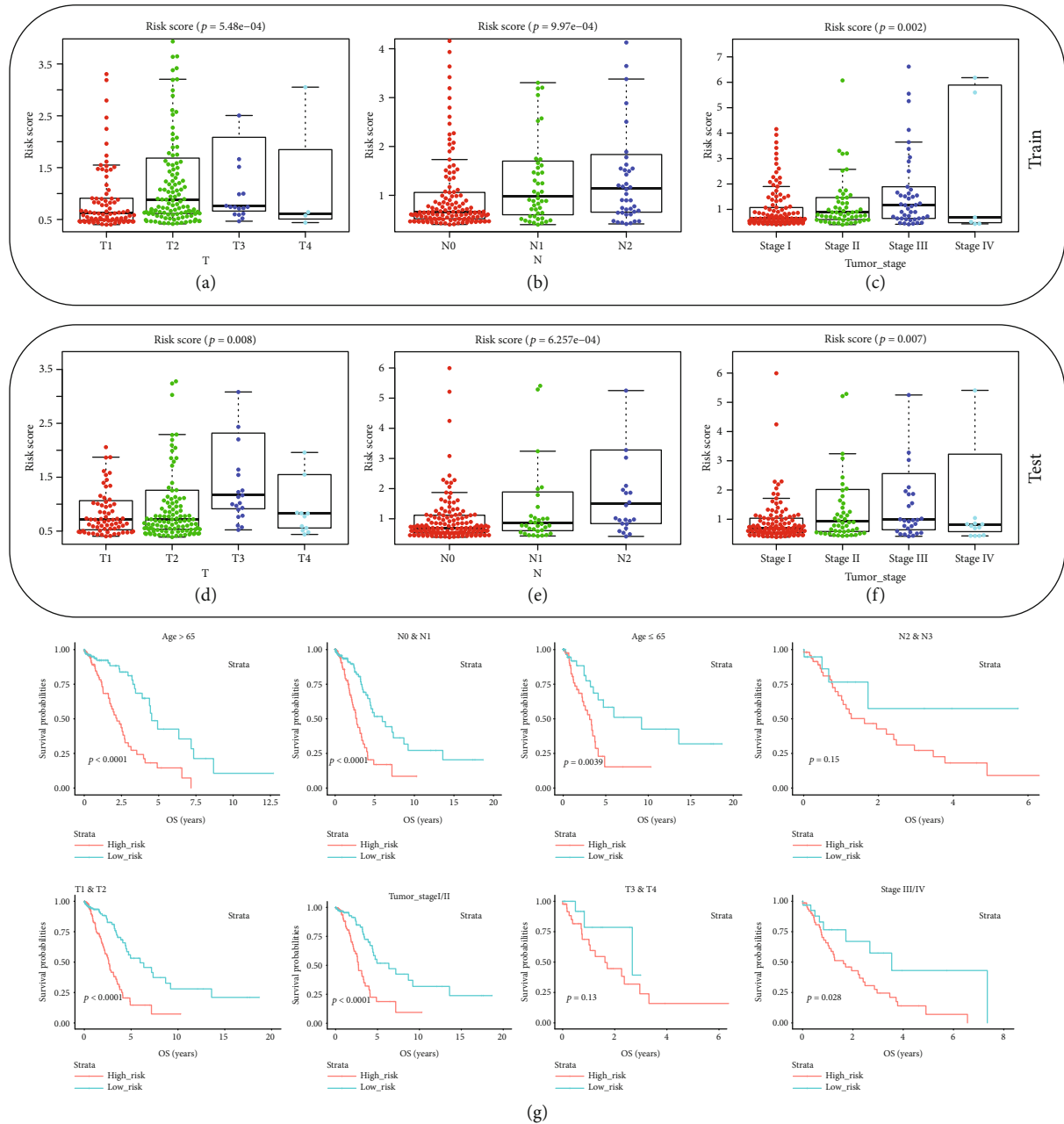


FIGURE 4: Prognostic impact of the model in clinical characteristics. The correlation of the 7-gene signature-based risk score with the TNM stage was analyzed in the (a–c) training set and (d–f) testing set. (g) OS curves were plotted for validation of the predictive ability of the 7-gene signature-based risk score in patients with different age and TNM stage.

factor-related protein 6), a member of the CTRP family, has shown its potential as an independent predictor for the prognosis of LUAD sufferers [23]. Additionally, although the role of B4GALNT2 (Beta-1,4-N-Acetyl-Galactosaminyltransferase 2) in LUAD has not been investigated, it has been observed to be highly related to gastric cancer metastasis [24]. However, the association between IVL (Involucrin) and the progression of LUAD has not been reported, which requires further study in the future. In view of the above studies, we could conclude that some of these signature genes exhibit a certain relationship with the prognosis of other cancers.

During the research, each patient in the training set and testing set was conferred a risk score and classified into the high-risk group and the low-risk group according to the median score. As suggested in OS curves, patients in the high-risk group had poorer survival. ROC curves were plotted, and the AUC values of 1-, 3-, and 5-year survival in two sets were all above 0.6, indicating that the 7-gene signature-based risk score model was capable of predicting LUAD prognosis. Notably, similar results were found in an independent dataset GSE26939 from the GEO database, demonstrating the validity and practicality of this 7-gene model. Furthermore, the association between this model

and clinical characteristics of LUAD patients was explored, finding that the model functioned well in predicting the prognosis of patients with different age (age > 65, age < 65) and TNM stage (N0&N1, T1&T2, and tumor stage I/II), while the effect in the GSE26939 was less remarkable.

5. Conclusions

In conclusion, we obtained 1,655 DEGs from the TCGA-LUAD dataset using the “edgeR” package and constructed a prognostic 7-gene signature-based model (containing NTSR1, RHOV, KLK8, TNS4, C1QTNF6, IVL, and B4GALNT2, seven genes) through univariate COX, Lasso, and multivariate COX regression analyses. The robust model we built helps to advance the clinical research on LUAD and better understand the mechanism underlying LUAD occurrence and progression.

Data Availability

All the data in my manuscript is available.

Conflicts of Interest

The authors declare no conflicts of interest.

Authors' Contributions

Yingqing Zhang contributed to the study design. Ming Zhang conducted the literature search. Xixi Gao required the data. Xiaoping Zhang and Xiaodong Lv wrote the article. Jialiang Liu performed data analysis and drafted. Yufen Xu and Zhixian Fang revised the article. Wenyu Chen gave the final approval of the version to be submitted. Yingqing Zhang, Xiaoping Zhang, and Xiaodong Lv are co-first authors and contributed equally to this article.

Acknowledgments

The authors are grateful to the funds from the Natural Science Foundation of Zhejiang Province (No. LQ20H160057), the Key Discipline of Jiaxing Respiratory Medicine Construction Project (No.2019-zc-04), the Science and technology project of Jiaxing (2019AY32030, 2019AD32126, and 2020AY30012), the Jiaxing Key Laboratory of Precision Treatment for Lung Cancer, the Early Diagnosis and Comprehensive Treatment of Lung Cancer Innovation Team Building Project, and the Clinical Research Project of Microwave Ablation Combined with Chemotherapy in The Treatment of Stage IIIb-IV Peripheral Non-small Cell Lung Cancer (2018AD32087). The authors gratefully acknowledge contributions from the GEO network, the TCGA network, and the GDC network. The authors appreciate the patients who have participated in TCGA, GDC, and GEO. The authors sincerely thank the researchers for providing their GEO databases, TCGA databases, and GDC information online; it is our pleasure to acknowledge their contributions.

Supplementary Materials

Supplementary Table 1: survival-related genes. Supplementary Table 2: matched clinical info of patients in independent dataset GSE26939. (*Supplementary Materials*)

References

- [1] R. L. Siegel, K. D. Miller, and A. Jemal, “Cancer statistics, 2019,” *CA: a Cancer Journal for Clinicians*, vol. 69, no. 1, pp. 7–34, 2018.
- [2] J. Shi, X. Hua, B. Zhu et al., “Somatic genomics and clinical features of lung adenocarcinoma: a retrospective study,” *PLoS Medicine*, vol. 13, no. 12, article e1002162, 2016.
- [3] A. Carrato, A. Vergnenègre, M. Thomas, K. McBride, J. Medina, and G. Cruciani, “Clinical management patterns and treatment outcomes in patients with non-small cell lung cancer (NSCLC) across Europe: EPICLIN-Lung study,” *Current Medical Research and Opinion*, vol. 30, no. 3, pp. 447–461, 2014.
- [4] K. D. Miller, R. L. Siegel, C. C. Lin et al., “Cancer treatment and survivorship statistics, 2016,” *CA: a Cancer Journal for Clinicians*, vol. 66, no. 4, pp. 271–289, 2016.
- [5] M. C. Chamberlain, C. S. Baik, V. K. Gadi, S. Bhatia, and L. Q. M. Chow, “Systemic therapy of brain metastases: non-small cell lung cancer, breast cancer, and melanoma,” *Neuro-Oncology*, vol. 19, no. 1, pp. i1–i24, 2016.
- [6] J. A. Roth, B. H. Goulart, A. Ravelo, H. Kolkey, and S. D. Ramsey, “Survival gains from first-line systemic therapy in metastatic non-small cell lung cancer in the U.S., 1990–2015: progress and opportunities,” *The Oncologist*, vol. 22, no. 3, pp. 304–310, 2017.
- [7] H. Wang, R. Gu, F. Tian et al., “PHLPP2 as a novel metastatic and prognostic biomarker in non-small cell lung cancer patients,” *Thoracic Cancer*, vol. 10, no. 11, pp. 2124–2132, 2019.
- [8] J. Y. Park, S. H. Jang, H. I. Kim et al., “Thyroid transcription factor-1 as a prognostic indicator for stage IV lung adenocarcinoma with and without EGFR-sensitizing mutations,” *BMC Cancer*, vol. 19, no. 1, p. 574, 2019.
- [9] J. Liu, Y. Li, X. Zhu et al., “Increased CX3CL1 mRNA expression level is a positive prognostic factor in patients with lung adenocarcinoma,” *Oncology Letters*, vol. 17, pp. 4877–4890, 2019.
- [10] M. M. Vasquez, C. Hu, D. J. Roe, Z. Chen, M. Halonen, and S. Guerra, “Least absolute shrinkage and selection operator type methods for the identification of serum biomarkers of overweight and obesity: simulation and application,” *BMC Medical Research Methodology*, vol. 16, no. 1, p. 154, 2016.
- [11] B. Li, Y. Cui, M. Diehn, and R. Li, “Development and validation of an individualized immune prognostic signature in early-stage nonsquamous non-small cell lung cancer,” *JAMA Oncology*, vol. 3, no. 11, pp. 1529–1537, 2017.
- [12] Y. Y. Park, E. S. Park, S. B. Kim et al., “Development and validation of a prognostic gene-expression signature for lung adenocarcinoma,” *PLoS One*, vol. 7, no. 9, article e44225, 2012.
- [13] S. Shukla, J. R. Evans, R. Malik et al., “Development of a RNA-Seq based prognostic signature in lung adenocarcinoma,” *Journal of the National Cancer Institute*, vol. 109, no. 1, p. djw200, 2016.

- [14] M. Alifano, F. Souaze, S. Dupouy et al., "Neurotensin receptor 1 determines the outcome of non-small cell lung cancer," *Clinical Cancer Research*, vol. 16, no. 17, pp. 4401–4410, 2010.
- [15] J. Zhong, X. Ren, Z. Chen et al., "miR-21-5p promotes lung adenocarcinoma progression partially through targeting SET/TAF-I α ," *Life Sciences*, vol. 231, p. 116539, 2019.
- [16] M. V. Shepelev and I. V. Korobko, "The *_RHOV_* gene is over-expressed in human non-small cell lung cancer," *Cancer Genetics*, vol. 206, no. 11, pp. 393–397, 2013.
- [17] C. Planque, Y. H. Choi, S. Guyetant, N. Heuzé-Vourc'h, L. Briollais, and Y. Courty, "Alternative splicing variant of kallikrein-related peptidase 8 as an independent predictor of unfavorable prognosis in lung cancer," *Clinical Chemistry*, vol. 56, no. 6, pp. 987–997, 2010.
- [18] A. Magklara, A. Scorilas, D. Katsaros et al., "The human *KLK8* (neuropsin/ovasin) gene: identification of two novel splice variants and its prognostic value in ovarian cancer," *Clinical Cancer Research*, vol. 7, no. 4, pp. 806–811, 2001.
- [19] K. Michaelidou, A. Ardavanis, and A. Scorilas, "Clinical relevance of the deregulated kallikrein-related peptidase 8 mRNA expression in breast cancer: a novel independent indicator of disease-free survival," *Breast Cancer Research and Treatment*, vol. 152, no. 2, pp. 323–336, 2015.
- [20] X. Liu, B. Quan, Z. Tian et al., "Elevated expression of *KLK8* predicts poor prognosis in colorectal cancer," *Biomedicine & Pharmacotherapy*, vol. 88, pp. 595–602, 2017.
- [21] S. Misono, N. Seki, K. Mizuno et al., "Molecular pathogenesis of gene regulation by the miR-150 duplex: miR-150-3p regulates *TNS4* in lung adenocarcinoma," *Cancers*, vol. 11, no. 5, p. 601, 2019.
- [22] W. He, D. Ju, Z. Jie, A. Zhang, X. Xing, and Q. Yang, "Aberrant CpG-methylation affects genes expression predicting survival in lung adenocarcinoma," *Cancer Medicine*, vol. 7, no. 11, pp. 5716–5726, 2018.
- [23] M. Han, B. Wang, M. Zhu, and Y. Zhang, "C1QTNF6 as a novel biomarker regulates cellular behaviors in A549 cells and exacerbates the outcome of lung adenocarcinoma patients," *In Vitro Cellular & Developmental Biology Animal*, vol. 55, no. 8, pp. 614–621, 2019.
- [24] Y. I. Kawamura, Y. Adachi, D. T. Curiel et al., "Therapeutic adenoviral gene transfer of a glycosyltransferase for prevention of peritoneal dissemination and metastasis of gastric cancer," *Cancer Gene Therapy*, vol. 21, no. 10, pp. 427–433, 2014.