



# A Robust 8-Gene Prognostic Signature for Early-Stage Non-small Cell Lung Cancer

Ru He<sup>1</sup> and Shuguang Zuo<sup>1,2\*</sup>

<sup>1</sup> Center for Translational Medicine, Huaihe Hospital of Henan University, Kaifeng, China, <sup>2</sup> Institute of Infection and Immunity, Huaihe Hospital of Henan University, Kaifeng, China

**Background:** The current staging system is imprecise for prognostic prediction of early-stage non-small cell lung cancer (NSCLC). This study aimed to develop a robust prognostic signature for early-stage NSCLC, allowing classification of patients with a high risk of poor outcome and specific treatment decision.

**Method:** In the present study, a comprehensive genome-wide profiling analysis was conducted using a retrospective pool of early-stage NSCLC patient data from the previous datasets of Gene Expression Omnibus (GEO) including GSE31210, GSE37745, and GSE50081 and The Cancer Genome Atlas (TCGA). Cox proportional hazards models were implemented to determine the association between gene expression levels and overall patient survival in each dataset. The common genes among all datasets were selected as candidate prognostic genes. A risk score model was developed and validated using four independent datasets and the entire cohort. The Kaplan-Meier with log-rank test was used to assess survival difference.

**Results:** A univariate Cox proportional hazards regression analysis for each dataset showed that a total of 2280 genes in GSE31210, 762 genes in GSE37745, 871 genes in GSE50081, and 666 genes in TCGA were identified as candidate protective genes, while overall 2131 genes in GSE31210, 913 in GSE37745, 1107 in GSE50081, and 997 in TCGA were identified as candidate risky genes. There were 8 common genes associated with overall survival, including 7 mRNA and 1 lncRNA. By using the Step-wise multivariate Cox analysis, an 8-gene prognostic signature (CDCP1, HMMR, TPX2, CIRBP, HLF, KBTBD7, SEC24B-AS1, and SH2B1) for early-stage NSCLC was developed. Patients in the high-risk group had shorter overall survival than those in the low-risk group. Multivariate regression and stratified analysis suggested that the prognostic power of the 8-gene signature was independent of other clinical factors. Furthermore, the 8-gene signature achieved AUC values of 0.726, 0.701, 0.725 and 0.650 in GSE31210, GSE37745, GSE50081 and TCGA, respectively. Moreover, the combination of the 8-gene signature and the stage resulted to a better patient classification for survival prediction and treatment decision.

**Conclusion:** This study developed a robust gene signature with great value for prognostic prediction in early-stage NSCLC, which may contribute to patient classification and personalized treatment decisions.

**Keywords:** non-small cell lung cancer, overall survival, risk score, stage, prognostic signature

## OPEN ACCESS

### Edited by:

Alfredo Addeo,  
Geneva University Hospitals  
(HUG), Switzerland

### Reviewed by:

Laura Mezquita,  
Institut Gustave Roussy, France  
Marzia Del Re,  
University of Pisa, Italy

### \*Correspondence:

Shuguang Zuo  
zuosg@icloud.com

### Specialty section:

This article was submitted to  
Thoracic Oncology,  
a section of the journal  
Frontiers in Oncology

**Received:** 03 May 2019

**Accepted:** 12 July 2019

**Published:** 31 July 2019

### Citation:

He R and Zuo S (2019) A Robust  
8-Gene Prognostic Signature for  
Early-Stage Non-small Cell Lung  
Cancer. *Front. Oncol.* 9:693.  
doi: 10.3389/fonc.2019.00693

## INTRODUCTION

Lung cancer is a highly lethal malignant disease, the second most common cancer in men and women, and the leading cause of cancer-related death worldwide (1). Non-small cell lung cancer (NSCLC), accounts for 85% of all lung cancers, and is the predominant histological type. Despite recent therapeutic advances, patients with NSCLC are still associated with bleak outcomes, due to lack of early diagnostic and predictive biomarkers (2). Pulmonary resection is the primary treatment for early-stage NSCLC, with a 5-year survival rate of about 60% (3). Recently, it has been shown that adjuvant chemotherapy confers a survival advantage of 4–15% for patients with resected stage II–III (4–7), but not for patients with stage I disease (8, 9). The limited survival advantage suggests the deficiency of the current staging system and the presence of unknown tumor factors. It is imperative to develop novel prognostic biomarkers for risk stratification and treatment optimization in early patients.

Recent advances in microarray profiling and genome-wide sequencing have facilitated the identification of molecular prognostic factors that are crucial for precise classification of human cancers and personalized treatment decisions. A large number of studies in early-stage NSCLC have demonstrated that genomic data generated from patients with long-term follow-up are superior to the current staging system in estimating risk of worse prognosis. In those studies, numerous gene signatures have been generated to classify NSCLC patients with different clinical outcomes (10–14). However, no reliable and consistent gene signatures have emerged from these efforts. Additionally, the vast majority of studies have focused on single molecules, either mRNAs or lncRNAs (10, 15). Numerous works have demonstrated that mRNA and lncRNA signatures could precisely predict the prognosis of cancers (16–18). lncRNAs, a type of ncRNAs, have sequence lengths of more than 200 nucleotides with little or no protein-coding function (19), but mRNAs have protein-coding ability. lncRNAs and mRNAs crosstalk by sharing miRNA response elements, thereby generating competing endogenous RNA network (20). Relative to protein-coding mRNAs, lncRNAs are more closely associated with the status of cancer (21, 22). The single-biomarker for evaluating cancer prognosis is less robust relative to the more widely reported multiple-biomarker-based models (23). However, few studies have identified prognostic and predictive signatures by combining both mRNAs and lncRNAs. The increasing availability of genome-wide gene expression data in NSCLC makes it feasible to identify a robust gene signature. In the present study, several published datasets from the Gene Expression Omnibus (GEO) and The Cancer Genome Atlas (TCGA) were mined, in order to produce a robust prognostic signature for early-stage NSCLC. An 8-gene signature with reliable prognostic power in early-stage NSCLC was identified, which might cover the shortage of the current staging system, improve patient stratification, and provide promise for more personalized therapeutic interventions.

## METHODS

### Patients and Study Design

The raw data of gene expression and corresponding clinical information of patients with early-stage NSCLC were downloaded from GEO and TCGA, respectively. In the study, three independent datasets were retrieved from GEO, including GSE31210 (24, 25), GSE37745 (26), and GSE50081 (27), and one dataset was employed from TCGA. After the samples without enough clinical information or with advanced disease were removed, a total of 1,331 patients were finally enrolled, including 226 patients from GSE31210, 165 from GSE37745, 181 from GSE50081, and 759 from TCGA. The gene expression data of the three GEO datasets were generated by Affymetrix U133 Plus 2.0 microarray platform, while the TCGA data were analyzed on the Illumina sequencing platform.

In the present study, initially, the candidate genes that were associated with the overall survival of early-stage NSCLC patients from each dataset were identified, and the credible prognostic genes of the four overlapping datasets were selected. Then, the prognostic signature was developed using a risk score model and validated using four datasets and the entire cohort. **Figure 1** illustrates the flow diagram of this study.

### Prognostic Signature

A univariate Cox proportional hazard regression model was used to assess the association of gene expression with the overall survival of NSCLC patients in each cohort. The hazard ratio (HR) from the univariate Cox regression analysis was used to identify candidate genes associated with the overall survival from each dataset. Genes with  $HR < 1$  were considered as protective genes and those with  $HR > 1$  were defined as risky genes. Meanwhile, genes with  $P < 0.05$  were considered statistically significant. In order to improve reliability, only common genes between the four datasets were screened to construct the prognostic signature.

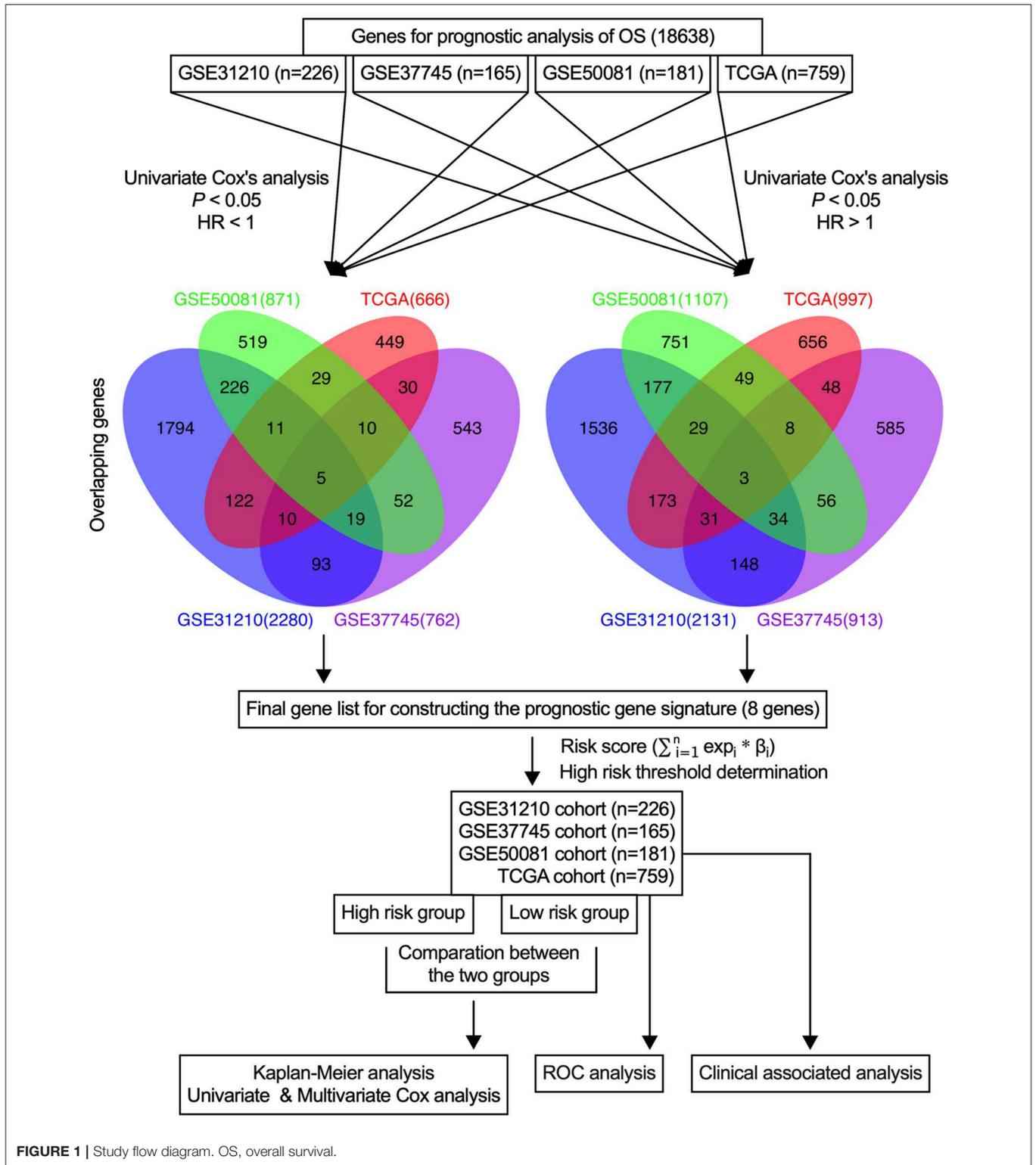
By combining the expression values of prognostic genes weighted by their regression coefficients, a risk score for each patient was constructed as follows:

$$Risk\ score = \sum_{i=1}^n \exp_i^* \beta_i$$

where  $n$  was the number of prognostic genes,  $\exp_i$  the expression value of gene  $i$ , and  $\beta_i$  the regression coefficient of gene  $i$  in the univariate Cox regression analysis. Using the median risk score as a cutoff value, NSCLC patients were classified into high- and low-risk groups. Moreover, the relationship between the prognosis signature and disease-free survival was investigated based on the three cohorts of GSE31210, GSE37745, and GSE50081.

### Statistical Analysis

The Kaplan-Meier method was used to assess the differences in survival time of low- and high-risk NSCLC patients, and the log-rank test was used to determine the statistical significance of observed differences between groups. Multivariable Cox



regression analysis and stratification analysis were used to assess whether the risk score was independent of other clinical features. The time-dependent receiver operating characteristic (ROC)

curve was used to measure the prognostic performance by comparing the areas under the ROC curves (AUC). Significance was defined as  $P < 0.05$ .

## RESULTS

### Prognostic Signature Generation

In this study, a univariate Cox proportional hazards regression analysis in each dataset was conducted, and candidate genes that were significantly correlated with the overall survival were identified. Under the cutoff values of  $P < 0.05$  and  $HR < 1$ , 2,280 genes in GSE31210, 762 genes in GSE37745, 871 genes in GSE50081, and 666 genes in TCGA were identified as candidate protective genes. Under the cut-off values of  $P < 0.05$  and  $HR > 1$ , 2,131 genes in GSE31210, 913 in GSE37745, 1,107 in GSE50081, and 997 in TCGA were identified as candidate risky genes. After combing the candidate protective genes in GSE31210, GSE37745, GSE50081, and TCGA, a total of 5 common genes were remained. Similarly, there were 3 common candidate risky genes after combing the identified candidate risky genes in the four datasets. By overlapping the four datasets, eight common genes (CDCP1, HMMR, TPX2, CIRBP, HLF, KBTBD7, SEC24B-AS1, and SH2B1) were finally identified, which were used to form the prognostic signature. The general information of the 8 genes is displayed in **Table 1**. Among them, 7 genes (CDCP1, HMMR, TPX2, CIRBP, HLF, KBTBD7, and SH2B1) were mRNA and one gene (SEC24B-AS1) was lncRNA. In **Table 2**, the prognostic correlation of the 8 genes with the overall survival of early-stage NSCLC patients in each dataset is shown.

### 8-Gene Prognostic Signature Validation

A risk score was constructed with the regression coefficients from the univariate Cox analysis, and a prognostic model was developed to predict overall survival. In the prognostic model, the risk score for each patient was calculated. The patients in each dataset were classified into high- and low-risk groups, based on the median risk score, which was used as the cutoff point. In **Figure 2**, the risk score distribution, gene expression, and the patients' survival status in each dataset were displayed, ranked according to the risk score values for the 8-gene signature. The

resulted data demonstrated that the patients in the high-risk group had a shorter overall survival than those in the low-risk group (GSE31210:  $HR = 4.74$ , 95%  $CI = 2.07-10.87$ ,  $P = 5.09e-05$ ; GSE37745:  $HR = 2.23$ , 95%  $CI = 1.54-3.23$ ,  $P = 1.23e-05$ ; GSE50081:  $HR = 2.33$ , 95%  $CI = 1.45-3.75$ ,  $P = 3.34e-04$ ; TCGA:  $HR = 1.59$ , 95%  $CI = 1.18-2.14$ ,  $P = 2.25e-03$ ) (**Figure 3** Left panel). Then, we evaluated the survival difference in 3 groups, including high-, moderate-, and low-risk groups. The results showed that the higher the risk score was, the worse the survival of patients was (GSE31210:  $P = 3.41e-05$ ; GSE37745:  $P = 1.76e-05$ ; GSE50081:  $P = 1.76e-05$ ; TCGA:  $P = 3.50e-06$ ) (**Figure 3** Middle panel). These results confirmed that risk score can be used as a prognostic indicator. The time-dependent ROC curves showed that the 8-gene signature achieved AUC values of 0.726, 0.701, 0.725, and 0.650 in GSE31210, GSE37745, GSE50081, and TCGA, respectively (**Figure 3** Right panel), suggesting a substantially effective performance for overall survival prediction.

From the eight genes, 3 were associated with high risk (CDCP1, HMMR, and TPX2;  $HR > 1$ ) and 5 appeared to be protective (CIRBP, HLF, KBTBD7, SEC24B-AS1, and SH2B1;  $HR < 1$ ). The expression of the 8 prognostic genes was detected and the differences between high- and low-risk groups were compared. Patients with high-risk scores tended to express risky genes, whereas patients in the low-risk group tended to express protective genes (**Figure 4**).

### The 8-Gene Prognostic Signature Is Independent of Other Clinicopathological Factors

In order to evaluate the contribution of the 8-gene signature as an independent prognostic factor of patient survival, a multivariate Cox regression analysis was performed using a stepwise method. Covariates included the gene signature and clinicopathological factors, such as age, gender, stage, histologic type, gene mutation, smoking, and performance status. The results showed that the predictive ability of the 8-gene signature was independent of other clinicopathological factors for overall survival of early-stage NSCLC patients in four independent datasets (GSE31210:  $HR = 3.51$ , 95%  $CI = 1.47-8.37$ ,  $P = 4.60E-03$ ; GSE37745:  $HR = 2.69$ , 95%  $CI = 1.82-3.97$ ,  $P = 6.00E-07$ ; GSE50081:  $HR = 1.92$ , 95%  $CI = 1.10-3.37$ ,  $P = 2.20E-02$ ; TCGA:  $HR = 1.47$ , 95%  $CI = 1.08-2.00$ ,  $P = 1.40E-02$ ) (**Table 3**) and the entire cohort ( $HR = 1.90$ , 95%  $CI = 1.55-2.33$ ,  $P = 8.70E-10$ ) (**Table 4**). Stage IB is the indication of adjuvant chemotherapy. Univariate and multivariate Cox regression model suggested that stage IA/IB in GSE37745 ( $HR = 1.71$ , 95%  $CI = 1.07-2.73$ ,  $P = 2.50E-02$  for univariate model, and  $HR = 1.69$ , 95%  $CI = 1.04-2.72$ ,  $P = 3.30E-02$  for multivariate model) was significantly correlated with overall survival of the patients, but stage IA/IB in other three cohorts did not show any significant association with overall survival (**Table 3**).

### Stratification Analysis

In the multivariate Cox regression analysis, several clinicopathological factors were also identified as independent

**TABLE 1** | General information of the 8 genes for constructing the prognostic signature.

Gene stable ID	Gene name	Gene type	Chromosome	Gene start (bp)	Gene end (bp)
ENSG00000163814	CDCP1	Protein coding	3	45082278	45146422
ENSG00000099622	CIRBP	Protein coding	19	1259384	1274880
ENSG00000108924	HLF	Protein coding	17	55265012	55325065
ENSG00000072571	HMMR	Protein coding	5	163460203	163491945
ENSG00000120696	KBTBD7	Protein coding	13	41189833	41194566
ENSG00000247950	SEC24B-AS1	antisense	4	109347475	109433817
ENSG00000178188	SH2B1	Protein coding	16	28846600	28874212
ENSG00000088325	TPX2	Protein coding	20	31739271	31801805

**TABLE 2** | Univariate regression analysis of 8 genes and overall survival of NSCLC patients in 4 datasets.

Genes	GSE31210		GSE37745		GSE5008		TCGA	
	HR (95% CI)	P						
CDCP1	2.17 (1.28–3.67)	3.80E-03	1.93 (1.42–2.63)	2.40E-05	1.42 (1.04–1.95)	3.00E-02	1.17 (1.02–1.35)	2.20E-02
CIRBP	0.25 (0.13–0.48)	3.00E-05	0.65 (0.45–0.95)	2.70E-02	0.6 (0.39–0.93)	2.30E-02	0.67 (0.53–0.85)	8.40E-04
HLF	0.7 (0.57–0.86)	6.40E-04	0.85 (0.74–0.99)	3.40E-02	0.76 (0.63–0.92)	5.80E-03	0.92 (0.85–0.99)	3.10E-02
HMMR	1.61 (1.21–2.13)	1.00E-03	1.35 (1.1–1.66)	3.80E-03	1.29 (1.02–1.63)	3.50E-02	1.16 (1.01–1.33)	3.60E-02
KBTBD7	0.39 (0.18–0.83)	1.60E-02	0.71 (0.52–0.97)	3.20E-02	0.64 (0.46–0.9)	9.60E-03	0.77 (0.61–0.97)	2.40E-02
SEC24B-AS1	0.3 (0.17–0.55)	9.90E-05	0.62 (0.4–0.97)	3.70E-02	0.48 (0.24–0.95)	3.50E-02	0.78 (0.65–0.94)	1.00E-02
SH2B1	0.34 (0.15–0.79)	1.20E-02	0.38 (0.19–0.74)	4.70E-03	0.32 (0.13–0.78)	1.30E-02	0.79 (0.65–0.96)	1.70E-02
TPX2	1.49 (1.18–1.89)	8.90E-04	1.17 (1.01–1.35)	3.70E-02	1.27 (1.08–1.49)	4.40E-03	1.12 (1–1.24)	5.00E-02

HR, hazard ratio; CI, confidence interval.

prognostic factors. Subsequently, a stratification analysis was carried out to evaluate whether the 8-gene signature could predict patient survival within the same clinical factor subgroup. Patients in the entire cohort were factitiously stratified based on clinical parameters, such as age ( $\leq 65$ / $> 65$ ), gender (female/male), stage (I/II), and histologic type (adeno/squamous). The results showed that the 8-gene signature could classify patients of the same stratum of age, gender, stage, and histologic type into high- and low-risk groups. Patients with high risk scores had a shorter overall survival than those with low risk scores in each stratum (Figure 5).

## Survival Prediction by Stage and 8-Gene Signature Combination

Tumor stage has great survival predictive value in clinical practice. In this study, stage and risk score were proven to be independent prognostic factors in all four independent datasets and the entire cohort. Therefore, the development of a prognostic model for survival prediction was attempted by combining the stage with the 8-gene signature. Based on the stage status and the risk score, patients were divided into six groups: Group 1 (Stage IA and Low risk), Group 2 (Stage IA and High risk), Group 3 (Stage IB and Low risk), Group 4 (Stage IB and High risk), Group 5 (Stage II and Low risk), and Group 6 (Stage II and High risk) (Figure 6). Based on the results shown in Figure 6, the patients in each stage were all classified into low- and high-risk groups, and the patients of each stage in high-risk group had poor prognosis. The results indicated the patients in Group 2 had worse outcomes than those in Group 1, Group 4 had worse outcomes than those in Group 3, and Group 6 had worse outcomes than those in Group 5 (Figure 6). However, there was no significant difference in overall survival between the patients in Group 2 and Group 3/5. Furthermore, no difference in overall survival was observed between Group 4 and Group 5/6 (Figure 6). These results suggest that patients with high risk score in stage IA might have similar prognosis as those with low risk score in stage IB and stage II, suggesting that adjuvant chemotherapy should also be used in patients with stage IA who have a high risk score.

Among the six groups, Group 1 showed the best prognosis, whereas Group 6 exhibited the worst. In future practice, patients

could be classified into six groups according to their stages and risk scores to predict clinical outcomes. Significantly, there was difference in overall survival between stage IA and IB in the combined dataset (Figure 6, HR = 2.07, 95% CI = 1.59–2.69,  $P = 3.33 \times 10^{-8}$ ).

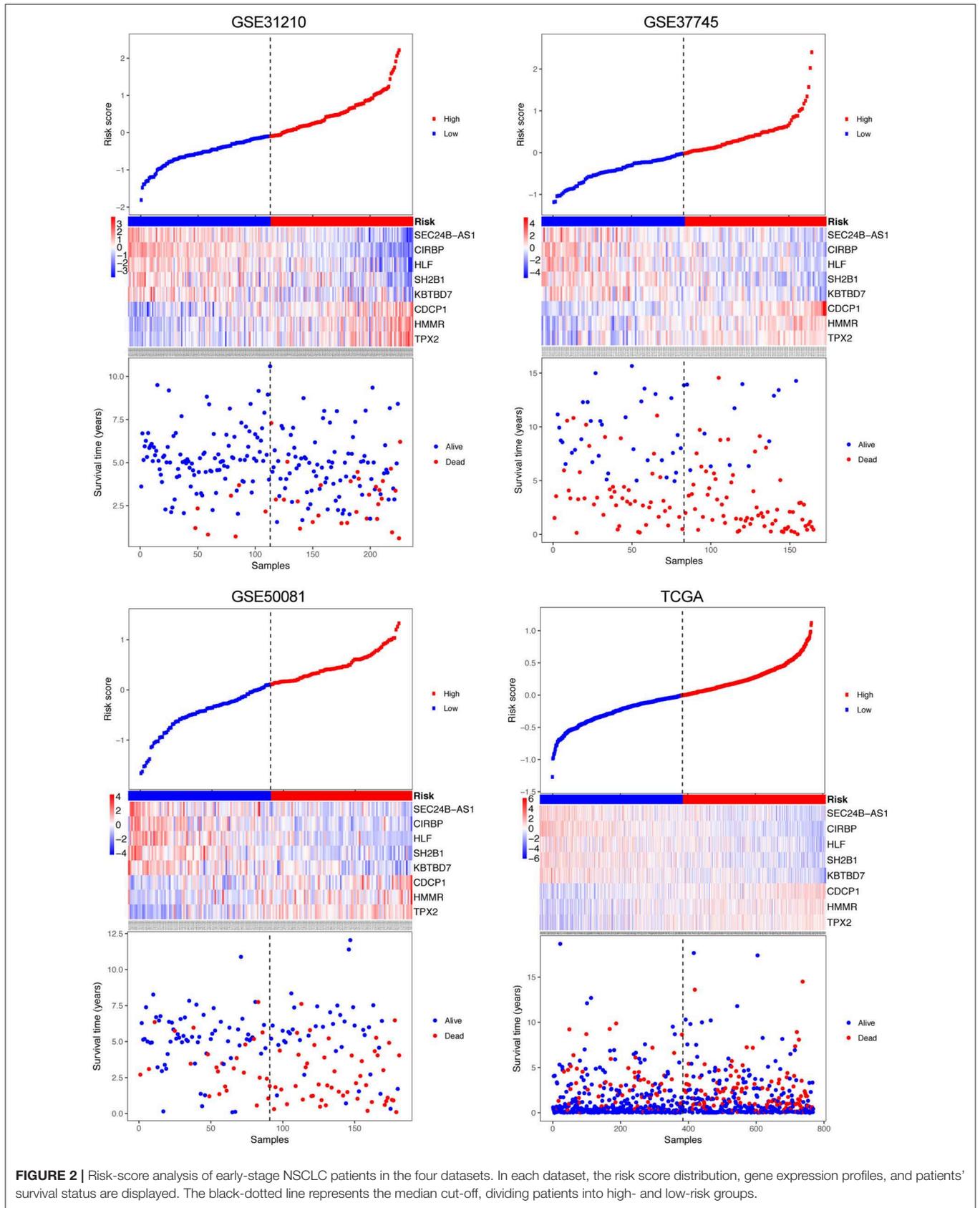
## Relationship Between the Prognosis Signature and Disease-Free Survival

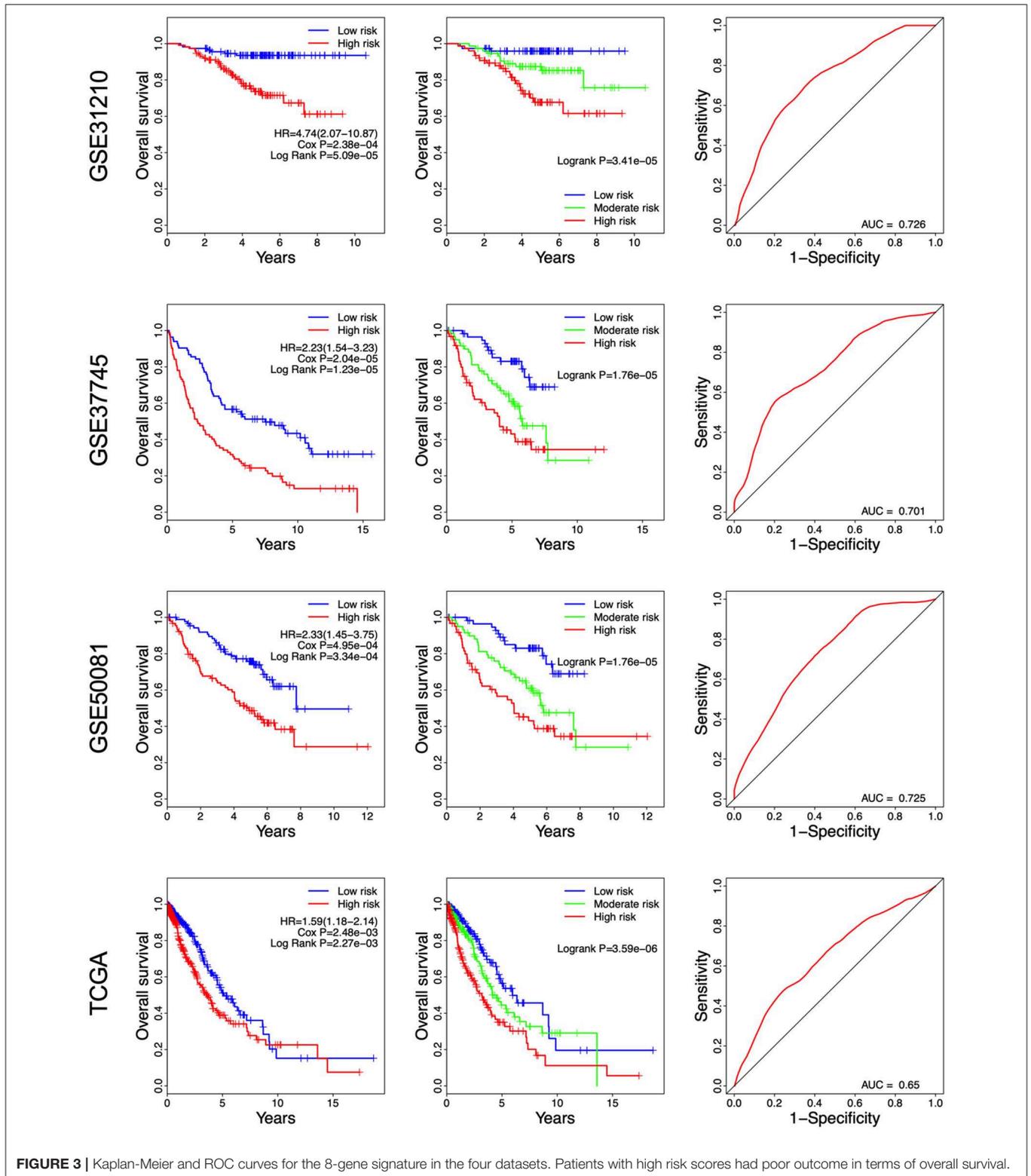
As shown in Figure 7, we found that the NSCLC patients in the high-risk group had a shorter disease-free survival, compared with those in the low-risk group (GSE31210: HR = 2.52, 95% CI = 1.48–4.27,  $P = 3.95 \times 10^{-4}$ ; GSE37745: HR = 2.05, 95% CI = 1.09–3.86,  $P = 0.023$ ; and GSE50081: HR = 3.94, 95% CI = 2.09–37.41,  $P = 4.48 \times 10^{-6}$ ). Higher AUC stands for a better performance. The time-dependent ROC curves showed that the AUC for the 8-gene signature achieved 0.605, 0.667, 0.651, 0.689, and 0.700 for the 1, 2, 3, 4, and 5 year survival in GSE31210, respectively. The 8-gene signature obtained 0.728, 0.692, 0.692, 0.659, and 0.677 for the 1, 2, 3, 4, and 5 year survival in GSE37745. Moreover, the AUC values of the 8-gene signature for the 1, 2, 3, 4, and 5 year survival in GSE50081 were respectively 0.744, 0.717, 0.693, 0.724, and 0.701 (Figure 7). These results suggest that there is a substantially effective performance for predicting disease-free survival.

## DISCUSSION

Increased understanding of the genomic changes of early-stage NSCLC promotes the discovery of prognostic and predictive signatures, and allows personalized treatment decisions. In this study, a novel 8-gene prognostic signature using genome-wide expression data from early-stage NSCLC patients was developed and validated. The developed 8-gene signature was able to identify early-stage NSCLC patients with high and low risk for poor prognosis. This signature may constitute an important step forward in treatment decision for early-stage NSCLC patients.

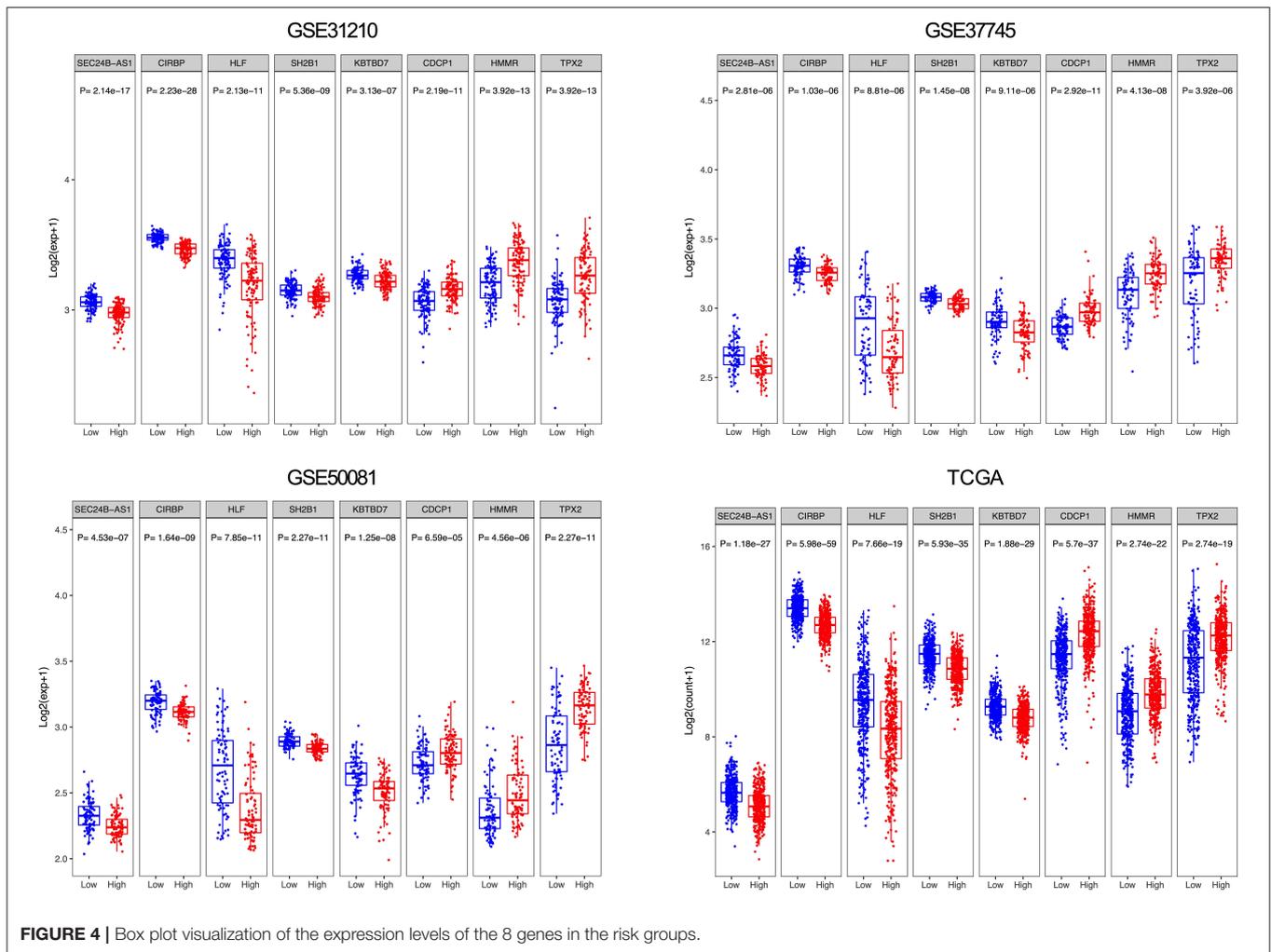
Previous studies have identified many molecular signatures that classify patients into different prognostic groups (10–14). However, these putative prognostic signatures demonstrated minimal overlap (10, 28). The discordant findings have been attributed to insufficient sample size, biological heterogeneity,





various expression platforms, and different statistical methods (10). In general, studies often use a training set to develop prognostic signatures (10, 12), which might lead to the

discordance. In the present study, survival-related genes were identified using a large patient cohort of four independent datasets. Only the common among the four datasets genes



were selected to build the gene signature, providing a more robust and reliable signature, relative to that derived from a single dataset, and partially handling the problem of discordance.

The mRNA CDCP1 was one of the 8-gene prognostic signature in our study. CDCP1, as a transmembrane protein, has been demonstrated to express in stem cells as well as hematopoietic cells (29). CDCP1 has been implicated to be highly expressed in many kinds of cancer cells, and to be related to over-proliferation, migration, invasion, and lymph node metastasis of lung cancer (30–32). Moreover, CDCP1 up-regulation is associated with the worse overall survival and recurrence-free survival of cancers (32, 33). HMMR has been demonstrated to inhibit cell proliferation of glioma in a dose-dependent way (34). HAMM over-expression in cancers has been implicated to cause centrosomal and mitotic dys-regulation, and to mediate apoptosis as well as cell cycle pathways (35). Moreover, HAMM has been suggested to have prognostic value, and affect the proliferative ability of chronic lymphocytic leukemia cells (36). Increased HAMM is correlated with poor prognosis in aggressive cancers (37, 38). A former study has reported that

there is a positive correlation between TPX2 up-regulation and lymph node metastasis, TNM stage, as well as poor prognosis of patients in cancers including cholangiocarcinoma, gastric cancer, and lung adenocarcinoma (39–41). Moreover, TPX2 silencing resulted in G2-M arrest, apoptosis and the suppression of cell migration and invasion of cancers (39, 42). TPX2 has been documented to mediate the cell growth and apoptosis via regulating PI3K/AKT/P21 signaling pathway in breast cancer (43). CIRBP high expression has been documented to have significantly better 5 year disease-free survival rate (44). CIRBP has been suggested to be a potential cell cycle regulator, and the loss of CIRBP might participate in the progression of endometrial carcinogenesis (45). Waters et al. (46) have suggested that HLF regulates the cell death, and is abnormally expressed in cancers. Chen et al. (47) have demonstrated that HLF-mediated miR-132 directly inhibits the expression of TTK, thereby playing inhibitory effects on cell growth, metastasis, as well as radio resistance of glioma. SH2B1, one member of the SH2B family, has been documented to serve as tumor activators in cancers. A previous study has implicated that SH2B1 is highly-expressed and linked with epithelial to mesenchymal

**TABLE 3 |** Univariate and multivariate Cox regression analyses of the gene signature and overall survival of NSCLC patients in 4 independent datasets.

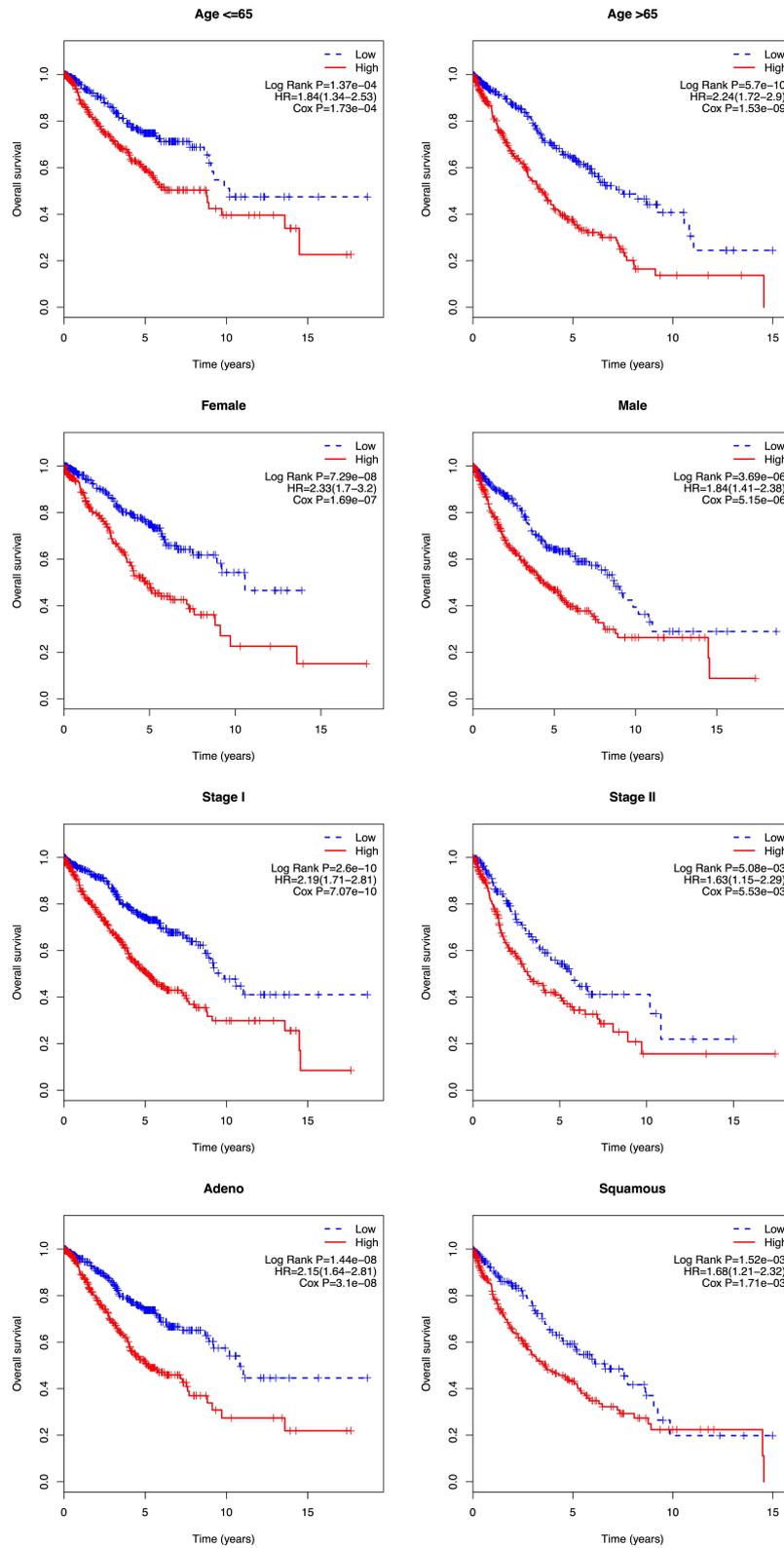
Variables	Patients (N)	Univariate analysis		Multivariate analysis		
		HR (95% CI)	P	HR (95% CI)	P	
<b>GSE31210</b>						
Age	< = 65/>65	176/50	2.58 (1.31–5.08)	6.00E-03	2.71 (1.31–5.63)	7.50E-03
Gender	Male/Female	121/105	1.52 (0.78–2.96)	2.20E-01		
ALK fusion	–/+	215/11	1.49 (0.36–6.24)	5.80E-01		
EGFR mutation	–/+	99/127	0.47 (0.24–0.93)	3.00E-02	0.77 (0.36–1.67)	5.10E-01
KRAS mutation	–/+	206/20	0.87 (0.27–2.85)	8.20E-01		
Myc	Low/High	207/17	0.70 (0.17–2.90)	6.20E-01		
Stage	IA/IB	114/54	2.47 (0.95–6.40)	6.30E-02	1.69 (0.64–4.46)	2.90E-01
Stage	IA/II	114/58	6.18 (2.68–14.26)	1.90E-05	4.89 (2.00–11.95)	4.90E-04
Smoking	No/Yes	115/111	1.64 (0.84–3.20)	1.50E-01		
Risk score	Low/High	113/113	4.74 (2.07–10.87)	2.40E-04	3.51 (1.47–8.37)	4.60E-03
<b>GSE37745</b>						
Age	< = 65/>65	82/83	1.51 (1.05–2.18)	2.60E-02	1.46 (1.00–2.13)	4.80E-02
Gender	Male/Female	74/91	1.11 (0.77–1.60)	5.90E-01		
Stage	IA/IB	40/90	1.71 (1.07–2.73)	2.50E-02	1.69 (1.04–2.72)	3.30E-02
Stage	IA/II	40/35	1.77 (1.02–3.06)	4.30E-02	1.81 (1.01–3.23)	4.50E-02
Histological type	Adeno/Large	89/21	0.97 (0.54–1.73)	9.10E-01		
Histological type	Adeno/Squamous	89/55	1.25 (0.84–1.85)	2.80E-01		
Performance status	0/1	87/66	2.04 (1.39–2.98)	2.40E-04	2.26 (1.53–3.34)	4.10E-05
Performance status	0/2–3	87/12	1.50 (0.74–3.04)	2.60E-01	1.47 (0.70–3.08)	3.10E-01
Risk score	Low/High	83/82	2.23 (1.54–3.23)	2.00E-05	2.69 (1.82–3.97)	6.00E-07
<b>GSE50081</b>						
Age	< = 65/>65	59/122	1.56 (0.93–2.61)	9.00E-02	1.30 (0.74–2.30)	3.60E-01
Gender	Male/Female	83/98	1.93 (1.19–3.14)	7.80E-03	1.73 (0.99–3.00)	5.30E-02
Stage	IA/IB	48/79	1.76 (0.93–3.34)	8.20E-02	1.78 (0.83–3.83)	1.40E-01
Stage	IA/II	48/54	2.46 (1.27–4.78)	7.80E-03	3.13 (1.39–7.01)	5.70E-03
Histological type	Adeno/Large	127/7	1.71 (0.62–4.74)	3.00E-01		
Histological type	Adeno/Other	127/4	1.84 (0.57–5.92)	3.00E-01		
Histological type	Adeno/Squamous	127/43	0.8 (0.46–1.39)	4.30E-01		
Smoking	No/Yes	24/136	1.39 (0.66–2.92)	3.90E-01		
Risk score	Low/High	91/90	2.33 (1.45–3.75)	4.90E-04	1.92 (1.10–3.37)	2.20E-02
<b>TCGA</b>						
Age	< = 65/>65	325/434	1.29 (0.95–1.75)	1.10E-01		
Gender	Male/Female	306/453	0.96 (0.71–1.29)	7.70E-01		
Stage	IA/IB	213/277	1.31 (0.87–1.97)	1.90E-01	1.23 (0.82–1.85)	3.20E-01
Stage	IA/II	213/269	2.04 (1.36–3.05)	5.30E-04	1.94 (1.28–2.92)	1.60E-03
Histological type	Adeno/Squamous	372/387	1.18 (0.88–1.60)	2.70E-01		
Risk score	Low/High	380/379	1.59 (1.18–2.14)	2.50E-03	1.47 (1.08–2.00)	1.40E-02

HR, hazard ratio; CI, confidence interval; Adeno, adenocarcinoma; Large, large cell carcinoma; Squamous, squamous cell carcinoma.

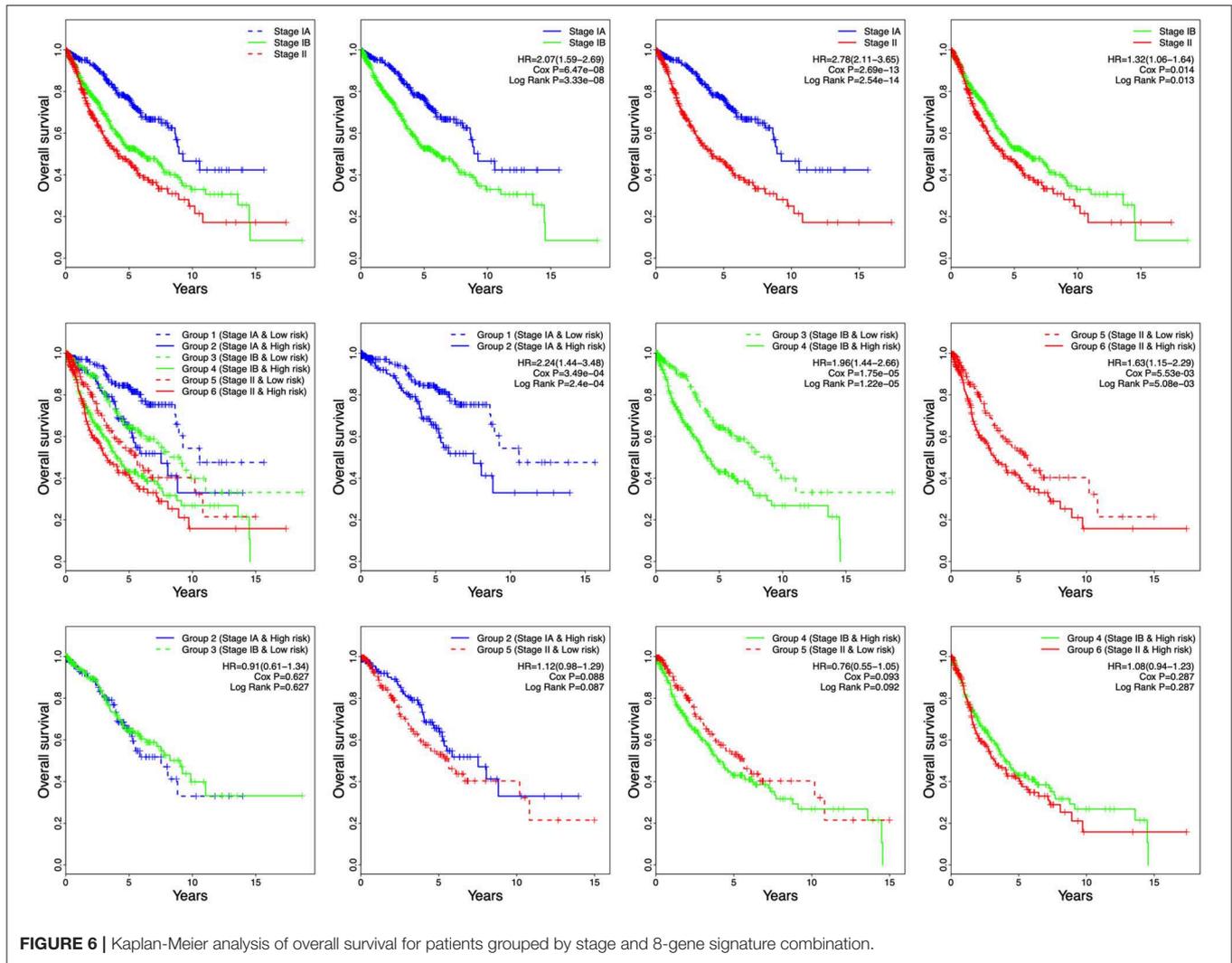
**TABLE 4 |** Univariate and multivariate Cox regression analyses of the gene signature and overall survival of NSCLC patients in entire cohort.

Variables	Patients (N)	Univariate analysis		Multivariate analysis		
		HR (95% CI)	P	HR (95% CI)	P	
Age	< = 65/>65	646/693	1.85 (1.51–2.25)	1.40E-09	1.77 (1.44–2.17)	4.00E-08
Gender	Male/Female	591/749	1.34 (1.10–1.63)	3.90E-03	1.11 (0.91–1.37)	3.00E-01
Stage	I/II	923/417	1.75 (1.43–2.14)	5.80E-08	1.62 (1.32–1.99)	3.80E-06
Histological type	Adeno/Large	819/28	1.61 (1.00–2.62)	5.20E-02	1.53 (0.94–2.49)	8.70E-02
Histological type	Adeno/Other	819/4	2.10 (0.67–6.57)	2.00E-01	1.83 (0.58–5.76)	3.00E-01
Histological type	Adeno/Squamous	819/489	1.76 (1.44–2.15)	3.00E-08	1.33 (1.08–1.64)	7.80E-03
Risk score	Low/High	671/669	2.07 (1.69–2.53)	1.60E-12	1.90 (1.55–2.33)	8.70E-10

HR, hazard ratio; CI, confidence interval; Adeno, adenocarcinoma; Large, large cell carcinoma; Squamous, squamous cell carcinoma.



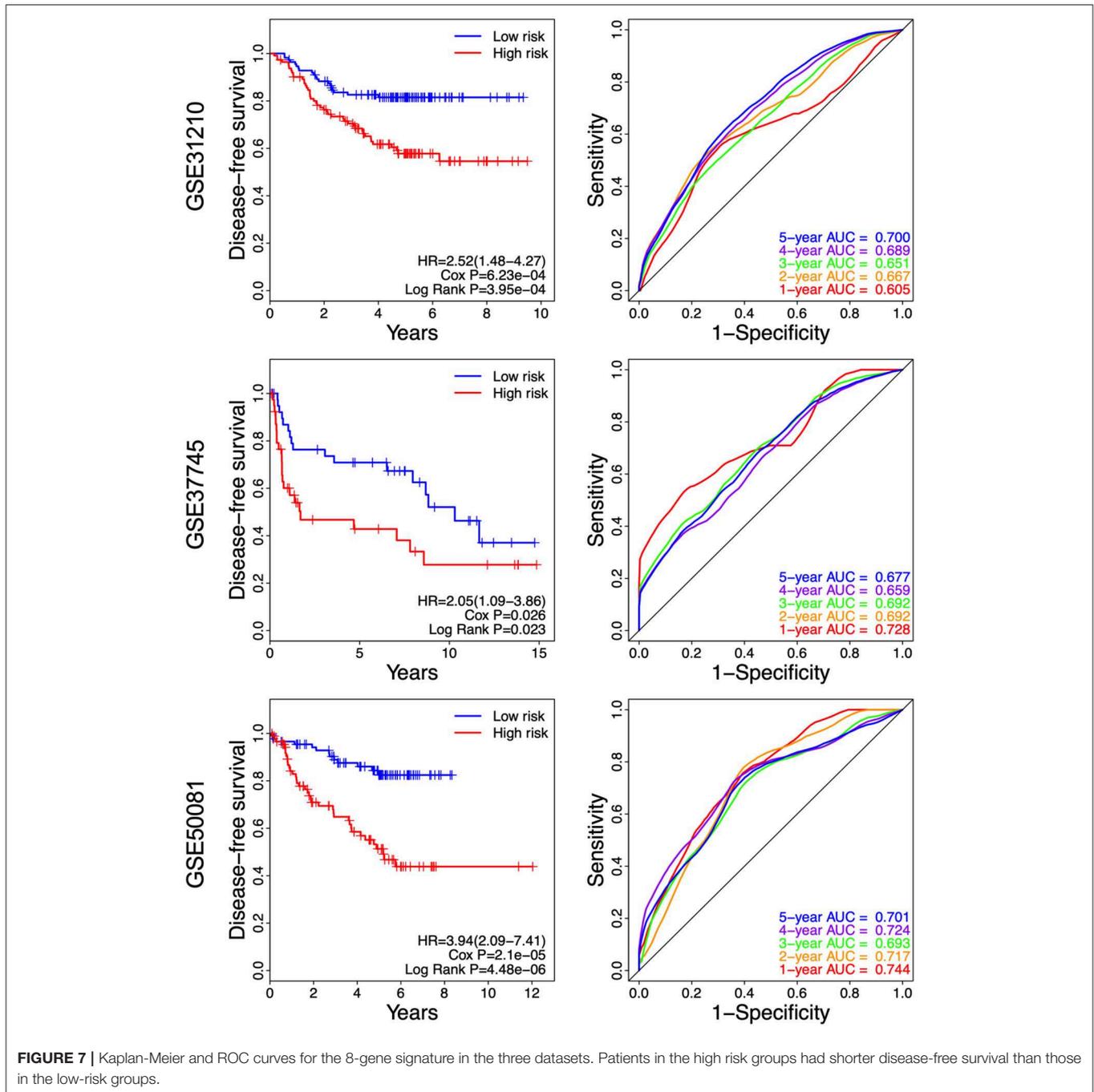
**FIGURE 5 |** Kaplan-Meier analysis of overall survival for patients stratified by age, gender, stage, and histological type.



transition biomarkers and poor prognosis in patients with lung adenocarcinoma, and SH2B1 has important roles on cell proliferation, migration, and invasion in A549 and H1299 cells (48). SH2B1 has been reported to be highly expressed in NSCLC tissues and cells, and SH2B1 high-expression has poor disease-free survival and overall survival (49). KBTBD7 has been found to be involved in inflammation and cardiac dysfunction, which is targeted by miR-21 (50). However, the roles of KBTBD7 and SEC24B-AS1 in cancer have not been investigated. Of note, the result of the AUC analyses in our study showed that the AUC values of the combination of 8 genes were more than 0.60 in both the overall survival and disease-free survival, suggesting that the combination of 8 genes could be regarded as a novel factor that may serve as a prognosis indicator for NSCLC patients. Stratification analysis indicated that the 8-gene signature predicted survival in most sub-groups and was independent of other clinical factors, such as age, gender, stage, and histology type. In our study, the 8-gene signature showed a great ability to stratify NSCLC patients into high- and low-risk

groups with significantly different overall survival. Thus, it could be an important asset in improving the prognosis and providing better prescriptions.

Currently, the tumor staging system remains the most powerful tool for survival prediction and treatment decision in NSCLC patients (51). Despite its great clinical value, its prognostic and predictive power is incompetent to guide patient management. In particular, the current staging system is far from accurate in predicting survival at the individual level, since half of the patients with early disease will eventually develop recurrent disease (51). This is directly linked to the decision of prescribing adjuvant chemotherapy after a pulmonary resection in early-stage NSCLC patients. Identifying early-stage patients with poor prognosis would consequently help specialists screen the appropriate candidates for adjuvant chemotherapy. Further development of genomic signatures is expected to assist patient stratification in clinical practice. In the present stratification analysis, the 8-gene signature showed prognostic value among stage IA, stage IB and stage II patients. It was



able to classify patients in the same stage into high- and low-risk groups with significantly different survival prospects, implying that the 8-gene signature can improve the accuracy of survival prediction. In addition, a prognostic model was developed by combining the stage with the 8-gene signature for survival prediction. These findings might help specialists select high-risk patients for adjuvant therapy in addition to surgery resection.

Significantly, in our study, univariate and multivariate Cox regression model suggested that stage IA/IB in GSE37745 was

significantly correlated with overall survival of the patients. Moreover, the patients in stage IB had worse overall survival than those in stage IA in the combined dataset. Strauss et al. have demonstrated that adjuvant chemotherapy is not standard care for stage IB NSCLC patients (52). However, another previous study has demonstrated that there is a remarkable survival improvement in stage IB NSCLC patients from Italy treated with adjuvant chemotherapy (53). These results that adjuvant chemotherapy is efficient for stage IB NSCLC patients with large tumors (51, 54).

These findings may have substantial clinical value for NSCLC. Remarkably, several limitations should be noted in our study. Firstly, data of ALK/EGFR/KRAS was only available in GSE31210, and there were no data of molecular status in the rest of the cohorts, thus, there was insufficient sample size to assess an association or not with the 8-gene signature. Secondly, our study was the retrospective nature of the research and had the heterogeneity of the techniques that have been used to analyze gene expression (Affymetrix U133 Plus 2.0 microarray platform and different Illumina sequencing platform). Thirdly, further studies should be carried out to determine the biological roles of these predictive mRNAs and lncRNAs relying on *in vitro* and *in vivo* data based on all kinds of experiment methods.

## CONCLUSIONS

A novel 8-gene signature for prognostic prediction in early-stage NSCLC patients was developed. The findings suggested that the 8-gene signature is a powerful predictor for overall survival of patients with early-stage NSCLC. Furthermore, the signature was independent of other clinical factors, such as stage. Additionally, a prognostic model combining the 8-gene signature with the stage was developed, which may conduce to treatment decisions

for individuals and hold promise for clinical practice in the near future.

## DATA AVAILABILITY

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

## AUTHOR CONTRIBUTIONS

RH: download the data and wrote the manuscript. SZ: conceived and designed the study, performed the analysis, and contributed to critical review of the manuscript. All authors read and approved the final manuscript.

## FUNDING

This work was supported by Natural Science Foundation of Henan Province (No. 162300410040), Outstanding Youth Science Foundation of Henan University (No. yqpy20140036), Science and Technology Development Program of Henan Province (No. 132300410274), and National Natural Science Foundation of China (No. 81301963).

## REFERENCES

- Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* (2018) 68:394–424. doi: 10.3322/caac.21492
- New M, Keith R. Early detection and chemoprevention of lung cancer. *F1000Res.* (2018) 7:61. doi: 10.12688/f1000research.12433.1
- Noone AM, Howlander N, Krapcho M, Miller D, Brest A, Yu M, et al. (eds.). *SEER Cancer Statistics Review, 1975–2015*. National Cancer Institute. Bethesda, MD. Available online at: [https://seer.cancer.gov/csr/1975\\_2015/](https://seer.cancer.gov/csr/1975_2015/)
- Arriagada R, Bergman B, Dunant A, Le Chevalier T, Pignon JP, Vansteenkiste J. Cisplatin-based adjuvant chemotherapy in patients with completely resected non-small-cell lung cancer. *N Engl J Med.* (2004) 350:351–60. doi: 10.1056/NEJMoa031644
- Douillard JY, Rosell R, De Lena M, Carpagnano F, Ramlau R, Gonzales-Larriba JL, et al. Adjuvant vinorelbine plus cisplatin versus observation in patients with completely resected stage IB–IIIA non-small-cell lung cancer (Adjuvant Navelbine International Trialist Association [ANITA]): a randomised controlled trial. *Lancet Oncol.* (2006) 7:719–27. doi: 10.1016/S1470-2045(06)70804-X
- Pisters KM, Evans WK, Azzoli CG, Kris MG, Smith CA, Desch CE, et al. Cancer care Ontario and American Society of Clinical Oncology adjuvant chemotherapy and adjuvant radiation therapy for stages I–IIIA resectable non small-cell lung cancer guideline. *J Clin Oncol.* (2007) 25:5506–18. doi: 10.1200/JCO.2007.14.1226
- Reck M, Heigener DF, Mok T, Soria JC, Rabe KF. Management of non-small-cell lung cancer: recent developments. *Lancet.* (2013) 382:709–19. doi: 10.1016/S0140-6736(13)61502-0
- Pignon JP, Tribodet H, Scagliotti GV, Douillard JY, Shepherd FA, Stephens RJ, et al. Lung adjuvant cisplatin evaluation: a pooled analysis by the LACE Collaborative Group. *J Clin Oncol.* (2008) 26:3552–9. doi: 10.1200/JCO.2007.13.9030
- Strauss GM. Management of early-stage lung cancer: past, present, and future adjuvant trials. *Oncology.* (2006) 20:1651–63; discussion 63–4, 66, 69–70 passim.
- Lau SK, Boutros PC, Pintilie M, Blackhall FH, Zhu CQ, Strumpf D, et al. Three-gene prognostic classifier for early-stage non small-cell lung cancer. *J Clin Oncol.* (2007) 25:5562–9. doi: 10.1200/JCO.2007.12.0352
- Moon H, Zhao Y, Pluta D, Ahn H. Subgroup analysis based on prognostic and predictive gene signatures for adjuvant chemotherapy in early-stage non-small-cell lung cancer patients. *J Biopharm Stat.* (2018) 28:750–62. doi: 10.1080/10543406.2017.1397006
- Boutros PC, Lau SK, Pintilie M, Liu N, Shepherd FA, Der SD, et al. Prognostic gene signatures for non-small-cell lung cancer. *Proc Natl Acad Sci USA.* (2009) 106:2824–8. doi: 10.1073/pnas.0809444106
- Roepman P, Jassem J, Smit EF, Muley T, Niklinski J, van de Velde T, et al. An immune response enriched 72-gene prognostic profile for early-stage non-small-cell lung cancer. *Clin Cancer Res.* (2009) 15:284–90. doi: 10.1158/1078-0432.CCR-08-1258
- Xie Y, Xiao G, Coombes KR, Behrens C, Solis LM, Raso G, et al. Robust gene expression signature from formalin-fixed paraffin-embedded samples predicts prognosis of non-small-cell lung cancer patients. *Clin Cancer Res.* (2011) 17:5705–14. doi: 10.1158/1078-0432.CCR-11-0196
- Lin T, Fu Y, Zhang X, Gu J, Ma X, Miao R, et al. A seven-long noncoding RNA signature predicts overall survival for patients with early stage non-small cell lung cancer. *Aging.* (2018) 10:2356–66. doi: 10.18632/aging.101550
- Ebata T, Hirata H, Kawauchi K. Functions of the tumor suppressors p53 and Rb in actin cytoskeleton remodeling. *BioMed Res Int.* (2016) 2016:1–10. doi: 10.1155/2016/9231057
- Li J, Chen Z, Tian L, Zhou C, He MY, Gao Y, et al. Original article: LncRNA profile study reveals a three-lncRNA signature associated with the survival of patients with oesophageal squamous cell carcinoma. *Gut.* (2014) 63:1700–10. doi: 10.1136/gutjnl-2013-305806
- Meng J, Li P, Zhang Q, Yang Z, Fu S. A four-long non-coding RNA signature in predicting breast cancer survival. *J Exp Clin Cancer Res.* (2014) 33:84. doi: 10.1186/s13046-014-0084-7
- Mercer TR, Dinger ME, Mattick JS. Long non-coding RNAs: insights into functions. *Nat Rev Genet.* (2009) 10:155–9. doi: 10.1038/nrg2521
- Qi X, Zhang DH, Wu N, Xiao JH, Wang X, Ma W. ceRNA in cancer: possible functions and clinical implications. *J Med Genet.* (2015) 52:710–8. doi: 10.1136/jmedgenet-2015-103334
- Cabili MN, Trapnell C, Goff L, Koziol M, Tazonvega B, Regev A, et al. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* (2011) 25:1915–27. doi: 10.1101/gad.17446611
- Gibb EA, Brown CJ, Lam WL. The functional role of long non-coding RNA in human carcinomas. *Mol Cancer.* (2011) 10:38. doi: 10.1186/1476-4598-10-38

23. Tian X, Zhu X, Yan T, Yu C, Shen C, Hong J, et al. Differentially expressed lncRNAs in gastric cancer patients: a potential biomarker for gastric cancer prognosis. *J Cancer*. (2017) 8:2575–86. doi: 10.7150/jca.19980
24. Okayama H, Kohno T, Ishii Y, Shimada Y, Shiraishi K, Iwakawa R, et al. Identification of genes upregulated in ALK-positive and EGFR/KRAS/ALK-negative lung adenocarcinomas. *Cancer Res*. (2012) 72:100–11. doi: 10.1158/0008-5472.CAN-11-1403
25. Yamauchi M, Yamaguchi R, Nakata A, Kohno T, Nagasaki M, Shimamura T, et al. Epidermal growth factor receptor tyrosine kinase defines critical prognostic genes of stage I lung adenocarcinoma. *PLoS ONE*. (2012) 7:e43923. doi: 10.1371/journal.pone.0043923
26. Botling J, Edlund K, Lohr M, Hellwig B, Holmberg L, Lambe M, et al. Biomarker discovery in non-small cell lung cancer: integrating gene expression profiling, meta-analysis, and tissue microarray validation. *Clin Cancer Res*. (2013) 19:194–204. doi: 10.1158/1078-0432.CCR-12-1139
27. Der SD, Sykes J, Pintilie M, Zhu CQ, Strumpf D, Liu N, et al. Validation of a histology-independent prognostic gene signature for early-stage, non-small-cell lung cancer including stage IA patients. *J Thorac Oncol*. (2014) 9:59–64. doi: 10.1097/JTO.0000000000000042
28. Ein-Dor L, Zuk O, Domany E. Thousands of samples are needed to generate a robust gene list for predicting outcome in cancer. *Proc Natl Acad Sci USA*. (2006) 103:5923–8. doi: 10.1073/pnas.0601231103
29. Vlad C, Kubelac P, Onisim A, Fetica B, Fulop A, Irimie A, et al. Expression of CDCP1 and ADAM12 in the ovarian cancer microenvironment. *J BUON*. (2016) 21:973–8.
30. Uekita T, Fujii S, Miyazawa Y, Iwakawa R, Narisawasaito M, Nakashima K, et al. Oncogenic Ras/ERK signaling activates CDCP1 to promote tumor invasion and metastasis. *Mol Cancer Res*. (2014) 12:1449–59. doi: 10.1158/1541-7786.MCR-13-0587
31. Zeng XJ, Wu YH, Luo M, Cong PG, Yu H. Inhibition of pulmonary carcinoma proliferation or metastasis of miR-218 via down-regulating CDCP1 expression. *Eur Rev Med Pharmacol Sci*. (2017) 21:1502–8.
32. Miyazawa Y, Uekita T, Hiraoka N, Fujii S, Kosuge T, Kanai Y, et al. CUB domain-containing protein 1, a prognostic factor for human pancreatic cancers, promotes cell migration and extracellular matrix degradation. *Cancer Res*. (2010) 70:5136–46. doi: 10.1158/0008-5472.CAN-10-0220
33. Ikeda JI, Oda T, Inoue M, Uekita T, Sakai R, Okumura M, et al. Expression of CUB domain containing protein (CDCP1) is correlated with prognosis and survival of patients with adenocarcinoma of lung. *Cancer Sci*. (2010) 100:429–33. doi: 10.1111/j.1349-7006.2008.01066.x
34. Akiyama Y, Jung S, Sallia B, Lee S, Hubbard S, Taylor M, et al. Hyaluronate receptors mediating glioma cell migration and proliferation. *J Neuro Oncol*. (2001) 53:115–27. doi: 10.1023/A:1012297132047
35. Maxwell CA, Keats JJ, Belch AR, Pilarski LM, Reiman T. Receptor for hyaluronan-mediated motility correlates with centrosome abnormalities in multiple myeloma and maintains mitotic integrity. *Cancer Res*. (2005) 65:850–60.
36. Giannopoulos K, Mertens D, Bühler A, Barth TFE, Idler I, Möller P, et al. The candidate immunotherapeutic target, the receptor for hyaluronic acid-mediated motility, is associated with proliferation and shows prognostic value in B-cell chronic lymphocytic leukemia. *Leukemia*. (2009) 23:519–27. doi: 10.1038/leu.2008.338
37. Mandana V, Kwon DH, Borowsky AD, Cornelia T, Leong HS, Lewis JD, et al. Cellular heterogeneity profiling by hyaluronan probes reveals an invasive but slow-growing breast tumor subset. *Proc Natl Acad Sci USA*. (2014) 111:E1731. doi: 10.1073/pnas.1402383111
38. Koelzer VH, Huber B, Mele V, Iezzi F, Trippel M, Karamitopoulou E, et al. Expression of the hyaluronan acid mediated motility receptor RHAMM in tumor budding cells identifies aggressive colorectal cancers. *Hum Pathol*. (2015) 46:1573–81. doi: 10.1016/j.humpath.2015.07.010
39. Liang B, Jia C, Huang Y, He H, Li J, Liao H, et al. TPX2 Level correlates with hepatocellular carcinoma cell proliferation, apoptosis, and EMT. *Dig Dis Sci*. (2015) 60:2360–72. doi: 10.1007/s10620-015-3730-9
40. Zhang MY, Liu XX, Li H, Li R, Liu X, Qu YQ. Elevated mRNA levels of AURKA, CDC20 and TPX2 are associated with poor prognosis of smoking related lung adenocarcinoma using bioinformatics analysis. *Int J Med Sci*. (2018) 15:1676–85. doi: 10.7150/ijms.28728
41. Hsu PK, Chen HY, Yeh YC, Yen CC, Wu YC, Hsu CP, et al. TPX2 expression is associated with cell proliferation and patient outcome in esophageal squamous cell carcinoma. *J Gastroenterol*. (2014) 49:1231–40. doi: 10.1007/s00535-013-0870-6
42. Yan Y, Jing Z, Yuegang W. Targeting of TRX2 by miR-330-3p in melanoma inhibits proliferation. *Biomed Pharmacother*. (2018) 107:1020–9. doi: 10.1016/j.biopha.2018.08.058
43. Chen M, Zhang H, Zhang G, Zhong A, Ma Q, Kai J, et al. Targeting TPX2 suppresses proliferation and promotes apoptosis via repression of the PI3k/AKT/P21 signaling pathway and activation of p53 pathway in breast cancer. *Biochem Biophys Res Commun*. (2018) 507:74–82. doi: 10.1016/j.bbrc.2018.10.164
44. Jang HH, Lee HN, Kim SY, Hong S, Lee WS. Expression of RNA-binding motif protein 3 (RBM3) and cold-inducible RNA-binding protein (CIRP) is associated with improved clinical outcome in patients with colon cancer. *Anticancer Res*. (2017) 37:1779–85. doi: 10.21873/anticancer.11511
45. Hamid AA, Masaki M, Jun F, Kanako N, Masatoshi K, Takashi K, et al. Expression of cold-inducible RNA-binding protein in the normal endometrium, endometrial hyperplasia, and endometrial carcinoma. *International J Gynecol Pathol*. (2003) 22:240–7. doi: 10.1097/01.PGP.0000070851.25718.EC
46. Waters KM, Sontag RL, Weber TJ. Hepatic leukemia factor promotes resistance to cell death: implications for therapeutics and chronotherapy. *Toxicol Appl Pharmacol*. (2013) 268:141–8. doi: 10.1016/j.taap.2013.01.031
47. Chen S, Wang Y, Ni C, Meng G, Sheng X. HLF/miR-132/TTK axis regulates cell proliferation, metastasis and radiosensitivity of glioma cells. *Biomed Pharmacother*. (2016) 83:898–904. doi: 10.1016/j.biopha.2016.08.004
48. Wang SO, Cheng Y, Gao Y, He Z, Zhou W, Chang R, et al. SH2B1 promotes epithelial-mesenchymal transition through the IRS1/PI3K-catenin signaling axis in lung adenocarcinoma. *Mol Carcinogenesis*. (2018) 57:640–52. doi: 10.1002/mc.22788
49. Zhang H, Duan CJ, Chen W, Wang SQ, Zhang SK, Dong S, et al. Clinical significance of SH2B1 adaptor protein expression in non-small cell lung cancer. *Asian Pacific J Cancer Prev*. (2012) 13:2355–62. doi: 10.7314/APJCP.2012.13.5.2355
50. Yang L, Bo W, Zhou Q, Wang Y, Liu X, Liu Z, et al. MicroRNA-21 prevents excessive inflammation and cardiac dysfunction after myocardial infarction through targeting KBTBD7. *Cell Death Dis*. (2018) 9:769. doi: 10.1038/s41419-018-0805-5
51. Goldstraw P, Crowley J, Chansky K, Giroux DJ, Groome PA, Rami-Porta R, et al. 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. doi: 10.1097/JTO.0b013e31812f3c1a
52. Strauss GM, Herndon JE II, Maddaus MA, Johnstone DW, Johnson EA, Harpole DH, et al. Adjuvant paclitaxel plus carboplatin compared with observation in stage IB non-small-cell lung cancer: CALGB 9633 with the Cancer and Leukemia Group B, Radiation Therapy Oncology Group, and North Central Cancer Treatment Group Study Groups. *J Clin Oncol*. (2008) 26:5043–51. doi: 10.1200/JCO.2008.16.4855
53. Mario R, Sabrina M, Patrizia F, Anastasia L, Davide M, Eugenio P, et al. Postsurgical chemotherapy in stage IB nonsmall cell lung cancer: long-term survival in a randomized study. *Int J Cancer J Int Du Cancer*. (2010) 119:955–60. doi: 10.1002/ijc.21933
54. López-Encuentra Á, Duque-Medina JL, Rami-Porta R, Cámara AGDL, Ferrando P. Staging in lung cancer: is 3 cm a prognostic threshold in pathologic stage I non-small cell lung cancer? : a multicenter study of 1,020 patients. *Chest*. (2002) 121:1515–20. doi: 10.1378/chest.12.1.1515

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 He and Zuo. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.