Original Research

Identification of a 7-miRNA signature for predicting the prognosis of patients with lung adenocarcinoma

Ruijun Liu, Zhiyi Guo, Jia Huang, Jiantao Li, Qiang Tan and Qingquan Luo

Lung Tumor Clinical Medicine Center, Shanghai Chest Hospital, Shanghai Jiao tong University, Shanghai 200030, China
Corresponding author: Ruijun Liu. Email: lrj_harvest@163.com

Impact Statement

Lung cancer is one of the deadliest cancers worldwide. Adenocarcinoma has been recognized as the most common kind of lung cancer, and patients with this kind of lung malignancy often develop a poor prognosis. Effective indicators are required to monitor the prognosis of patients with lung adenocarcinoma (LUAD). MicroRNAs have been used as biomarkers due to their instrumental function in the progression and metastasis of tumors. Our research developed a prognosis risk system based on seven microRNAs, which effectively and independently predicts LUAD patient prognosis. The findings of our work may provide a reference for the evaluation of the clinical prognosis of patients with LUAD.

Abstract

The role of microRNAs (miRNAs) in tumor diagnosis and patients' prognosis has recently gained extensive research attention. This study was designed to analyze miRNA in lung adenocarcinoma (LUAD) using bioinformatics analysis and to identify novel biomarkers to predict overall survival (OS) for LUAD patients. Differential miRNA expression analysis was performed on LUAD, and normal tissues were extracted from The Cancer Genome Atlas (TCGA). Univariate Cox risk regression and least absolute shrinkage and selection operator (LASSO) Cox analysis were used to screen prognostic miRNAs and develop a risk score model. The prognostic performance of the system was examined utilizing the Kaplan–Meier and receiver operating characteristic (ROC) curves. Independent prognostic factors of LUAD were determined by multivariate Cox regression analysis. Nomogram was constructed according to the independent prognostic factors to evaluate the patients' one-, three- and five-year OS. A 7-miRNA signature based on miR-584-5p, miR-31-3p, miR-490-3, miR-4661-5p, miR-30e-5p, miR-582-5p, and miR-148a-3p was established. To categorize patients into high- and low-risk groups, the risk score was computed. The OS of the low-risk group was significantly longer than the high-risk group, and the signature showed high sensitivity and specificity in anticipating the one-, three- and five-year OS. The system was an independent factor in predicting the OS of LUAD patients and performed better when combined with the N stage in nomogram. A 7-miRNA signature developed in this study could accurately predict LUAD survival.

Keywords: MiRNA, lung adenocarcinoma, prognosis, risk score

***Experimental Biology and Medicine* 2022; 247: 641–657. DOI: 10.1177/15353702211067450**

Introduction

Lung cancer is the second most frequently diagnosed cancer in 2020, accounting for 11.4% of all newly diagnosed cancer cases worldwide. At the same time, lung cancer, which is also a leading cause of cancer death, makes up 18.0% of all cancer-related deaths in 2020. The morbidity and mortality of male patients are two times higher than those of women.[1] Lung adenocarcinoma (LUAD) is a major lung cancer histological subtype, with a more invasive nature and more morphological heterogeneity than other types of lung cancer.[2] The incidence of LUAD is experiencing a steady increase in current smokers, ex-smokers, and even non-smokers, and the prognosis of a majority of patients is poor, with a five-year survival probability of just 15%.[3] Therefore, effective prognostic strategies are needed to enhance the LUAD patients' survival probabilities.

Tumor, node, metastasis (TNM) staging system is a common component employed to evaluate the LUAD patients' prognosis.[4] However, the TNM staging system is limited to anatomy rather than the biological behaviors of the disease.[5,6] Recent studies demonstrated that the identification of molecular biomarkers can help prognostic evaluation of LUAD patients.[7] At present, a large number of prognostic markers, such as CD8,[8] G6PD,[9] and HSP90 β,[10] have been identified in LUAD. The improvement of the study of these coding genes as prognostic indicators has driven increasing research attention to the function performed by non-coding RNAs in tumor diagnosis and prognosis. Wang *et al.*[11] identified four prognostic markers of lncRNA, namely, LINC00578, TDRKH-AS1, RBPMS-AS1, and RP11470 M17.2, in patients with LUAD. Li *et al.*[12] established a seven immune-related lncRNA model that might effectively anticipate the LUAD patients' prognosis. In addition, lncRNA GCC2AS1 is considered to be a diagnostic and prognostic indicator for LUAD.[13] It is reported that microRNAs (miRNAs) have a stable and robust expression pattern in clinical samples,

pointing to its potential of acting as a prognostic indicator. At present, many miRNA signatures have been proposed, and their uses are increasingly studied in clinical trials.[14] In terms of the prognosis of LUAD, most studies only focused on the identification of a single miRNA marker, which does not necessarily produce reliable predictive results.[15] There is still a lack of research on multi-miRNA markers in the prognostic prediction of LUAD.

In this research, we thoroughly examined the clinical information and miRNA expression data of patients with LUAD obtained from The Cancer Genome Atlas (TCGA) database, developed a prognostic model based on miRNAs, and verified it in the test set, validation set, and the entire dataset. It is possible that our model might serve as a novel standard for prognostic assessment of LUAD patients.

Materials and methods

LUAD dataset preparation and patient grouping

The transcriptome gene expression profiles along with the clinical data of LUAD patients were acquired from TCGA (https://portal.gdc.cancer.gov/repository), and 500 LUAD samples and 46 normal samples were obtained (Table 1), which were categorized into training and validation sets. To minimize random allocation bias interfering with the stability of subsequent modeling, all data were randomly placed again into the groups for a total of 100 times before subsequent analyses. In this study, group sampling was conducted following a ratio of 1:1 for the training set to the validation set. Based on this ratio, the 500 LUAD samples were classified into a training set ($n = 250$) and a verification set ($n = 250$). After conducting a chi-square test, it was confirmed that there existed no significant differences in smoking history, TNM stage, sex, clinical stage, age, overall survival (OS), radiotherapy, and chemotherapy between the two groups (Table 2). We also downloaded miRNA expression profiles and survival data of 32 samples in the GSE63805 cohort from Gene Expression Omnibus (GEO). In addition, sample information of serum datasets GSE137140, GSE31568, GSE68951, and GSE40738 was obtained from GEO.

Table 1. Clinicopathological information of LUAD samples.

Clinical features	TCGA-LUAD
Type	
Normal	46
Tumor	500
OS	
0	319
1	181
T stage	
T1	169
T2	265
T3	45
T4	18
TX	3
N stage	
N0	324
N1	94
N2	69
N3	2
NX	11
M stage	
M0	334
M1	22
MX	144
Stage	
I	271
II	119
III	80
IV	23
X	7
Smoking	
1	72
2	119
3	129
4	162
5	4
7	14
Gender	
Male	232
Female	268
Age (years)	
⩽65	236
>65	254
Unknown	10
Chemotherapy	
Yes	174
No	326
Radiotherapy	
Yes	57
No	361
Unknown	82

LUAD: lung adenocarcinoma; TCGA: The Cancer Genome Atlas; OS: overall survival.

Identification of miRNA signature related to LUAD prognosis

Limma package[16] was used for analyzing differentially expressed miRNAs. The screening criteria were error detection rate (FDR) < 0.05, $|$ multiple change (FC) $| \geqslant 2$. Then, univariate Cox regression analysis was conducted in the training database to thoroughly screen the miRNA strongly associated with the patients' OS. Then, least absolute shrinkage and selection operator (LASSO) analysis in the R software glmnet[17] was performed to narrow the number of miRNAs, and the miRNA variables were streamlined according to Akaike information criteria (AIC).

Development and validation of the risk score model

The risk score model was constructed according to the miRNAs screened by Cox and LASSO. The performance of the risk scoring model was verified in the test set, the internal and external validation sets, and the entire dataset. The risk score of each LUAD patient was computed using the model and subsequently standardized. LUAD patients were categorized into two groups, namely the high- and low-risk groups with 0 as the threshold. The Kaplan–Meier survival analysis was used to assess the differences in survival between the two groups. The receiver operating characteristic (ROC) curves of one, three, and five years were generated, and the

Table 2. Sample information of training set and verification set.

Clinical features	TCGA train	TCGA test	*P*
OS			
0	166	153	0.2641
1	84	97	
T stage			
T1	80	89	0.6702
T2	138	127	
T3	24	21	
T4	7	11	
TX	1	2	
N stage			
N0	177	147	0.04989
N1	39	55	
N2	30	39	
N3	0	2	
NX	4	7	
M stage			
M0	172	162	0.5282
M1	9	13	
MX	69	75	
Stage			
I	147	124	0.1465
II	55	64	
III	34	46	
IV	9	14	
X	5	2	
Smoking			
1	31	41	0.493
2	61	58	
3	59	70	
4	90	72	
5	2	2	
7	7	7	
Gender			
Male	114	118	0.7879
Female	136	132	
Age (years)			
⩽65	119	117	1
>65	128	126	
Unknown	3	7	
Chemotherapy			
Yes	80	94	0.2223
No	170	156	
Radiotherapy			
Yes	25	32	0.5086
No	179	182	
Unknown	46	36	

LUAD: lung adenocarcinoma; TCGA: The Cancer Genome Atlas; OS: overall survival.

accuracy of survival prediction was examined by computing the area under the curve (AUC). We also assessed whether the model was independent of other clinicopathological parameters by proportional hazard regression of univariate and multivariate Cox, in which clinic stage, N stage, T stage, gender, M stage, age, and risk type were used as covariates.

Immune microenvironment analysis for risk groups

Immune score, stromal score, and ESTIMATE score were evaluated by applying the ESTIMATE algorithm to gene expression data from TCGA-LUAD. The microenvironment cell populations-counter (MCP-counter) in R package was utilized to estimate the abundance of 10 infiltrating immune cells in patients with different risks. Moreover, the enrichment scores of the transforming growth factor (TGF-β), epithelial–mesenchymal transition (EMT), and angiogenesis hallmark gene sets were also computed utilizing the single-sample gene set enrichment analysis (ssGSEA) method. In addition, ssGSEA was carried out using the R software "GSVA" to evaluate immune cell enrichment scores between low- and high-risk populations.

Establishment of a prognostic nomogram

The survival rate of patients was predicted by establishing a nomogram, which took into account the clinical characteristics of the N stage and risk score. Then, calibration curves were drawn to predict one-, three-, and five-year survival probability to evaluate the uniformity between the actual and the anticipated survival rates of the nomogram. Finally, the decision curve analysis (DCA)[18] was undertaken to verify the clinical practicability of the nomogram.

Functional enrichment analysis and target prediction

Target analysis of miRNA in the model was done through TargetScan (http://www.targetscan.org/vert_71/), miRDB (http://mirdb.org/index.html), and MiRTarBase (http://mirtarbase.cuhk.edu.cn/php/index.php). Then, through using the R software package WebGestalt,[19] the miRNA target genes were analyzed by Gene Ontology (GO) functional enrichment analysis and Kyoto Encyclopedia of Genes and Genomics (KEGG) pathway analysis.

Statistical analysis

The R software (http://www.Rproject.org) was employed to conduct all the statistical analyses. An unpaired *t*-test was utilized to screen miRNAs with differential expression. OS between low- and high-risk patients was compared with the aid of the Kaplan–Meier survival analysis. Hence, *P*-value less than 0.05 was considered to be statistically significant.

Results

Identification of differentially expressed miRNAs

Our research route was shown in Figure 1. As per the threshold, 295 differentially expressed miRNAs (including 231 differentially upregulated miRNAs and 64 differentially downregulated miRNA) were identified between tumor samples and normal samples (Figure 2(a) and (b)).

Development and verification of the 7-miRNA prognostic model

Univariate Cox regression analysis was employed to analyze the miRNAs associated with OS in the training set. The results showed that there were 10 miRNAs in the range of $P < 0.01$ (Supplementary Table S1). LASSO Cox regression reduced the number of 10 prognostic miRNAs

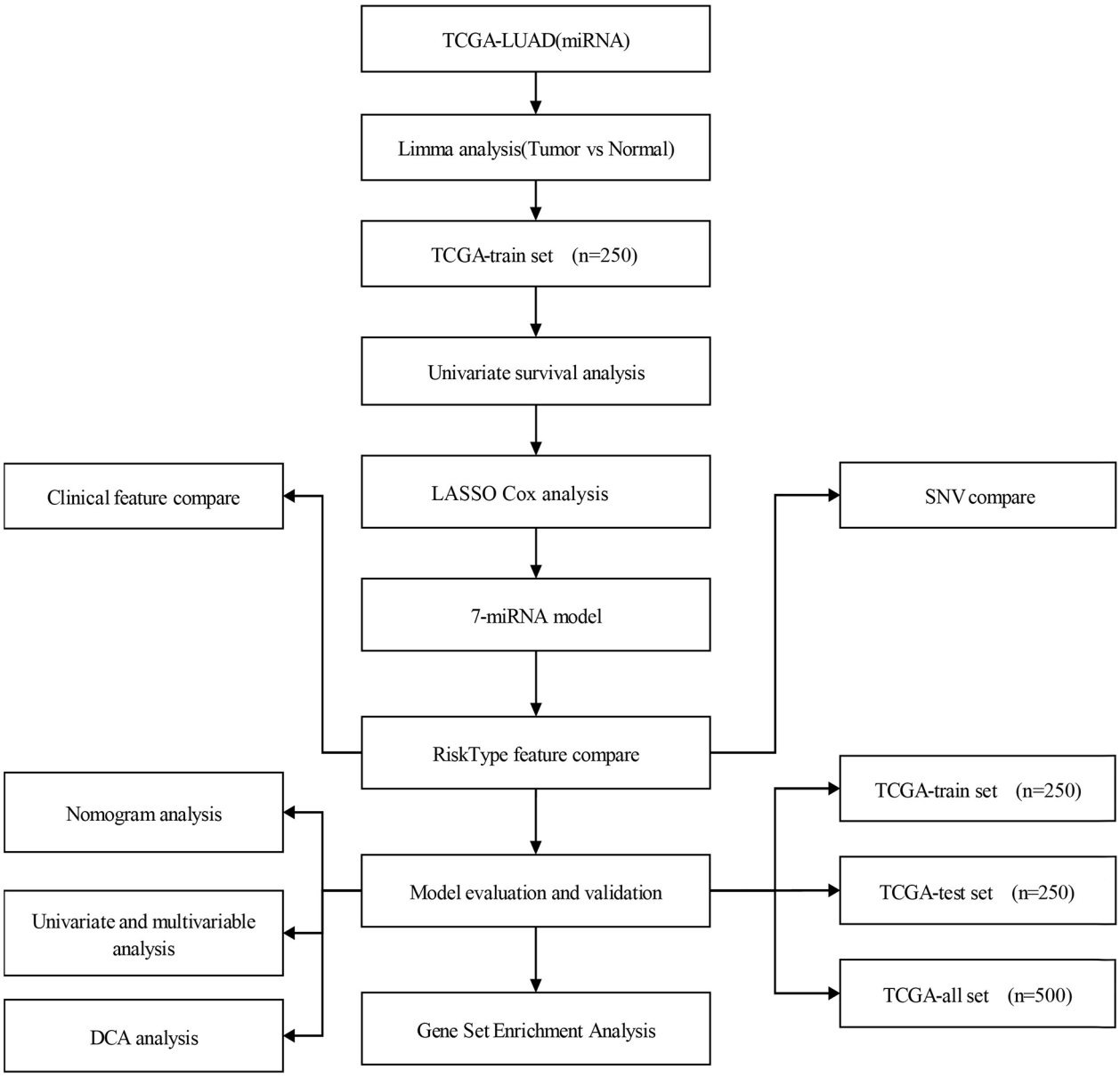

Figure 1. A roadmap for this study.

down to 8 and calculated the regression coefficient of each miRNA (Figure 3(a) and (b)). According to AIC, seven miRNAs, namely miR-584-5p, miR-582-5p, miR-4661-5p, miR-30e-5p, miR-490-3p, miR-31-3p, and miR-148a-3p, were determined as prognostic miRNAs for LUAD. Next, according to the regression coefficient and the candidate miRNAs expression (weighted by the relative coefficient acquired using multivariate Cox regression), the prognostic signature of LUAD was constructed to compute each patient's risk score in the training set. After standardization of the risk score, the patients were categorized into a low-risk group ($n = 132$) and a high-risk group ($n = 118$) with zero as the segmentation point. We discovered that patients having high-risk scores tended to express high levels of the risk miRNAs, while those with low-risk scores expressed higher levels of protective miRNAs (Figure 4(a)). The results of the survival study revealed that the OS of patients with high risks was significantly poorer as opposed to that of patients with low risks (Figure 4(b)). The one-, three-, and five-year ROC curves of OS demonstrated that AUC were all greater than 0.7 (Figure 4(c)), indicating that the system based on 7-miRNA signature had a high specificity and sensitivity.

We further verified the performance of the signature in the validation set and the entire dataset to determine its robustness. The patients were categorized into low- and high-risk groups following the same steps as the training set in the internal and external validation sets and the entire dataset (Figure 4(d) and (g)). The findings of the Kaplan–Meier analysis illustrated that the patients' survival outcome in the high-risk group was poorer as opposed to that in the low-risk group for the internal validation set and entire TCGA-LUAD cohort (Figure 4(e) and (h)). Time-dependent ROC curve analysis revealed that

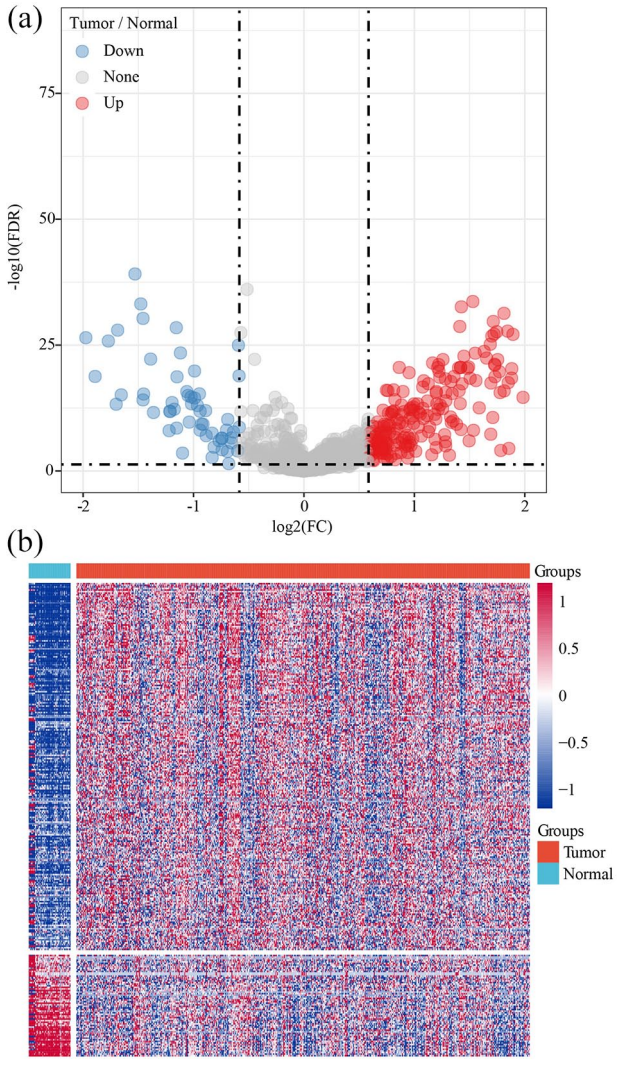

Figure 2. Differential expression miRNA analysis. (a) Volcano map of different miRNA, where each dot represents a miRNA; the red dot represents a differential upregulated miRNA and the blue dot represents a differential downmodulated miRNA. (b) Expression heat map of 295 differentially expressed miRNAs. (A color version of this figure is available in the online journal.)

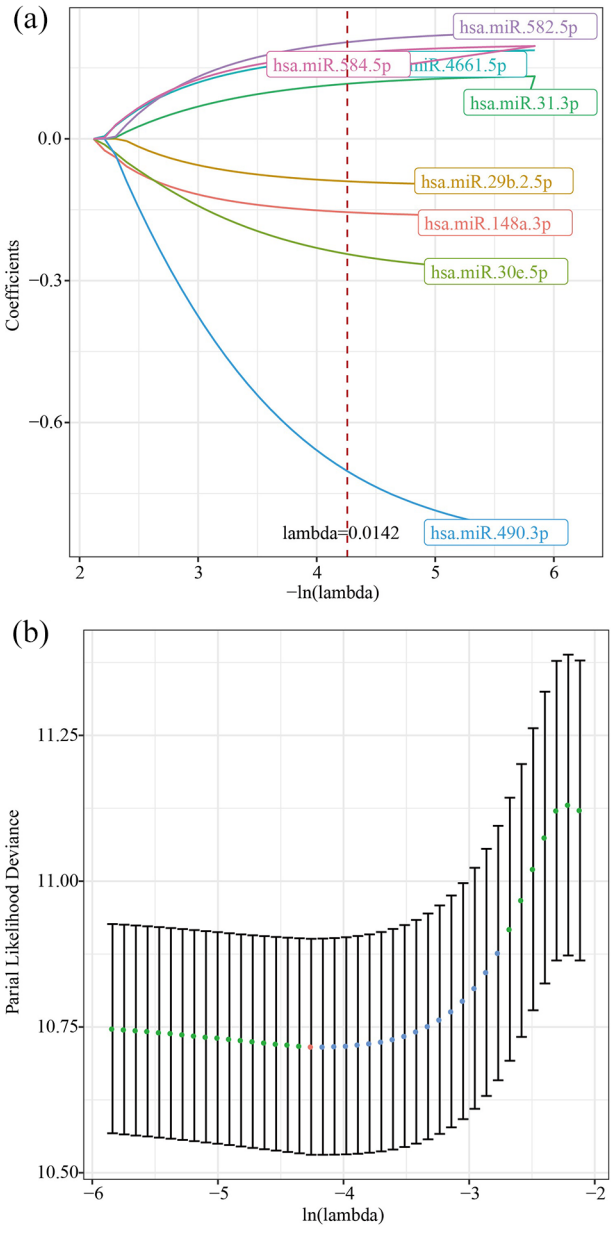

Figure 3. The optimal prognostic miRNA was screened by LASSO regression analysis. (a) For the changing trajectory of each independent variable, the horizontal axis shows the independent variable λ log value, whereas the vertical axis indicates the independent variable's coefficient. (b) The confidence interval under each λ 5x cross-validation to adjust the parameter selection in the LASSO model. (A color version of this figure is available in the online journal.)

the 7-miRNA score model performed well in anticipating LUAD patients' one-, three-, and five-year survival (Figure 4(f) and (i)). There were only four miRNAs of the risk model in the GSE63805 cohort, and the risk score of each LUAD sample in this cohort was computed according to the expression levels of the four miRNAs and risk regression coefficients. For each LUAD sample, a higher risk score indicated a poor prognosis (Figure S1A). The AUC for three- and five-year OS was 0.82 and 0.73, respectively (Figure S1B).

The 7-miRNA signature was related to the pathological features of LUAD

We determined whether the 7-miRNA signature was associated with the clinical characteristics of patients with LUAD. In the stratified patients, there were significant differences in risk scores among different TNM stages, tumor stages, and gender; moreover, higher scores were found in high TNM stage, high clinical stage, radiotherapy, and male patients (Figure 5). Our data also revealed that there was a significant correlation between the 7-miRNA signature and clinicopathological features.

Different clinical characteristics and immune microenvironment features between low- and high-risk groups

To identify the differences in clinical performance between low- and high-risk groups, the samples in the risk groups were evaluated by analysis of variance. The results demonstrated that there were significant differences in male/female

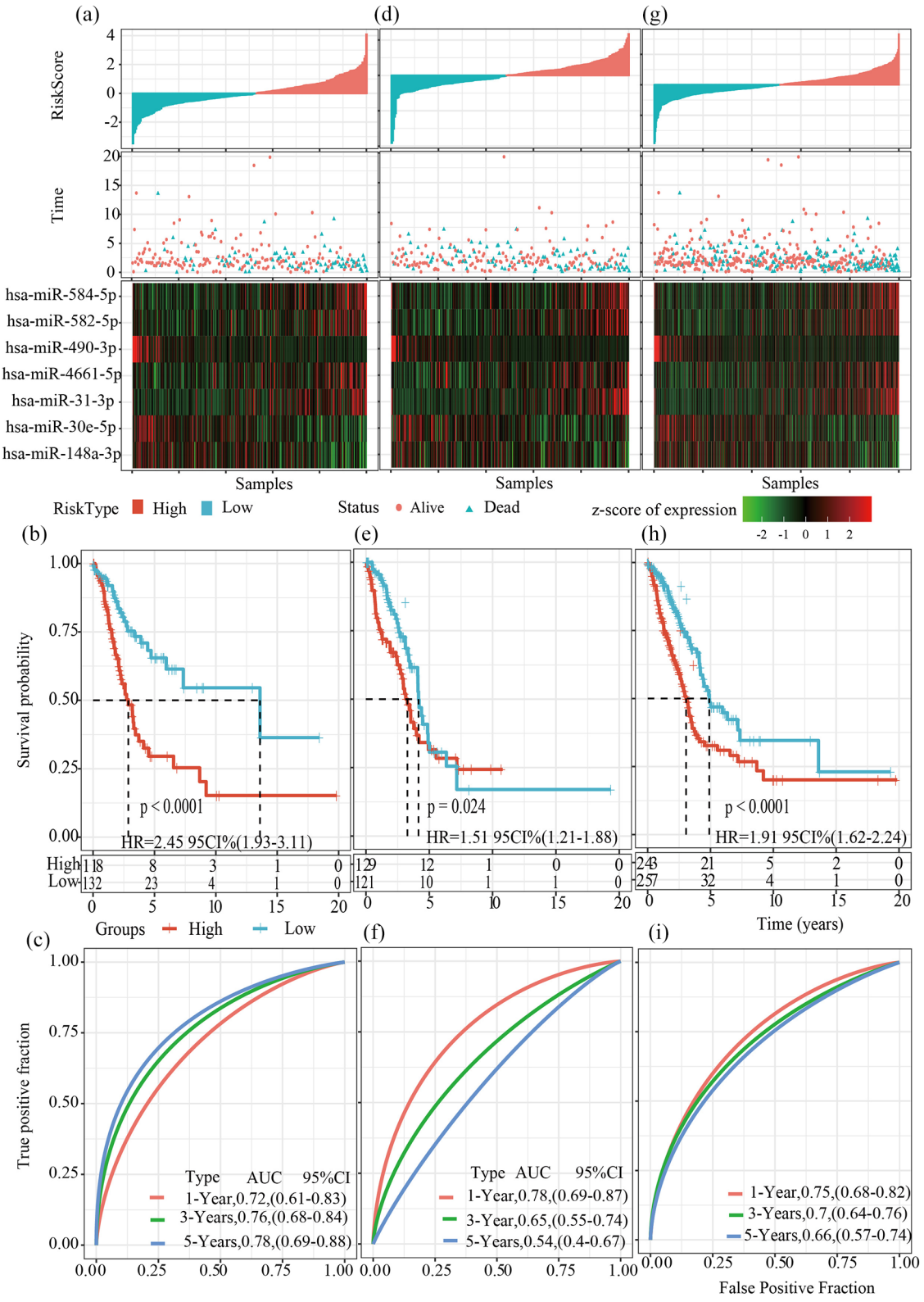

Figure 4. Validation of the 7-miRNA signature in the LUAD training set, validation set, and the entire set. (a), (d), (e) Distribution of risk scores, levels of 7-miRNA expression, and survival of LUAD patients in the training set, validation set, and combination set. (b), (e), (h) Kaplan–Meier survival analysis based on TCGA-LUAD training set, validation set, and whole dataset. (c), (f), (i) The time-dependent ROC curve of a risk score for patient survival in predictive training, validation, and combinatorial set. (A color version of this figure is available in the online journal.)

ratio, survival ratio, TNM stage, clinical stage, and chemotherapy ratio between low- and high-risk groups. There were more male patients in the high-risk group, with a greater proportion of chemotherapy and mortality. The probability of patients in Stages T2–T4, N2, and M1 was also considerably greater as opposed to that in the low-risk group. In

Figure 5. The relationship between the 7-miRNA signature and pathological features. (a) to (i) Distribution of risk scores according to M stage, N stage, gender, clinic stage, T stage, smoking, age, chemotherapy, and radiotherapy. (A color version of this figure is available in the online journal.)

addition, the proportion of Stages III and IV high-risk group patients was noticeably higher in contrast with that in the low-risk group (Figure 6). The difference in the immune cell composition in the tumor microenvironment of each risk score group was displayed in the form of a violin plot. MCP-counter analysis showed significant differences in the scores of 5 of the 10 infiltrating immune cells between various risk groups. The scores of natural killer (NK) cells and fibroblasts

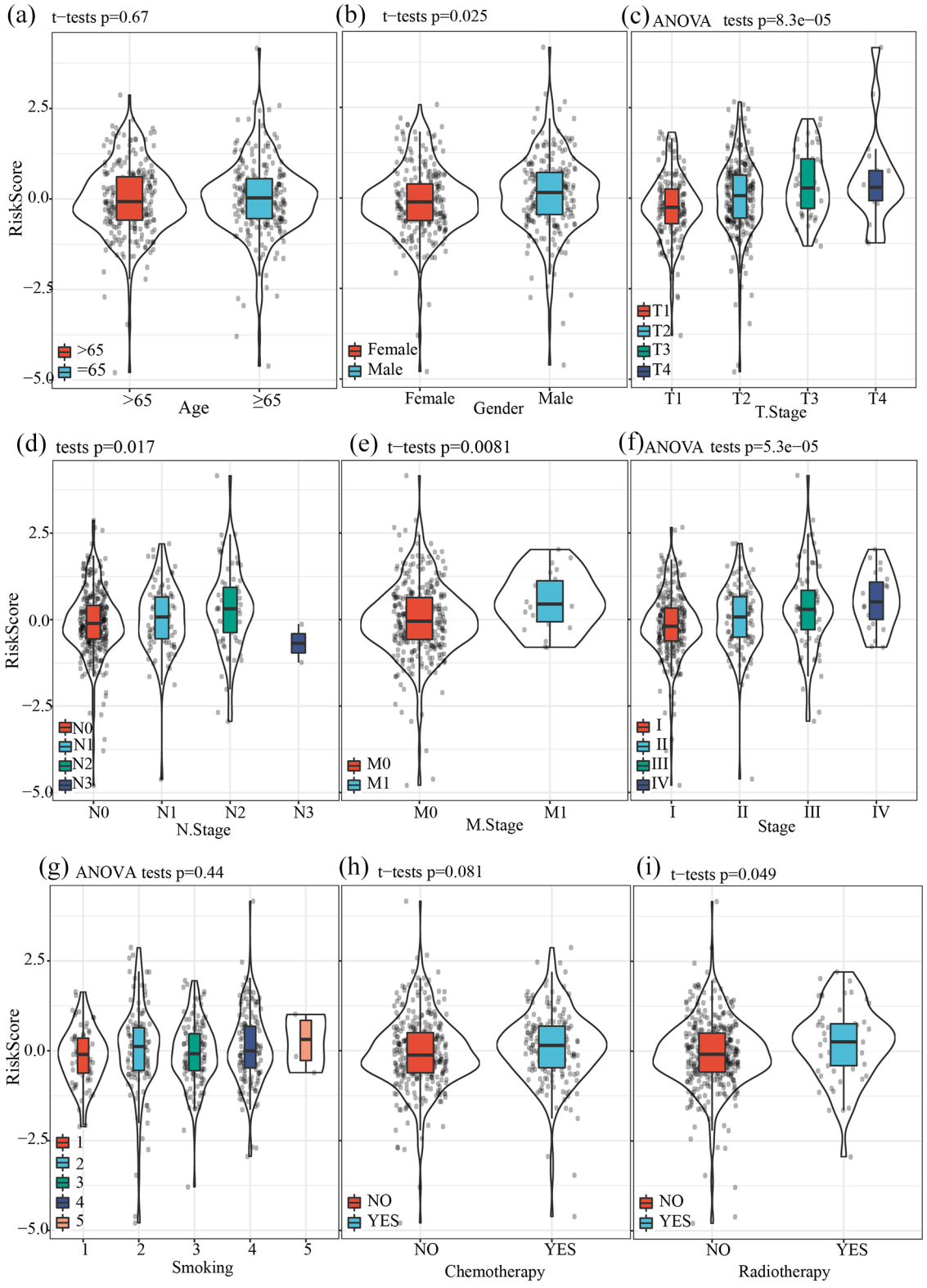

Figure 6. The difference in (a) age, (b) gender, (c) survival status, (d) T stage, (e) N stage, (f) M stage, (g) clinic stage, (h) smoking, (i) chemotherapy, and (j) radiotherapy in high- and low-risk groups. (A color version of this figure is available in the online journal.)

were substantially higher in the high-risk group, while myeloid dendritic cells, B lineage, and T cells had a higher score in the low-risk group (Figure 7(a)). The score generated by ssGSEA was analyzed for differences between the two risk groups, and 16 of the 28 immune cells showed different scores between the two risk groups (Figure 7(b)). The two risk groups also showed significantly different immune scores and ESTIMATE scores (Figure 7(c)). Furthermore, 22 of the

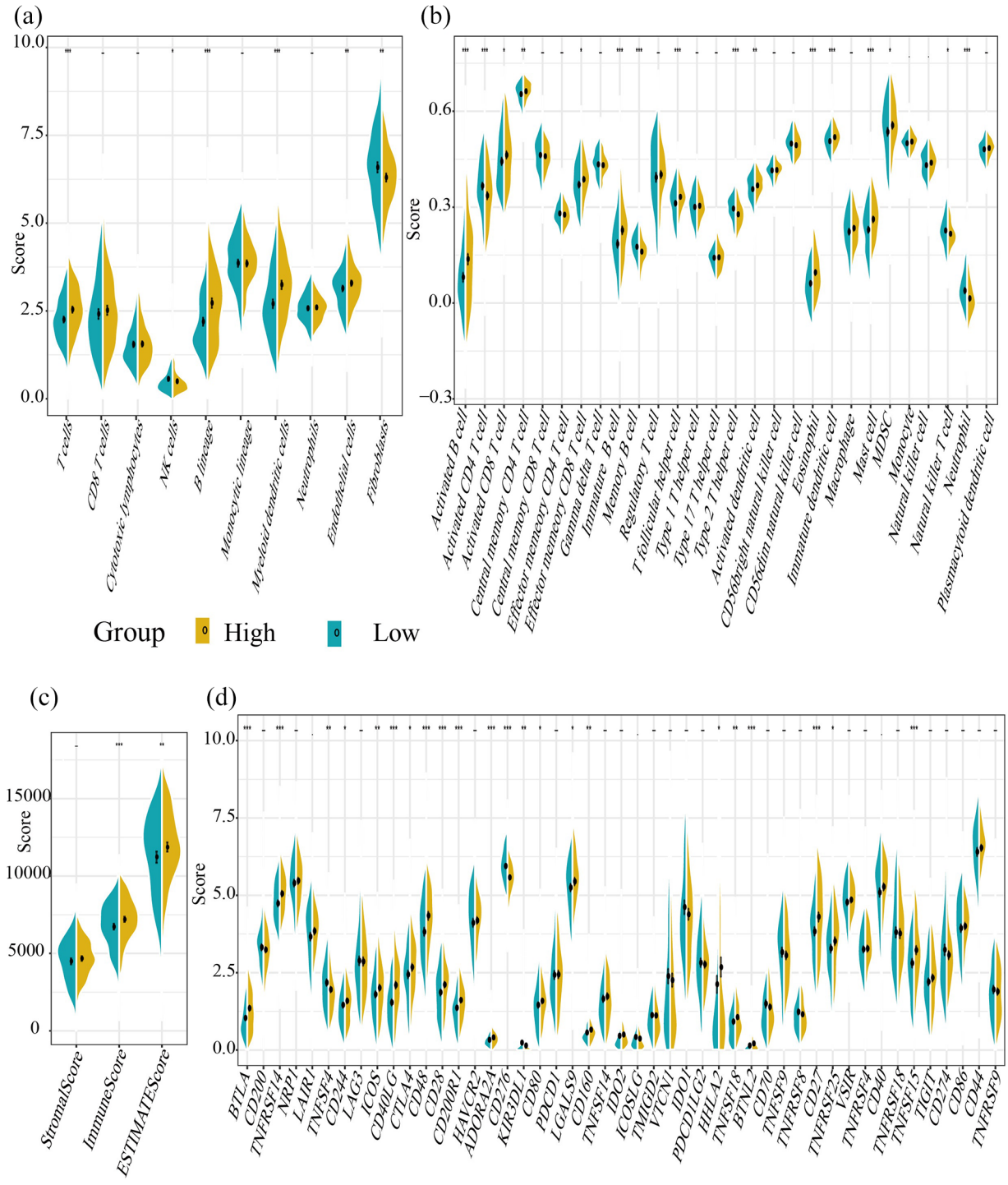

Figure 7. Analysis of immune characteristics in each risk group. (a) MCP-counter analyzed the score differences of 10 infiltrating immune cells between the low- and high-risk groups. (b) The ssGSEA score differences of 28 immune cells between two risk groups. (c) Differences in the stromal score, immune score, and ESTIMATE score between the two subgroups. (d) Differential expression analysis of 47 immune checkpoints in low- and high-risk groups. (A color version of this figure is available in the online journal.)

47 immune sites were expressed differentially between the low- and high-risk groups (Figure 7(d)).

The 7-miRNA signature was an independent indicator affecting the prognosis of LUAD patients

We used Cox regression analysis to examine the association between clinical features and patient survival throughout the dataset. The findings of univariate Cox analysis showed that clinic stage, risk type, N stage, M stage, chemotherapy, and T stage were substantially associated with the survival of LUAD patients (Figure 8(a)). Multivariate Cox analysis illustrated that N stage, chemotherapy, and risk type might serve as independent prognostic indicators for LUAD (Figure 8(b)). The following survival analysis showed that

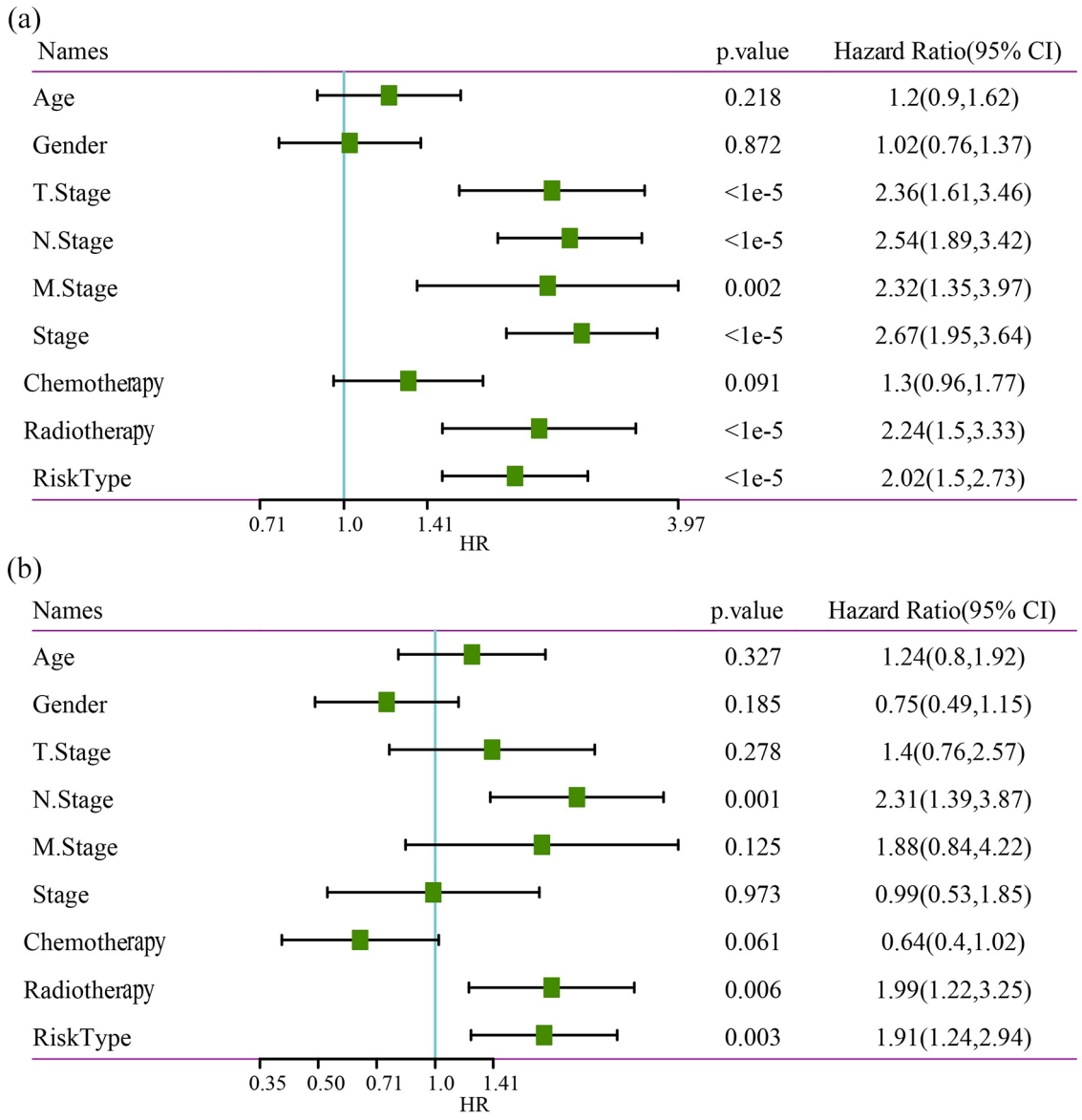

Figure 8. Cox regression analyses of prognostic indicators and patients' OS. (a) Univariate and (b) multivariate Cox regression analysis were employed to analyze clinicopathological factors (including risk type) and OS in the whole dataset. (A color version of this figure is available in the online journal.)

the 7-miRNA model showed a high performance in stratifying of age ⩾65 years male and female, M0, M1, chemotherapy, N4, N0, N1, N2, T2, T3, clinical stage I + II, clinical stage III + IV, non-chemotherapy, and non-radiotherapy (Figure 9). These results suggested that our 7-miRNA signature independently acted as a prognostic indicator for anticipating the prognosis of LUAD patients.

The establishment of nomogram and calibration curves

We used risk score and N stage to build a nomogram to anticipate the patients' one-, three-, and five-year OS. The risk score was shown to be the most significant factor influencing clinical status, according to the nomogram (Figure 10(a)). The calibration chart demonstrated that the one-year and three-year deviation correction lines were greatly close to the ideal curve. Therefore, the performance of the nomogram was highly effective in predicting one-year and three-year OS (Figure 10(b)). The DCA chart showed that the nomogram had more net income than the N stage (Figure 10(c)), suggesting a high accuracy of the nomogram in the prediction of LUAD OS.

Evaluation of the preponderance of 7-miRNA signature in predicting prognosis of LUAD patients

In addition, 7-miRNA signatures were compared with other previous LUAD prognosis miRNA signatures in this study. Four studies were retrieved from NCBI PubMed,[14,20–22] and the risk score of each LUAD sample was computed based on the signatures in each report. Through the generated survival curve, we found that the OS of high-risk patients classified according to these four signatures was substantially reduced in contrast with that of low-risk patients (Figure 11(a) to (d)). Time-dependent ROC curves showed that the

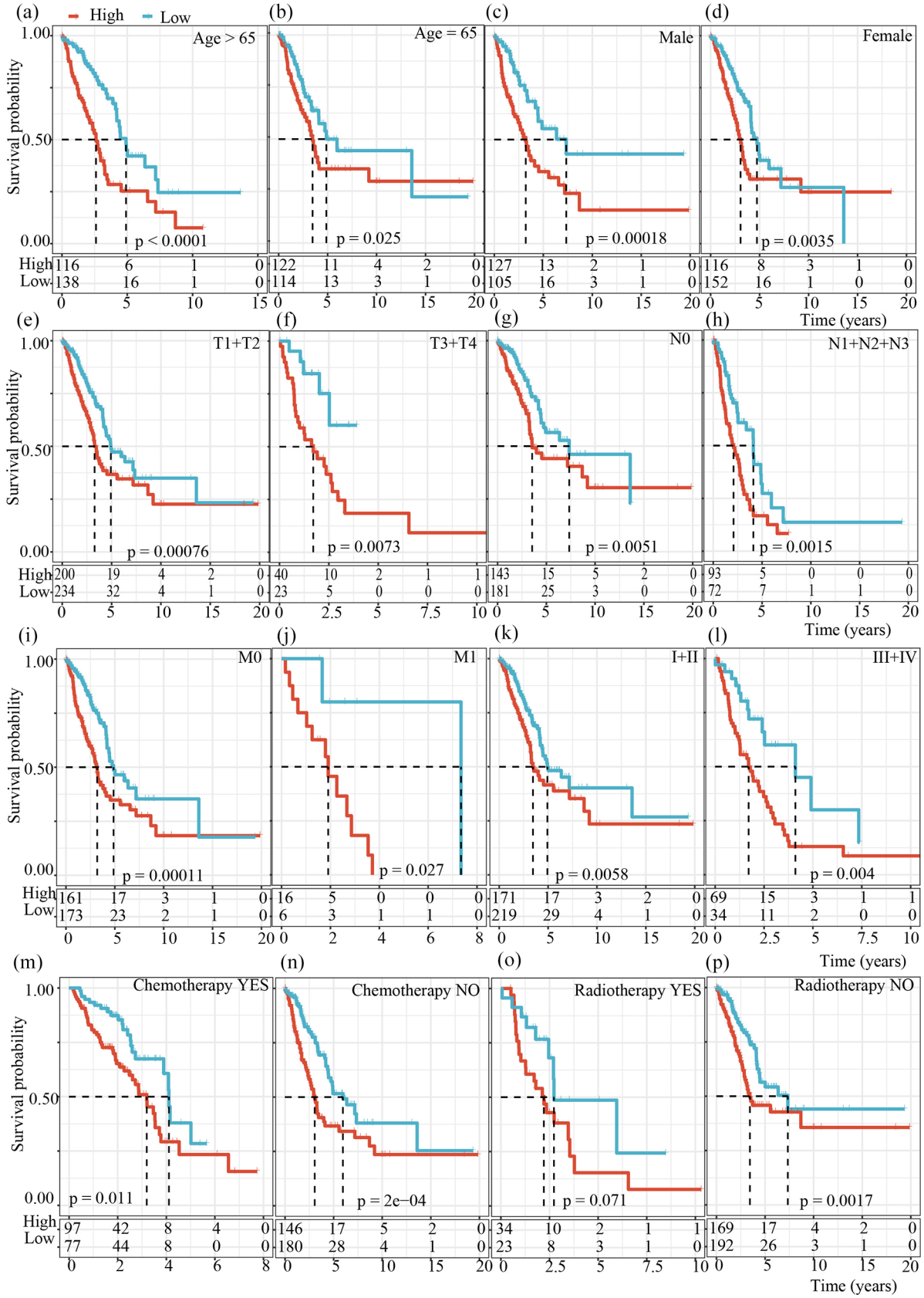

Figure 9. Prognosis of risk scores for the LUAD patients categorized by the clinical feature. (a) Age >65 years. (b) Age = 65 years. (c) Male. (d) Female. (e) T1 + T2. (f) T3 + T4. (g) N0. (h) N1 + N2 + N3. (i) M0. (j) M1. (k) Clinic stage I + II. (l) Clinic stage III + IV. (m) Chemotherapy. (n) No chemotherapy. (o) Radiotherapy. (p) No radiotherapy. (A color version of this figure is available in the online journal.)

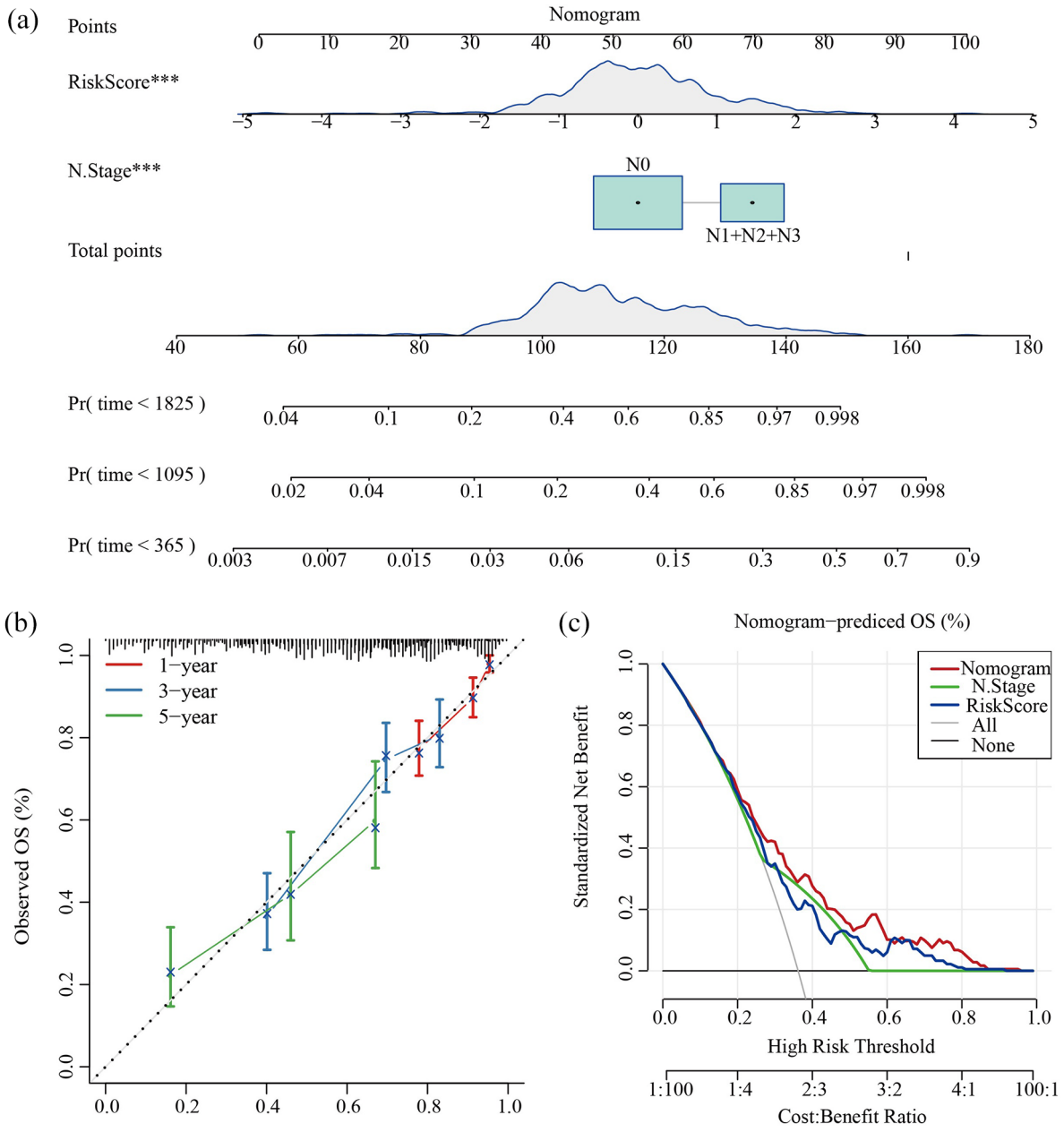

Figure 10. Nomogram in predicting OS for LUAD patients. (a) Construction of nomogram based on independent prognostic factors risk score and N stage. (b) The one-, three-, and five-year OS are shown by the calibration plots of the nomogram. The 45° dashed line represents the ideal reference line. (c) DCA was used to confirm the effect of the nomogram. (A color version of this figure is available in the online journal.)

7-miRNA signature in this study predicted the prognosis of patients in the entire TCGA-LUAD with an AUC of 0.75, 0.7, and 0.66 at one, three, and five years, respectively, which was generally higher than the four previous signatures (Figure 11(e) to (h)).

Analysis of serum expression and regulation of miRNAs

We also analyzed the expression of seven miRNAs in the risk model in normal serum and tumor serum. The findings illustrated that, in GSE13704, the expression levels of seven miRNAs in LUAD patients' serum were significantly higher than that in the control serum (Figure 12(a)). In the other three serum datasets (GSE31568, GSE68951, and GSE40738), seven miRNAs showed few or no expression differences between tumor serum and normal serum (Figure 12(b) to (d)), which was different from the results obtained in TCGA (Figure 12(e)).

To explore the regulation pathway of the miRNAs in the 7-miRNA signature, we first predicted the targets of miRNAs using TargetScan, miRDB, and MiRTarBase and summarized the target motifs predicted in the database. Then, we examined the functions and pathways related to these target genes. GO cellular component (CC) and biological process (BP) analysis revealed that they were mainly concentrated in BPs and molecular functions (MFs) related to synaptic formation and their regulatory activities (Figure 13(a) and (b)). GO MF analysis showed that the predicted target genes were strongly

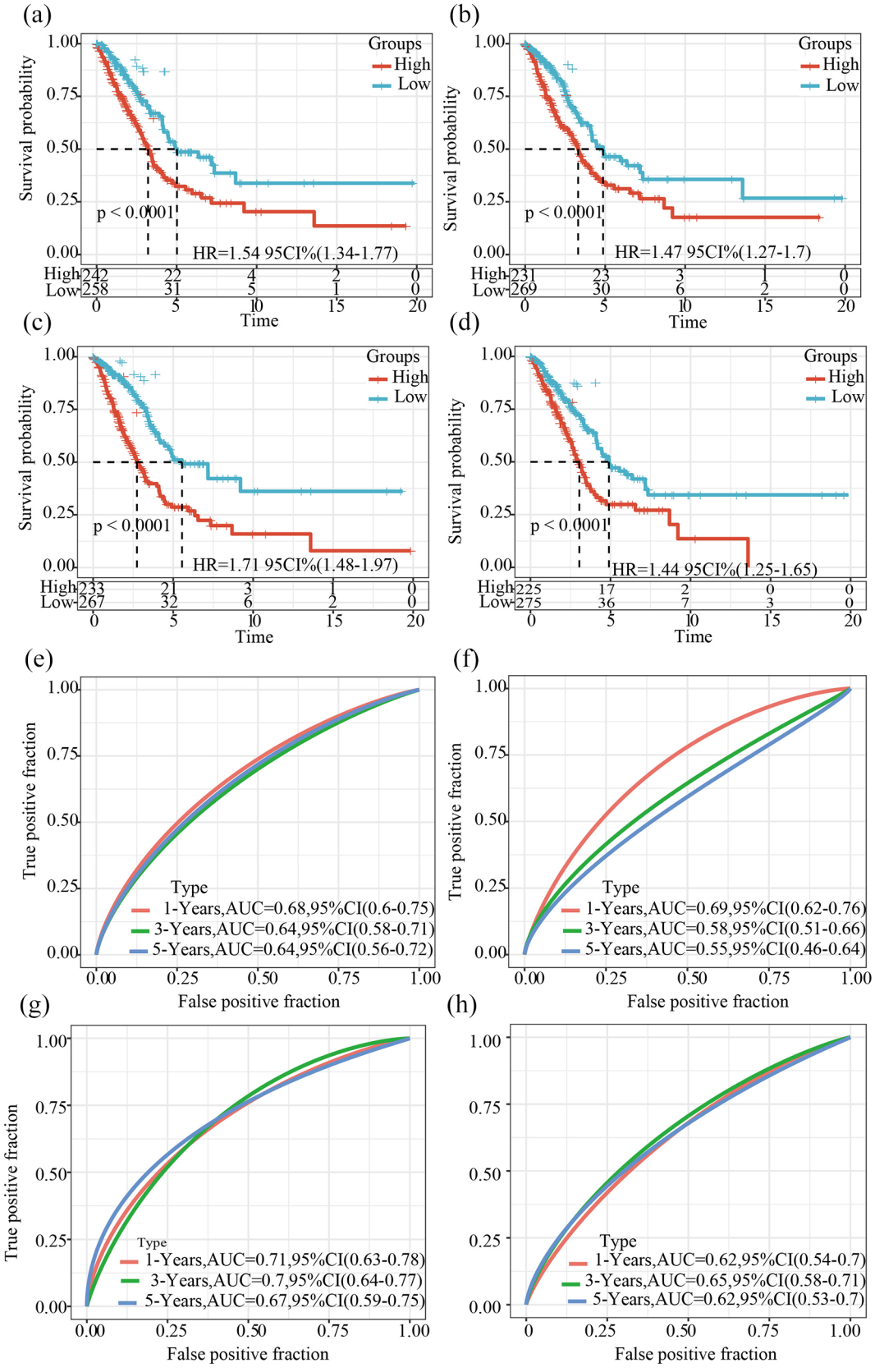

Figure 11. Evaluation of the preponderance of 7-miRNA signature in predicting prognosis of LUAD. (a) Kaplan–Meier curves of patients based on 4-miRNA signature grouping. (b) Survival curves of TCGA-LUAD samples grouped by 8-miRNA signature. (c) The death risk of TCGA-LUAD samples was evaluated according to the 6-miRNA risk model. (d) The miRNA risk score model composed of miR-5844, miR-29b-1, miR-148a, and miR-375 predicted the survival outcome of samples in TCGA-LUAD. (e) Prognostic ROC curve of TCGA-LUAD samples was based on 4-miRNA signature. (f) The ROC curve of TCGA-LUAD sample OS was predicted based on an 8-miRNA signature. (g) The ROC curve of TCGA-LUAD sample OS was predicted according to the 6-miRNA risk model. (h) Evaluation of the predictive performance of the risk score model was composed of miR-5844, miR-29b-1, miR-148a, and miR-375. (A color version of this figure is available in the online journal.)

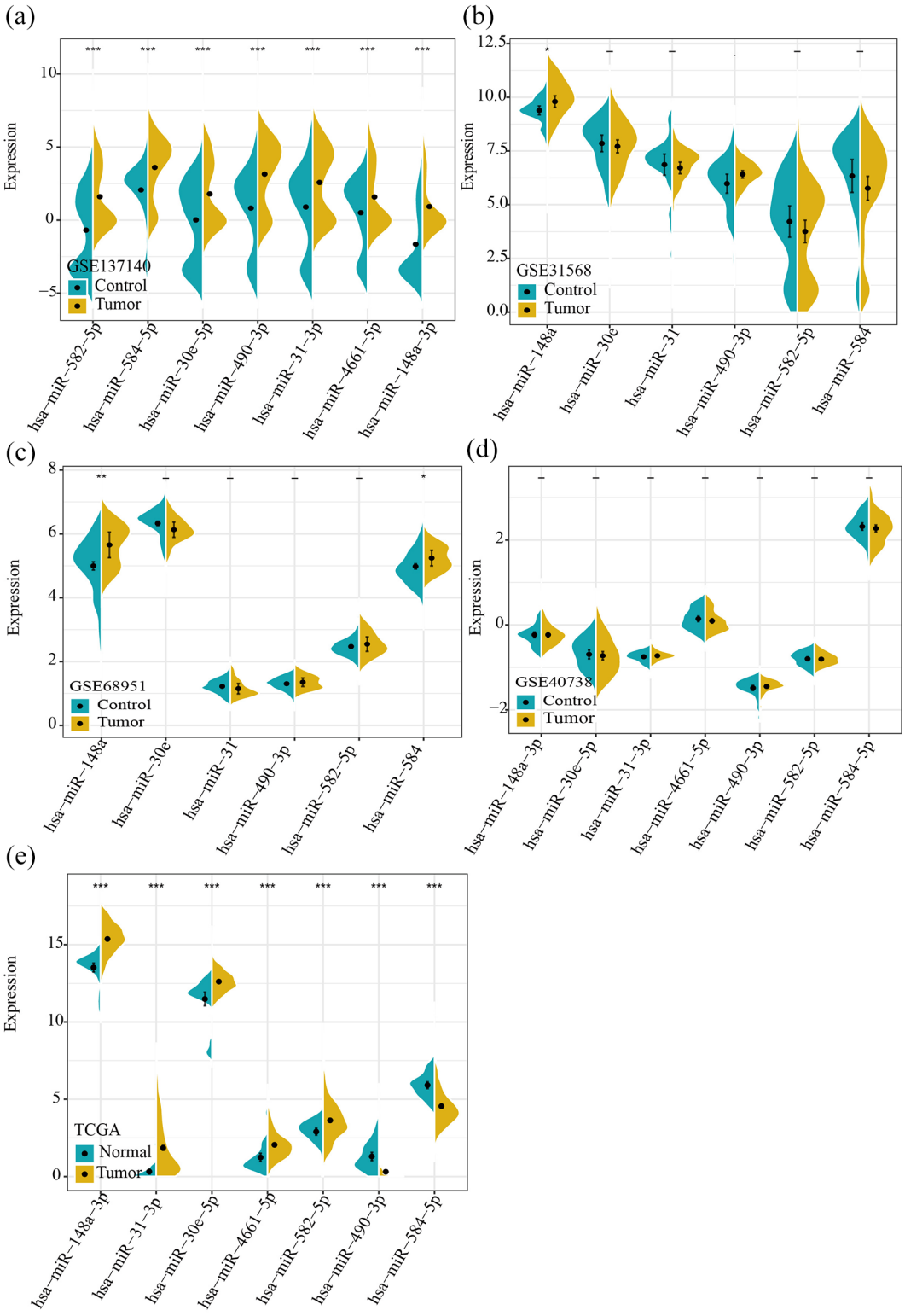

Figure 12. Expression analysis of seven miRNAs in risk model in normal serum and tumor serum. (a) The expression difference of seven miRNAs between LUAD serum and control serum in GSE13704. (b) Expression of seven miRNAs between normal serum and tumor serum in GSE31568 (c) Expression of seven miRNAs in normal serum and tumor serum of GSE68951. (d) In GSE40738, the expression of seven miRNAs was analyzed between LUAD serum and control serum. (e) Differential expression analysis of miRNA between tumor and normal tissues in TCGA-LUAD cohort. (A color version of this figure is available in the online journal.)

enriched to Tau protein binding, core promoter binding, core promoter sequence-specific DNA binding, and so on (Figure 13(c)). The study of the KEGG pathway demonstrated that the target groups of miRNAs were significantly enriched in circadian rhythm, aldosterone-regulated sodium reabsorption, and small cell lung cancer pathways (Figure 13(d)). These results suggested that the target genes for the seven miRNAs were involved in non-small cell tumor growth.

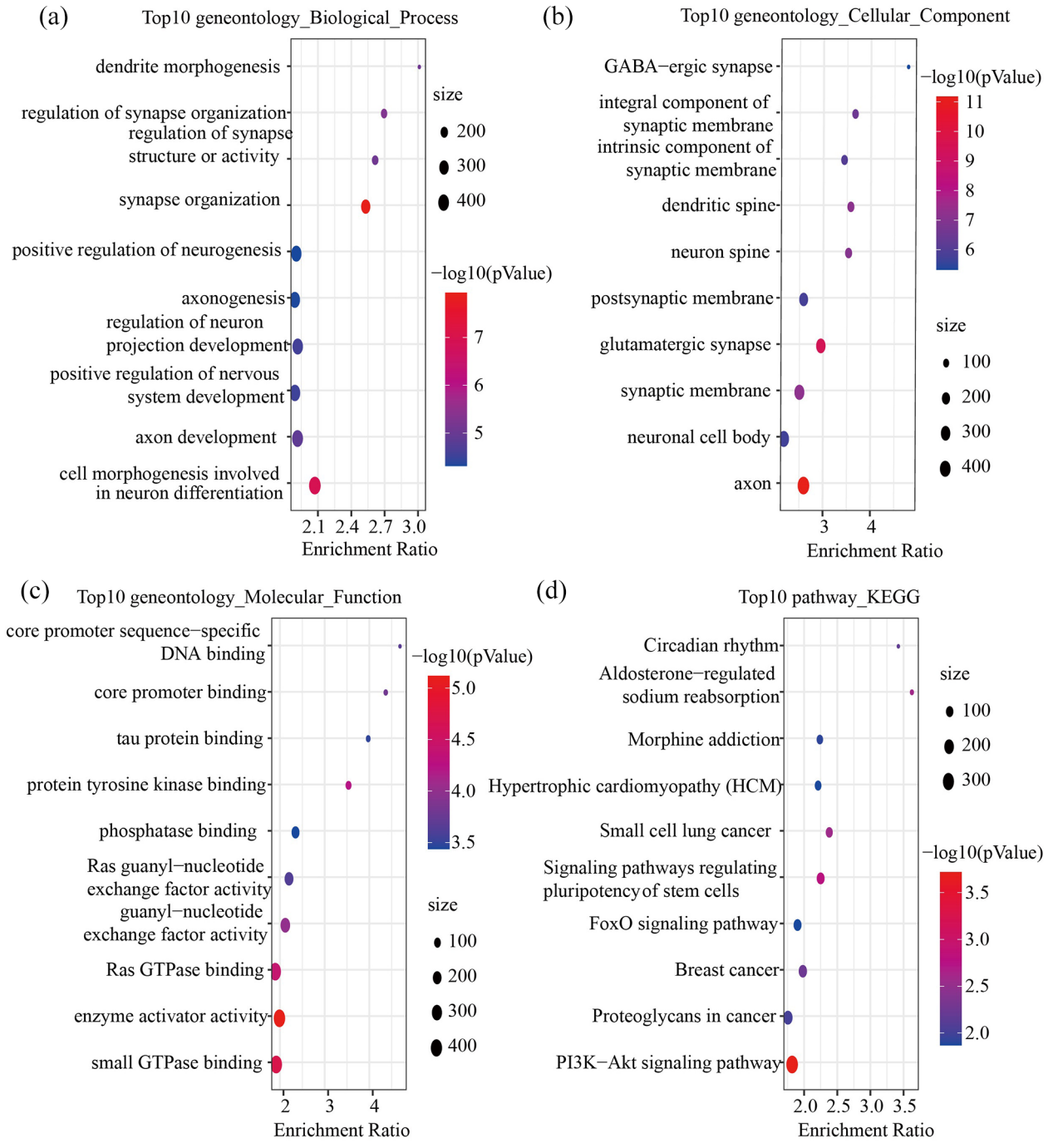

Figure 13. Go and KEGG enrichment analysis of miRNA target genes. (a) Go BP analysis of miRNA target genes. (b) GO CC analysis of miRNA target genes. (c) GO molecular function analysis of miRNA target genes. (d) KEGG pathway enrichment analysis of miRNA target genes. (A color version of this figure is available in the online journal.)

Discussion

The identification of reliable prognostic biomarkers is a major area of interest within the field of tumor research. MiRNAs are seen as a group of highly promising biomarkers of cancer as they are resistant to degradation in many tissue types and are easy to measure; more importantly, miRNAs are associated with the presence of tumors or clinically related cancer characteristics.[23,24] The retrospective study of Barger and Nana-Sinkam[25] concluded that in lung cancer, miRNA expression profile can distinguish histological subtypes, predict chemotherapy response, and relate to metastasis, prognosis, and survival of patients. Researches on miRNAs in the diagnosis of LUAD are gradually emerging. Based on a large number of samples and data from multiple sources, a novel 16-miRNA signature is developed as a promising biomarker for determining LUAD pathological staging.[26] Yang *et al.*[27] established a 4-miRNA signature for the diagnosis of lung cancer, and its efficacy has been successfully verified by RT-PCR with a large number of subjects. Thus, it can be seen

that miRNAs have great potential in the clinical application of LUAD.

In this research, we aimed to develop a prognostic miRNA signature for LUAD. From TCGA, we obtained the data of 500 LUAD samples along with 46 normal samples and screened 295 differentially expressed miRNAs. LASSO and Univariate Cox regression analyses illustrated that there were seven prognosis-related miRNAs, including miR-31-3p, miR-584-5p, miR-30e-5p, miR-582-5p, miR-4661-5p, miR-490-3p, and miR-148a-3p, from the 295 differentially expressed miRNAs. The role of some of the seven miRNAs in lung cancer has been previously explored. It was reported that miRNA-148a-3p inhibits proliferation and epithelial–mesenchymal transformation of non-small cell lung cancer (NSCLC) by modulating Ras/MAPK/ERK signal transduction.[28] A bioinformatics research report indicated that miR-31-3p is associated with OS of LUAD patients, as it could promote lymph node metastasis and poor prognosis.[29] MiR-30e-5p inhibits carcinogenesis by downmodulating USP22-mediated Sirt1/JAK/STAT3 signaling in NSCLC.[30] Furthermore, miR-582-5p and miR-584-5p are also reported to be independent prognostic biomarkers for patients with NSCLC.[31,32] MiR-490-3p also suppresses the proliferation and the metastasis of cells by downregulating Wnt/β-catenin signal pathway in LUAD.[33] Therefore, all of the above-mentioned miRNAs perform a critical function in the progression of cancer, suggesting the combination of them in a joint signature may represent a new prognostic marker for LUAD.

We used these seven miRNAs to build a risk score model and verified it in three cohorts. According to the risk score calculated by the score model, the patients were classified into groups. The ROC analysis and Kaplan–Meier survival analysis verified that the 7-miRNA signature had high sensitivity and specificity in the prognosis of LUAD. Multivariate Cox regression analysis confirmed that the signature independently predicted the prognosis of LUAD patients and was significantly correlated with clinicopathological features of LUAD. More importantly, we also found that the 7-miRNA signature performed better in predicting OS when in a nomogram combined with the N stage.

As miRNAs perform their biological functions by regulating target mRNAs, mRNAs may be more closely related to cancer progression.[34] In this study, we predicted seven target genes of miRNA and analyzed the functions and pathways of these target genes by GO and KEGG. The results of GO MF analysis illustrated that the predicted target genes were considerably enriched in Tau protein binding, core promoter binding, core promoter sequence-specific DNA and binding, and these processes are integral to cancer development. KEGG analysis intuitively illustrated that the target genes of these seven miRNAs were significantly enriched in NSCLC. Thus, we assumed that the seven target genes of miRNA may be related to LUAD tumor development and that their potential functions and regulatory mechanisms should be further studied in the future. However, this research had some shortcomings that needed to be addressed. As the sample size used was not rich enough, the signature requires further validation in a larger independent dataset. Second, this study was based on bioinformatics analysis with a public database and has not been tested in clinical practice. In addition, the training and test sets basically came from one database and need to be verified in a separate queue from another database.

To sum up, we developed a 7-miRNA signature that can efficaciously and independently anticipate the survival and prognosis of patients with LUAD. Moreover, the signature combined with the N stage in a nomogram could more accurately predict the prognosis of patients with LUAD. The findings of this study might provide the direction for evaluating the clinical prognosis of patients with LUAD.

AUTHORS' CONTRIBUTIONS

All authors participated in the design, interpretation of the studies, and analysis of the data and review of the article. R.J.L., Q.T., and Q.Q.L. designed the research; Z.Y.G. and J.H. acquired the data; R.J.L. and J.T.L. analyzed and interpreted the data; Q.T. and Z.Y.G. performed the statistical analysis. The first article was written by R.J.L. The article was reviewed and modified by Q.T. and R.J.L. All of the authors approved the article, and they all agreed to publish the article.

DECLARATION OF CONFLICTING INTERESTS

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

FUNDING

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID ID

Ruijun Liu https://orcid.org/0000-0001-9353-9133

SUPPLEMENTAL MATERIAL

Supplemental material for this article is available online.

REFERENCES

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2021;**71**:209–49
2. Zhang Y, Lin Q, Xu T, Deng W, Yu J, Liao Z, Yue J. Out of the darkness and into the light: new strategies for improving treatments for locally advanced non-small cell lung cancer. *Cancer Lett* 2018;**421**:59–62
3. Spella M, Stathopoulos GT. Immune resistance in lung adenocarcinoma. *Cancers (Basel)* 2021;**13**:384
4. Goldstraw P, Crowley J, Chansky K, Giroux DJ, Groome PA, Rami-Porta R, Postmus PE, Rusch V, Sobin L, International Association for the Study of Lung Cancer International Staging Committee, Participating Institutions. The IASLC Lung Cancer Staging Project: proposals for the revision of the TNM stage groupings in the forthcoming (seventh) edition of the TNM Classification of malignant tumours. *J Thorac Oncol* 2007;**2**:706–14
5. Ball D. TNM in non-small cell lung cancer: a staging system for all oncologists or just for surgeons? *Ann Transl Med* 2019;**7**:S103
6. Mlecnik B, Bindea G, Pages F, Galon J. Tumor immunosurveillance in human cancers. *Cancer Metastasis Rev* 2011;**30**:5–12
7. Leng X, Wei S, Mei J, Deng S, Yang Z, Liu Z, Guo C, Deng Y, Xia L, Cheng J, Zhao K, Gan F, Li C, Merrell KW, Molina JR, Metro G,

Liu L. Identifying the prognostic significance of B3GNT3 with PD-L1 expression in lung adenocarcinoma. *Transl Lung Cancer Res* 2021;**10**:965–80
8. Ye SL, Li XY, Zhao K, Feng T. High expression of CD8 predicts favorable prognosis in patients with lung adenocarcinoma: a cohort study. *Medicine (Baltimore)* 2017;**96**:e6472
9. Nagashio R, Oikawa S, Yanagita K, Hagiuda D, Kuchitsu Y, Igawa S, Naoki K, Satoh Y, Ichinoe M, Murakumo Y, Saegusa M, Sato Y. Prognostic significance of G6PD expression and localization in lung adenocarcinoma. *Biochim Biophys Acta Proteins Proteom* 2019;**1867**:38–46
10. Xu JY, Zhang C, Wang X, Zhai L, Ma Y, Mao Y, Qian K, Sun C, Liu Z, Jiang S, Wang M, Feng L, Zhao L, Liu P, Wang B, Zhao X, Xie H, Yang X, Zhao L, Chang Y, Jia J, Wang X, Zhang Y, Wang Y, Yang Y, Wu Z, Yang L, Liu B, Zhao T, Ren S, Sun A, Zhao Y, Ying W, Wang F, Wang G, Zhang Y, Cheng S, Qin J, Qian X, Wang Y, Li J, He F, Xiao T, Tan M. Integrative proteomic characterization of human lung adenocarcinoma. *Cell* 2020;**182**:245–61.e17
11. Wang L, Zhao H, Xu Y, Li J, Deng C, Deng Y, Bai J, Li X, Xiao Y, Zhang Y. Systematic identification of lincRNA-based prognostic biomarkers by integrating lincRNA expression and copy number variation in lung adenocarcinoma. *Int J Cancer* 2019;**144**:1723–34
12. Li JP, Li R, Liu X, Huo C, Liu TT, Yao J, Qu YQ. A seven immune-related lncRNAs model to increase the predicted value of lung adenocarcinoma. *Front Oncol* 2020;**10**:560779
13. Yu F, Liang M, Wu W, Huang Y, Zheng J, Zheng B, Chen C. Upregulation of long non-coding RNA GCC2-AS1 facilitates malignant phenotypes and correlated with unfavorable prognosis for lung adenocarcinoma. *Front Oncol* 2020;**10**:628608
14. Hayes J, Peruzzi PP, Lawler S. MicroRNAs in cancer: biomarkers, functions and therapy. *Trends Mol Med* 2014;**20**:460–9
15. Li W, Liu S, Su S, Chen Y, Sun G. Construction and validation of a novel prognostic signature of microRNAs in lung adenocarcinoma. *PeerJ* 2021;**9**:e10470
16. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 2015;**43**:e47
17. Engebretsen S, Bohlin J. Statistical predictions with glmnet. *Clin Epigenetics* 2019;**11**:123
18. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making* 2006;**26**:565–74
19. Wang J, Vasaikar S, Shi Z, Greer M, Zhang B. WebGestalt 2017: a more comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit. *Nucleic Acids Res* 2017;**45**:W130–7
20. Lin Y, Lv Y, Liang R, Yuan C, Zhang J, He D, Zheng X, Zhang J. Four-miRNA signature as a prognostic tool for lung adenocarcinoma. *Onco Targets Ther* 2018;**11**:29–36
21. Li X, Shi Y, Yin Z, Xue X, Zhou B. An eight-miRNA signature as a potential biomarker for predicting survival in lung adenocarcinoma. *J Transl Med* 2014;**12**:159
22. Chen B, Gao T, Yuan W, Zhao W, Wang TH, Wu J. Prognostic value of survival of microRNAs signatures in non-small cell lung cancer. *J Cancer* 2019;**10**:5793–804
23. Ferracin M, Negrini M. Micromarkers 2.0: an update on the role of microRNAs in cancer diagnosis and prognosis. *Expert Rev Mol Diagn* 2015;**15**:1369–81
24. Di Leva G, Croce CM. MiRNA profiling of cancer. *Curr Opin Genet Dev* 2013;**23**:3–11
25. Barger JF, Nana-Sinkam SP. MicroRNA as tools and therapeutics in lung cancer. *Respir Med* 2015;**109**:803–12
26. Yang Z, Yin H, Shi L, Qian X. A novel microRNA signature for pathological grading in lung adenocarcinoma based on TCGA and GEO data. *Int J Mol Med* 2020;**45**:1397–408
27. Yang X, Su W, Chen X, Geng Q, Zhai J, Shan H, Guo C, Wang Z, Fu H, Jiang H, Lin J, Lagisetty KH, Zhang J, Li Y, Yang S, Massion PP, Beer DG, Chang AC, Ramnath N, Chen G. Validation of a serum 4-microRNA signature for the detection of lung cancer. *Transl Lung Cancer Res* 2019;**8**:636–48
28. Xie Q, Yu Z, Lu Y, Fan J, Ni Y, Ma L. MicroRNA-148a-3p inhibited the proliferation and epithelial-mesenchymal transition progression of non-small-cell lung cancer via modulating Ras/MAPK/Erk signaling. *J Cell Physiol* 2019;**234**:12786–99
29. Wang Y, Shang S, Yu K, Sun H, Ma W, Zhao W. MiR-224, miR-147b and miR-31 associated with lymph node metastasis and prognosis for lung adenocarcinoma by regulating PRPF4B, WDR82 or NR3C2. *PeerJ* 2020;**8**:e9704
30. Xu G, Cai J, Wang L, Jiang L, Huang J, Hu R, Ding F. MicroRNA-30e-5p suppresses non-small cell lung cancer tumorigenesis by regulating USP22-mediated Sirt1/JAK/STAT3 signaling. *Exp Cell Res* 2018;**362**:268–78
31. Wang LL, Zhang M. MiR-582-5p is a potential prognostic marker in human non-small cell lung cancer and functions as a tumor suppressor by targeting MAP3K2. *Eur Rev Med Pharmacol Sci* 2018;**22**:7760–7
32. Guo T, Zheng C, Wang Z, Zheng X. MiR5845p regulates migration and invasion in nonsmall cell lung cancer cell lines through regulation of MMP14. *Mol Med Rep* 2019;**19**:1747–52
33. Li Z, Jiang D, Yang S. MiR-490-3p inhibits the malignant progression of lung adenocarcinoma. *Cancer Manag Res* 2020;**12**:10975–84
34. Takahashi RU, Prieto-Vila M, Kohama I, Ochiya T. Development of miRNA-based therapeutic approaches for cancer patients. *Cancer Sci* 2019;**110**:1140–7

(Received August 24, 2021, Accepted November 30, 2021)