



## OPEN ACCESS

EDITED BY  
Ángeles Carlos Reyes,  
Instituto Nacional de Enfermedades  
Respiratorias-México (INER), Mexico

REVIEWED BY  
Weimin Zhong,  
Xiamen Fifth Hospital, China  
Desi Shang,  
Harbin Medical University, China

\*CORRESPONDENCE  
Juan-Juan Fu,  
fu.juanjuan@scu.edu.cn

<sup>†</sup>These authors have contributed equally  
to this work and share first authorship

## SPECIALTY SECTION

This article was submitted to Cancer  
Genetics and Oncogenomics,  
a section of the journal  
Frontiers in Genetics

RECEIVED 28 April 2022

ACCEPTED 29 July 2022

PUBLISHED 02 September 2022

## CITATION

Miao T-W, Chen F-Y, Du L-Y, Xiao W and  
Fu J-J (2022), Signature based on RNA-  
binding protein-related genes for  
predicting prognosis and guiding  
therapy in non-small cell lung cancer.  
*Front. Genet.* 13:930826.  
doi: 10.3389/fgene.2022.930826

## COPYRIGHT

© 2022 Miao, Chen, Du, Xiao and Fu.  
This is an open-access article  
distributed under the terms of the  
[Creative Commons Attribution License  
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). 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.

# Signature based on RNA-binding protein-related genes for predicting prognosis and guiding therapy in non-small cell lung cancer

Ti-Wei Miao<sup>1†</sup>, Fang-Ying Chen<sup>2†</sup>, Long-Yi Du<sup>1</sup>, Wei Xiao<sup>1</sup> and Juan-Juan Fu<sup>1\*</sup>

<sup>1</sup>Department of Integrated Traditional Chinese and Western Medicine, West China Hospital, Sichuan University, Chengdu, China, <sup>2</sup>West China School of Medicine, West China Hospital, Sichuan University, Chengdu, China

**Background:** Studies have reported that RNA-binding proteins (RBPs) are dysregulated in multiple cancers and are correlated with the progression and prognosis of disease. However, the functions of RBPs in non-small cell lung cancer (NSCLC) remain unclear. The present study aimed to explore the function of RBPs in NSCLC and their prognostic and therapeutic value.

**Methods:** The mRNA expression profiles, DNA methylation data, gene mutation data, copy number variation data, and corresponding clinical information on NSCLC were downloaded from The Cancer Genome Atlas, Gene Expression Omnibus, and the University of California Santa Cruz Xena databases. The differentially expressed RBPs were identified between tumor and control tissues, and the expression and prognostic value of these RBPs were systemically investigated by bioinformatics analysis. A quantitative polymerase chain reaction (qPCR) was performed to validate the dysregulated genes in the prognostic signature.

**Results:** A prognostic RBP-related signature was successfully constructed based on eight RBPs represented as a risk score using least absolute shrinkage and selection operator (LASSO) regression analysis. The high-risk group had a worse overall survival (OS) probability than the low-risk group ( $p < 0.001$ ) with 1-, 3-, and 5-year area under the receiver operator characteristic curve values of 0.671, 0.638, and 0.637, respectively. The risk score was associated with the stage of disease ( $p < 0.05$ ) and was an independent prognostic factor for NSCLC when adjusted for age and UICC stage ( $p < 0.001$ , hazard ratio (HR): 1.888). The constructed nomogram showed a good predictive value. The *P53*, focal adhesion, and NOD-like receptor signaling pathways were the primary pathways in the high-risk group (adjusted  $p$  value  $< 0.05$ ). The high-risk group was correlated with increased immune infiltration ( $p < 0.05$ ), upregulated relative expression levels of programmed cell death 1 (*PD1*) ( $p = 0.015$ ), cytotoxic T-lymphocyte-associated protein 4 (*CTLA4*) ( $p = 0.042$ ), higher gene mutation frequency, higher tumor mutational burden ( $p = 0.034$ ), and better chemotherapy response ( $p < 0.001$ ). The

signature was successfully validated using the GSE26939, GSE31210, GSE30219, and GSE157009 datasets. Dysregulation of these genes in patients with NSCLC was confirmed using the qPCR in an independent cohort ( $p < 0.05$ ).

**Conclusion:** An RBP-related signature was successfully constructed to predict prognosis in NSCLC, functioning as a reference for individualized therapy, including immunotherapy and chemotherapy.

#### KEYWORDS

RNA-binding proteins, non-small cell lung cancer, prognosis, immune infiltration, therapy, bioinformatics

## Introduction

Non-small cell lung cancer (NSCLC) is the leading cause of cancer-related death worldwide with an incidence rate of 1.3 million cases per year (Bray et al., 2018). NSCLC remains asymptomatic during the early stage, and only 25% of patients with NSCLC are diagnosed at this stage (Ni et al., 2018). However, approximately 80% of patients are diagnosed in metastatic stages with a 5-year survival rate below 15% (Goldstraw et al., 2016). Despite constant progress in novel therapies for NSCLC, including targeted therapy, immunotherapy, and chemotherapy, the therapeutic efficacy is still unsatisfactory (Facchinetti et al., 2016). Thus, the identification of effective biomarkers to accurately predict the overall survival (OS) probability and guide therapy in NSCLC is of great importance.

RNA-binding proteins (RBPs) are a group of proteins that function with an RNA-binding domain to distinguish and bind to target RNAs, including coding RNAs and non-coding RNAs (Hong, 2017). To date, more than 1,500 RBPs have been identified in the human genome through high-throughput screening (Cook et al., 2011; Gerstberger et al., 2014). RBPs modulate the fate of binding RNAs by regulating transcription, editing, splicing, polyadenylation, translocation, and turnover (Müller-McNicoll and Neugebauer, 2013). In recent years, studies in genetics and proteomics have shown that most RBPs exhibit functional abnormalities in lung cancer. Quaking I-5 (*QKI-5*), RALY heterogeneous nuclear ribonucleoprotein (*RALY*), and KH-type splicing regulatory protein (*KHSRP*) promote cancer cell proliferation and invasion, and they are associated with OS probability in NSCLC (Yan et al., 2019; Liang et al., 2020a; Song et al., 2020). Musashi1 (*MSI1*) promotes NSCLC malignancy and chemoresistance (Lang et al., 2017). RNA-binding motif protein 47 (*RBM47*) inhibits NSCLC metastasis through modulation of *AXIN1* mRNA stability and Wnt/ $\beta$ -catenin signaling (Shen et al., 2020). However, the molecular mechanism of RBPs and their prognostic predictive ability in NSCLC remain unknown. A systematic analysis of RBPs has reported novel and comprehensive insights into the underlying mechanism during cancer progression, and

prognostic signatures using RBPs have been constructed in multiple tumors (Li et al., 2020a; Li et al., 2020b; Li et al., 2020c; Kang et al., 2020; Wang et al., 2020). However, a prognostic RBP-related signature in NSCLC has not been reported. Thus, the present study aimed to explore the function of RBPs in NSCLC and their prognostic and therapeutic value.

In the present study, mRNA expression profiles, DNA methylation data, gene mutation data, copy number variation data, and clinical information on NSCLC were obtained from The Cancer Genome Atlas (TCGA, <https://portal.gdc.cancer.gov/repository>), Gene Expression Omnibus (GEO, <https://www.ncbi.nlm.nih.gov/geo/>), and the University of California Santa Cruz (UCSC) Xena (<https://xenabrowser.net/>) databases. The differentially expressed RBPs (DERBPs) were screened and then applied to perform functional enrichment analysis and to construct a prognostic RBP-related signature. A nomogram was created, and Kaplan–Meier survival analysis and Cox regression analysis were performed to explore its prognostic value. The underlying molecular mechanisms between the different risk groups were investigated using gene set enrichment analysis (GSEA). The correlations of risk score with clinical characteristics, DNA methylation levels, tumor mutational burden (TMB), immune infiltration, and chemotherapy sensitivity were analyzed using R software packages. Finally, dysregulated expression levels of these genes were validated using the quantitative polymerase chain reaction (qPCR).

## Materials and methods

### Data preprocessing and identification of DERBPs

The following data were downloaded from TCGA database: mRNA expression profiles of 1,037 NSCLC samples and 108 control samples; DNA methylation data on 807 NSCLC samples and 71 control samples; gene mutation data on 1,059 NSCLC samples; and clinical information (age, gender,

TABLE 1 Clinical features of patients with NSCLC from TCGA and GEO databases.

Clinical characteristic	TCGA cohort (978)	GSE31210 (226)	GSE26939 (114)	GSE30219 (264)	GSE157009 (248)
Age (years)					
≥65	590 (60.33%)	62 (0.27%)	61 (53.51%)	104 (39.39%)	182 (73.39%)
<65	388 (39.67%)	164 (72.57%)	53(46.49%)	159 (60.23%)	66 (26.61%)
Unknown	0	0	0	1 (0.38%)	0
Gender					
Male	585 (59.82%)	105 (46.46%)	52 (46.61%)	223 (84.47%)	160 (64.52%)
Female	393 (40.185)	121 (53.54)	62 (54.39%)	41 (15.53%)	88 (35.48%)
Unknown	0	0	0	0	0
T classification					
T1–T2	819 (83.74%)	226 (100%)	—	218 (82.58%)	230 (92.74%)
T3–T4	156 (15.95%)	0	—	44 (16.67%)	18 (7.26%)
Unknown	3 (0.31%)	0	—	2 (0.76%)	0
N classification					
N0	629 (64.31%)	—	—	188 (71.21%)	—
N1–N3	333 (34.05%)	—	—	74 (28.03)	—
Unknown	16 (1.64%)	—	—	2 (0.76%)	—
M classification					
M0	722 (73.82%)	—	—	257 (97.35%)	—
M1	31 (3.17%)	—	—	4 (1.52)	—
Unknown	225 (23.01%)	—	—	3 (1.14%)	—
UICC stage					
Stage I–II	772 (78.94%)	226 (100%)	81 (71.05%)	208 (78.79%)	—
Stage III–IV	194 (19.84%)	0	20 (17.54%)	52 (19.70%)	—
Unknown	12 (1.23%)	0	13 (11.40%)	4 (1.52%)	—

TCGA, The Cancer Genome Atlas; GEO, Gene Expression Omnibus; NSCLC, non-small cell lung cancer; UICC, Union for International Cancer Control.

smoking history, Union for International Cancer Control (UICC) stage, survival time, and survival status) on 1,027 tumor samples. Moreover, the GSE31210, GSE26939, GSE30219, and GSE157009 datasets were obtained from the GEO database. The GSE31210, GSE30219, and GSE157009 datasets were generated using the GPL570 [HG-U133\_Plus\_2] Affymetrix Human Genome U133 Plus 2.0 Array, whereas the GSE26939 dataset was generated using the GPL9053 Agilent-UNC-custom-4X44K. The GSE31210 dataset included mRNA expression profiles of 226 NSCLC samples and 20 control samples, as well as clinical information on 226 NSCLC samples. The GSE26939 dataset included mRNA expression profiles of 116 NSCLC samples and 0 control samples, as well as clinical information on 116 NSCLC samples. The GSE30219 dataset included mRNA expression profiles of 293 NSCLC samples and 14 control samples, as well as clinical information on 293 NSCLC samples. The GSE157009 dataset included mRNA expression profiles of 249 NSCLC samples and 0 control samples, as well as clinical information on 249 NSCLC samples. Copy number

variation data on 1079 NSCLC samples were downloaded from the UCSC Xena database. The inclusion criteria were as follows: 1) NSCLC; 2) mRNA expression profiles, and 3) complete follow-up data. Therefore, 978 patients from TCGA, 226 patients from the GSE31210 dataset, 114 patients from the GSE26939 dataset, 264 patients from the GSE30219 dataset, and 248 patients from the GSE157009 dataset were enrolled in the present study. The baseline characteristics of the patients are shown in Table 1. A total of 1,542 RBPs were included in the present study (Cook et al., 2011; Gerstberger et al., 2014). DERBPs were identified based on  $|\log_2 \text{fold change (FC)}| \geq 0.7$  and adjusted  $p$  value  $< 0.05$  using the “limma” package in R (Version 4.0.2) when two groups were compared. The “ggplot2” and “gplots” packages in R were used to generate the volcano plots and heatmap.

## Functional enrichment analyses

Functional enrichment analyses of the DERBPs were performed by Gene Ontology (GO) and Kyoto Encyclopedia

of Genes and Genomes (KEGG) pathway analyses using the “clusterProfiler,” “org.Hs.eg.db,” “enrichplot,” “ggplot2,” and “GOplot” packages in R. GO analysis included biological processes (BPs), cellular components (CCs), and molecular functions (MFs). The adjusted  $p$  value  $<0.05$  was considered statistically different.

## Construction and validation of the prognostic RBP-related signature

The prognostic RBPs were screened using DERBPs through univariate Cox regression analysis, and  $p < 0.05$  was selected as the statistical threshold. The least absolute shrinkage and selection operator (LASSO) regression analysis was applied to identify hub DERBPs to minimize the risk of overfitting among the signatures. The changing trajectory of each independent variable was first analyzed, and fivefold cross-validation was used to build a model and analyze the confidence interval under each lambda value. The following formula was utilized: risk score = expression for each gene  $\times$  coefficient for each gene. The patients were divided into low- and high-risk groups, according to the median risk score. A Kaplan–Meier survival curve was constructed between the two risk groups and was compared using the log-rank test. A receiver operating characteristic (ROC) curve was used to assess the predictive value of the Kaplan–Meier survival curves, and 1-, 3-, and 5-year area under the curve (AUC) values were calculated. The prognostic signature was validated using the following independent cohorts: GSE31210, GSE26939, GSE30219, and GSE157009.

## Correlation of the risk score with clinical characteristics

The risk score was compared in different age groups ( $\geq 65$  and  $<65$  years), genders (female and male), UICC stages (stage I–II and stage III–IV), T stages (T1–2 and T3–4), N stages (N0 and N1–3), and M stages (M0 and M1) by the Mann–Whitney test in TCGA cohort. Univariate and multivariable Cox regression analyses were performed to screen independent prognostic factors for NSCLC in TCGA, GSE31210, and GSE26939 datasets by R software.

## Nomogram and calibration plots

A nomogram was applied to forecast the likelihood of OS probability using independent prognostic factors for NSCLC through the “rms” package in R. Calibration plots of the

nomogram were generated to evaluate the conformity of the nomogram predicted and actual OS probability.

## Gene set enrichment analysis

GSEA is a computational method that determines whether a previously defined set of genes shows a significant difference between two biological states (Subramanian et al., 2005). The “c2.cp.kegg.v7.5.1.symbols.gmt” file was downloaded from the GSEA database (<http://www.gsea-msigdb.org/gsea/index.jsp>). The “limma,” “GSEABase,” “GSVA,” and “pheatmap” packages in R were applied to perform KEGG pathway analysis between the two risk groups. The adjusted  $p$  value  $<0.05$  was considered statistically significant.

## DNA methylation, TMB, and copy number variation analyses

DNA methylation data on NSCLC were obtained by Strawberry Perl (5.32.1.1–64-bit). DNA methylation levels of prognostic genes were extracted using the “limma” package in R and compared between the two risk groups by the Mann–Whitney test. Gene mutation data on lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC) were allocated in different “maf” files in TCGA database, and the gene mutation frequencies in the two risk groups in LUAD and LUSC were evaluated using the “maftools” package in R. The correlation of risk score with TMB was evaluated by the Mann–Whitney test and Spearman’s rank correlation analysis. Kaplan–Meier survival curve analysis was performed on the high-risk score + high TMB group and the low-risk score + low TMB group, and the curves were compared by the log-rank test by R software. Copy number variation matrixes of NSCLC were obtained by Strawberry Perl. The copy number variation levels and positions of prognostic genes were analyzed using the “RCircos” package in R.

## Immune infiltration analysis

The mRNA expression matrix of NSCLC was converted into a tumor microenvironment (TME) score matrix using the “limma” and “estimate” packages in R. TME scores, including immune score, stromal score, and estimate score, were compared between the two risk groups using the “reshape2” and “ggpubr” packages in R. Immune infiltration profiles were compared between the two risk groups and visualized by a violin plot using the “vioplot” package in R. The correlations of immune cells with risk scores were evaluated using the “limma,” “reshape2,” “tidyverse,” “ggplot2,” “ggpubr,” and “ggExtra” packages in R. Comparisons were performed by Spearman’s

rank correlation analysis based on  $p$ -value  $< 0.05$  and  $|r| > 0.1$ . Relative expression levels of immune checkpoint inhibitors (programmed cell death 1 (*PDI*), programmed cell death ligand 1 (*PDL1*), and cytotoxic T-lymphocyte-associated protein 4 (*CTLA4*)) were compared between the two risk groups.

## Sensitivity of chemotherapy drugs

Nine chemotherapy drugs, namely, axitinib (Solomon et al., 2022), dasatinib (Kim et al., 2021), docetaxel (Xiao et al., 2022), erlotinib (Wang et al., 2022), gemcitabine (Guo et al., 2022), metformin (Arrieta et al., 2022), paclitaxel (Saito et al., 2022), parthenolide (Li et al., 2020d; Sun et al., 2020), and shikonin (Pan et al., 2021), were screened from the previous literature that demonstrated their antitumor efforts on lung cancer, and they were selected for the present study. The half inhibitory concentration (IC<sub>50</sub>) of the chemotherapy drugs was compared between the two risk groups in NSCLC using the “pRRophetic” package in R. A  $p$ -value  $< 0.05$  was considered statistically significant.

## RNA extraction and qPCR validation

Lung tissue samples from patients with NSCLC were obtained from the West China Hospital of Sichuan University. Histologically normal tissues were used as controls. Total RNA was extracted from the lung tissues (14 control and 13 NSCLC samples) using the E. Z.N.A. HP Total RNA Kit (OMEGA, United States), according to the manufacturer’s protocol. Complementary DNA (cDNA) was synthesized using the PrimeScript™ RT reagent Kit (TaKaRa, Japan), following the manufacturer’s instructions. Quantitative PCR was performed in triplicate using the Iq™ SYBR Green SuperMix (BIO-RAD, United States), according to the manufacturer’s protocol. The relative gene expression levels were normalized via the  $\beta$ -actin Ct value, applying the  $2^{-\Delta\Delta Ct}$  relative quantification method. The following qPCR primers were used:

*ZC3H12C*-forward, 5'-GGCTTTTGAGTCGGACGGTA-3'; *ZC3H12C*-reverse, 5'-TCAGGGGGCATGAACCTTGTC-3'; *SMAD9*-forward, 5'-GTTTGTACGAGGAGCCCA-3'; *SMAD9*-reverse, 5'-AGGGTCGGTGAACCCATCTA-3'; *MRPL15*-forward, 5'-GAGAGGTGTGACCATCCAGC-3'; *MRPL15*-reverse, 5'-TTGGAATGGGTTGTCCACGAA-3'; *MBNL2*-forward, 5'-ATACGGCAGACGGCTTTCAG-3'; *MBNL2*-reverse, 5'-CTCTGCCTGTCCTTCCATT-3'; *FASTKD3*-forward, 5'-GATGGAAACCTGCCTGACA-3'; *FASTKD3*-reverse, 5'-CCAGGTTTCAGCAACAGGTA-3'; *SNRPB*-forward, 5'-AAGGGAAGAGAAGCGAGTCC-3'; *SNRPB*-reverse, 5'-GCAAGTGGAACTCGAGCAAT-3';

*IGF2BP1*-forward: 5'-TAGCTCCTTTATGCAGGCTCC-3'; *IGF2BP1*-reverse, 5'-CGGGAGAGCTGTTTGATGTG-3'; *INTS7*-forward, 5'-CACTATCAGGGACCATCGCC-3';

*INTS7*-reverse, 5'-GGTAACAGCACTCTTGGGCT-3';  $\beta$ -actin-forward, 5'-CCACGAACTACCTTCAACTCC-3';  $\beta$ -actin-reverse, 5'-GTGATCTCCTTCTGCATCCTGT-3'.

## Statistical analysis

Statistical analysis was performed by R (Version 4.0.2) and GraphPad Prism (Version 7.00) software. Levels of mRNA expression were expressed as the median (interquartile range), according to the data distribution type. Comparisons between the two groups were determined by the Mann-Whitney test for nonparametric data. Survival curves were compared by the log-rank test. Correlation analysis was performed by Spearman’s rank correlation analysis.  $p < 0.05$  was considered statistically significant.

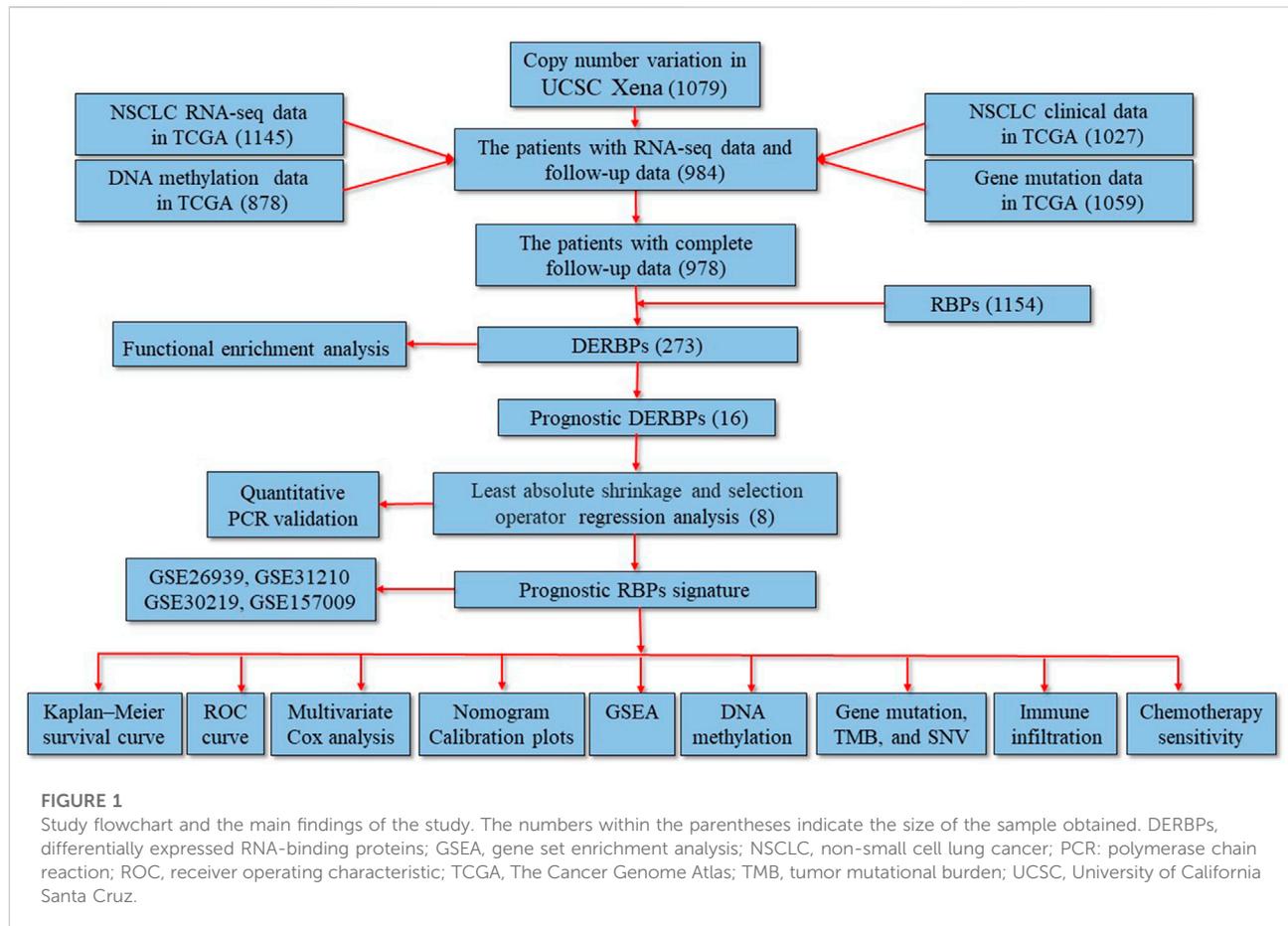
## Results

### Identification of DERBPs and functional enrichment analysis

The study flowchart is shown in Figure 1. A total of 273 DERBPs (173 upregulated RBPs and 100 downregulated RBPs) were identified when NSCLC samples were compared to control samples (Figures 2A,B). GO analysis using 273 DERBPs identified the following enriched terms: BPs, including the ncRNA metabolic process, ncRNA processing, and regulation of the mRNA metabolic process (adjusted  $p$  value  $< 0.05$ , Figure 2C); CCs, including those located in cytoplasmic ribonucleoprotein granules, ribonucleoprotein granules, and ribosomes (adjusted  $p$  value  $< 0.05$ , Figure 2C); and MFs, including the catalytic activity, acting on RNA, mRNA 3'-UTR binding, and single-stranded RNA binding (adjusted  $p$  value  $< 0.05$ , Figure 2C). KEGG pathway analysis identified ribosome biogenesis in eukaryotes, the mRNA surveillance pathway, and influenza A as the primary pathways in NSCLC (adjusted  $p$  value  $< 0.05$ , Figure 2D).

### Construction and validation of the prognostic RBP-related signature

A total of 16 prognostic RBPs were screened from 273 DERBPs using univariate Cox regression analysis ( $p < 0.05$ , Figure 3A). An eight-RBP signature (zinc finger CCCH-type containing 12C (*ZC3H12C*), mitochondrial ribosomal protein L15 (*MRPL15*), muscleblind-like splicing regulator 2 (*MBNL2*), SMAD family member 9 (*SMAD9*), insulin-like growth factor 2 mRNA binding protein 1 (*IGF2BP1*), FAST kinase domains 3 (*FASTKD3*), integrator complex subunit 7 (*INTS7*), and small nuclear ribonucleoprotein polypeptides B



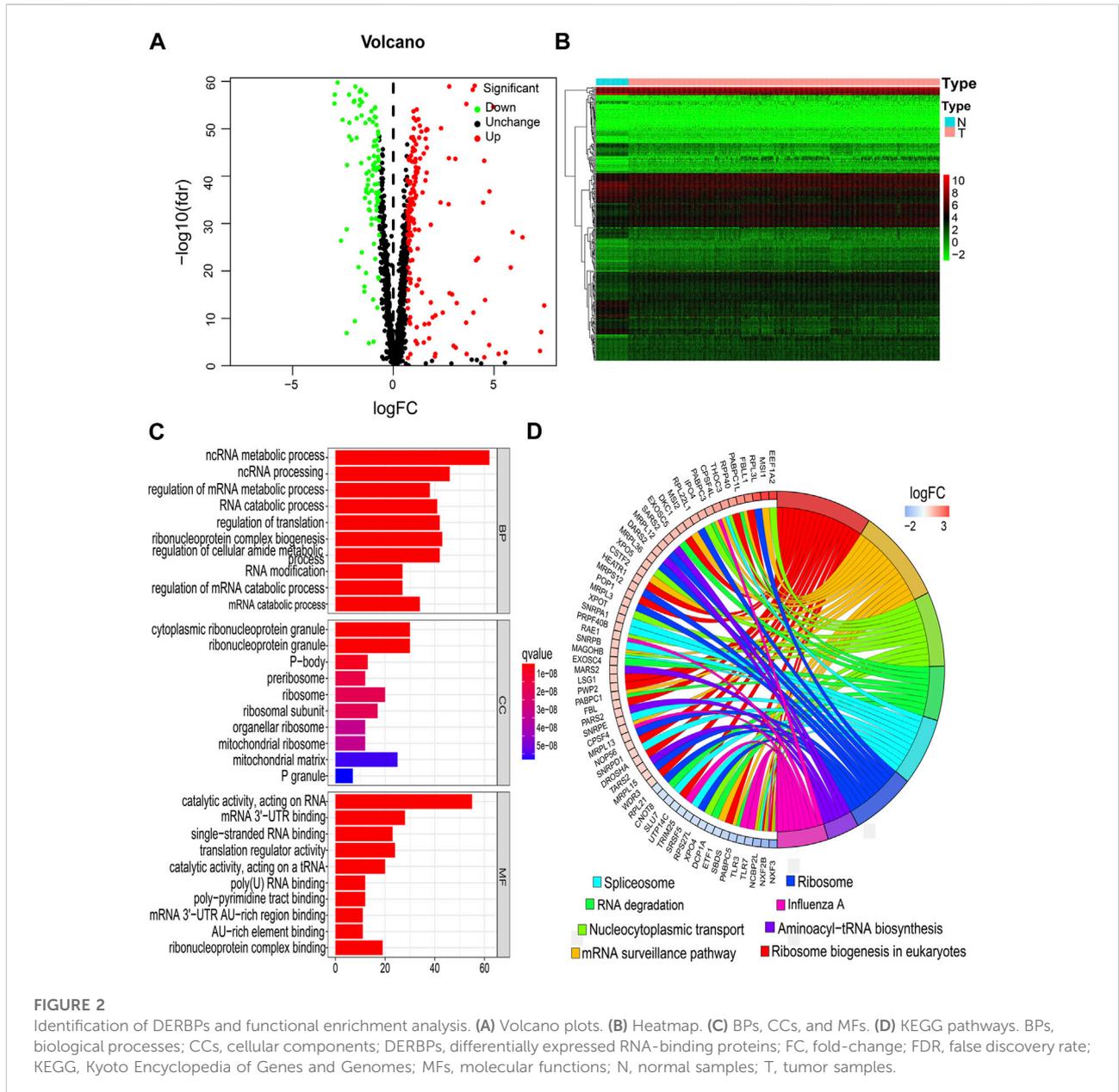
and B1 (*SNRPB*)) were generated utilizing LASSO regression analysis (Figures 3B–D). The formula for calculating the risk score was as follows: Risk score = (0.20 \* *ZC3H12Cexp*) + (0.31 \* *MRPL15exp*) + (0.45 \* *MBNL2exp*) + (-0.27 \* *SMAD9exp*) + (0.15 \* *IGF2BP1exp*) + (-0.37 \* *FASTKD3exp*) + (0.22 \* *INTS7exp*) + (0.34 \* *SNRPBexp*).

All patients with NSCLC were divided into low-or high-risk groups, according to the median risk score of 0.9867 (Figure 3G). The high-risk group had a worse OS probability than the low-risk group ( $p < 0.001$ ; Figure 3E), and the 1-, 3-, and 5-year AUC values were 0.671, 0.638, and 0.637, respectively (Figure 3F). The high-risk group had higher mortality than the low-risk group (Figure 3G). A heatmap was generated to show the different expression profiles of the eight RBPs between the two risk groups (Figure 3H). Kaplan–Meier survival curve analysis showed that the OS probability was significantly worse in the high-risk group than in the low-risk group in the GSE31210 ( $p < 0.001$ ; Figure 4A), GSE26939 ( $p < 0.001$ ; Figure 4B), GSE30219 ( $p < 0.001$ ; Figure 4C), and GSE157009 ( $p < 0.001$ ; Figure 4D) datasets. The 5-year AUC value was 0.655 in the GSE31210 dataset (Figure 4E), 0.630 in the GSE26939 dataset (Figure 4F), 0.595 in the GSE30219 dataset (Figure 4G), and 0.548 in the GSE157009 dataset (Figure 4H). The risk score and

survival status in the GSE31210, GSE26939, GSE30219, and GSE157009 datasets are shown in Figures 4I–K and 4M, respectively.

## Association of risk scores with clinical characteristics

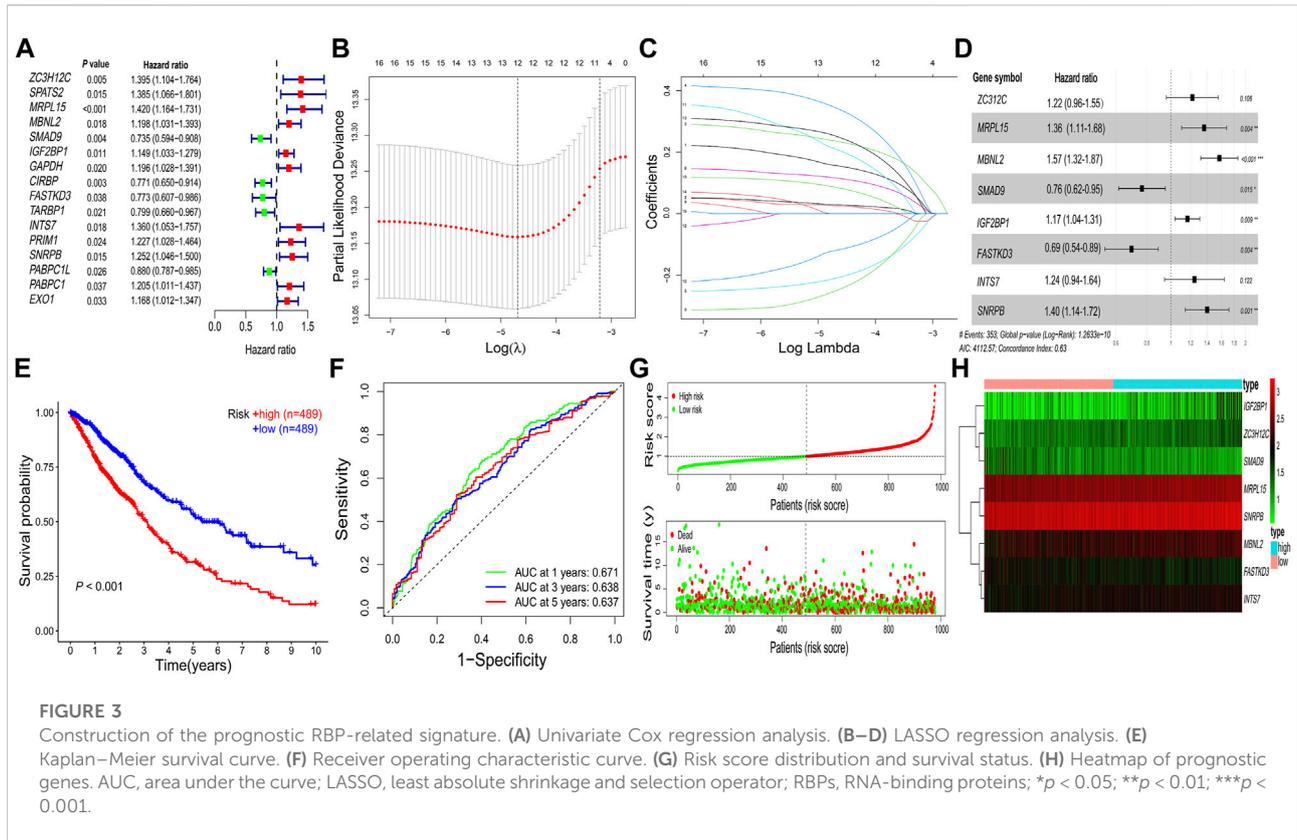
The risk score in patients with stage III–IV cancer was increased compared to that in patients with stage I–II cancer ( $p = 0.004$ , Figure 5A). The risk score was higher in patients with T3–4 cancer than in patients with T1–2 cancer ( $p = 0.003$ , Figure 5B), and the risk score was increased in patients with N1–3 cancer compared to patients with N0 cancer ( $p = 0.009$ , Figure 5C). The risk score of male patients was higher than that of female patients ( $p = 0.020$ , Figure 5D), and the risk score did not differ in different age groups ( $p = 0.340$ , Figure 5E) or M stages ( $p = 0.921$ , Figure 5F). Univariate Cox regression analysis showed that age, UICC stage, and risk score were correlated with prognosis of NSCLC in TCGA database (hazard ratio [HR]: 1.014,  $p = 0.025$ ; HR: 1.475,  $p < 0.001$ ; and HR: 1.939,  $p < 0.001$ , Figure 5G), which was confirmed by multivariate Cox regression analysis (for age: HR: 1.025,  $p < 0.001$ ; for UICC



stage: HR: 1.457,  $p < 0.001$ ; and for risk score: HR: 1.888,  $p < 0.001$ , Figure 5J). The risk score was correlated with prognosis in the GSE26939 dataset by univariate Cox regression analysis (HR: 1.647,  $p = 0.020$ , Figure 5H) and multivariate Cox regression analysis (HR: 1.765,  $p = 0.013$ , Figure 5K). The UICC stage and risk score were correlated with the prognosis of NSCLC in the GSE31210 dataset by univariate Cox regression analysis (HR: 4.232,  $p < 0.001$ ; HR: 1.000,  $p = 0.004$ , Figure 5I) and multivariate Cox regression analysis (HR: 3.734,  $p < 0.001$ ; HR: 1.000,  $p = 0.069$ , Figure 5L).

### Nomogram and calibration plots

A nomogram was constructed using the independent prognostic factors (age, UICC stage, and risk score) to predict OS probability after 1, 3, and 5 years, which was calculated by plotting a vertical line between the total point axis and each prognostic axis (Figure 6A). Calibration plots of the nomogram showed high conformity of the nomogram predicted and actual OS probability at 1, 3, and 5 years (Figure 6B).



## GSEA

The results of the GSEA showed that the P53, focal adhesion, and NOD-like receptor signaling pathways were the primary enriched pathways in the high-risk group of patients with NSCLC (adjusted  $p$ -value  $< 0.05$ , Figure 7).

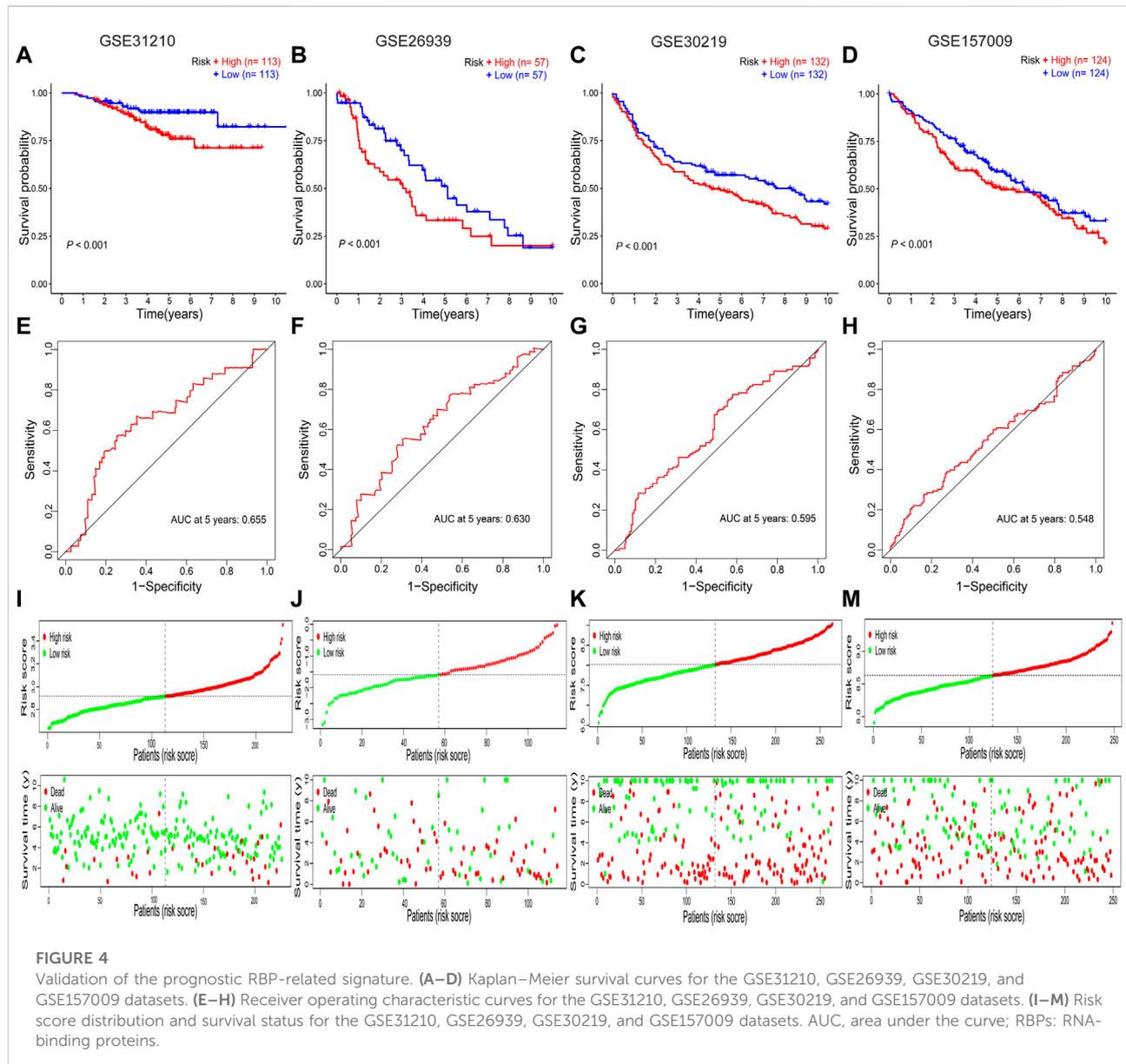
## DNA methylation, TMB, and copy number variation analyses

The DNA methylation levels of the eight prognostic genes were not significantly different between the two risk groups (Supplementary Figure S1). Mutation analysis revealed that TP53 and TTN were the most frequently mutated genes in both risk groups of LUAD (Figures 8A,B) and LUSC (Figures 8C,D). The high-risk group had a higher TMB than the low-risk group in NSCLC ( $p = 0.034$ , Figure 8E), and Spearman's rank correlation analysis showed the positive correlation of TMB with risk score ( $\rho = 0.12$  and  $p < 0.001$ ; Figure 8F). The Kaplan–Meier survival curve showed that the OS probability was lower in the high-risk score + high TMB group than in the low-risk score + low TMB group ( $p < 0.001$ , Figure 8G). Copy number variations were significantly increased in FASTKD3, MBNL2, INTS7, IGF2BP1, and MRPL15, and they were

significantly decreased in NRPB, ZC3H12C, and SMAD9 (Figure 8H). The positions of the prognostic genes in the chromosome are illustrated in Figure 8I.

## Immune infiltration analysis

Because the stromal score ( $p < 0.001$ ), immune score ( $p < 0.01$ ), and estimate score ( $p < 0.001$ ) were all increased in the high-risk group compared to the low-risk group (Figure 9B), the correlations of immune cell with risk score were evaluated. The boxplot showed that the high-risk group had increased activated memory CD4<sup>+</sup> T cells ( $p < 0.001$ ), resting natural killer (NK) cells ( $p < 0.001$ ), M<sub>0</sub> and M<sub>1</sub> macrophages ( $p < 0.01$ ), neutrophils ( $p < 0.01$ ), reduced memory B cells ( $p < 0.001$ ), follicular helper T cells ( $p < 0.001$ ), regulatory T (Treg) cells ( $p < 0.01$ ), monocytes ( $p < 0.05$ ), resting dendritic cells ( $p < 0.05$ ), and resting mast cells ( $p < 0.05$ ) compared to the low-risk group (Figure 9A). The risk score was positively correlated with activated memory CD4<sup>+</sup> T cells ( $\rho = 0.16$ ;  $p < 0.001$ ), resting NK cells ( $\rho = 0.11$ ;  $p < 0.001$ ), M<sub>1</sub> macrophages ( $\rho = 0.16$ ;  $p < 0.01$ ), and neutrophils ( $\rho = 0.12$ ;  $p < 0.01$ ) (Figures 9C–F), and it was negatively correlated with memory B cells ( $\rho = -0.15$ ;  $p < 0.001$ ), follicular helper T cells ( $\rho = -0.13$ ;  $p < 0.001$ ), Treg cells ( $\rho = -0.12$ ;  $p < 0.01$ ), and monocytes ( $\rho = -0.1$ ;  $p = 0.004$ ) (Figures 9G–J). In addition, the relative expression levels of



*PDI* ( $p = 0.015$ ) and *CTLA4* ( $p = 0.042$ ) were higher in the high-risk group than the low-risk group (Figures 9K,L). However, the relative expression levels of *PDL1* were not significantly different between the two risk groups ( $p = 0.087$ , Figure 9M).

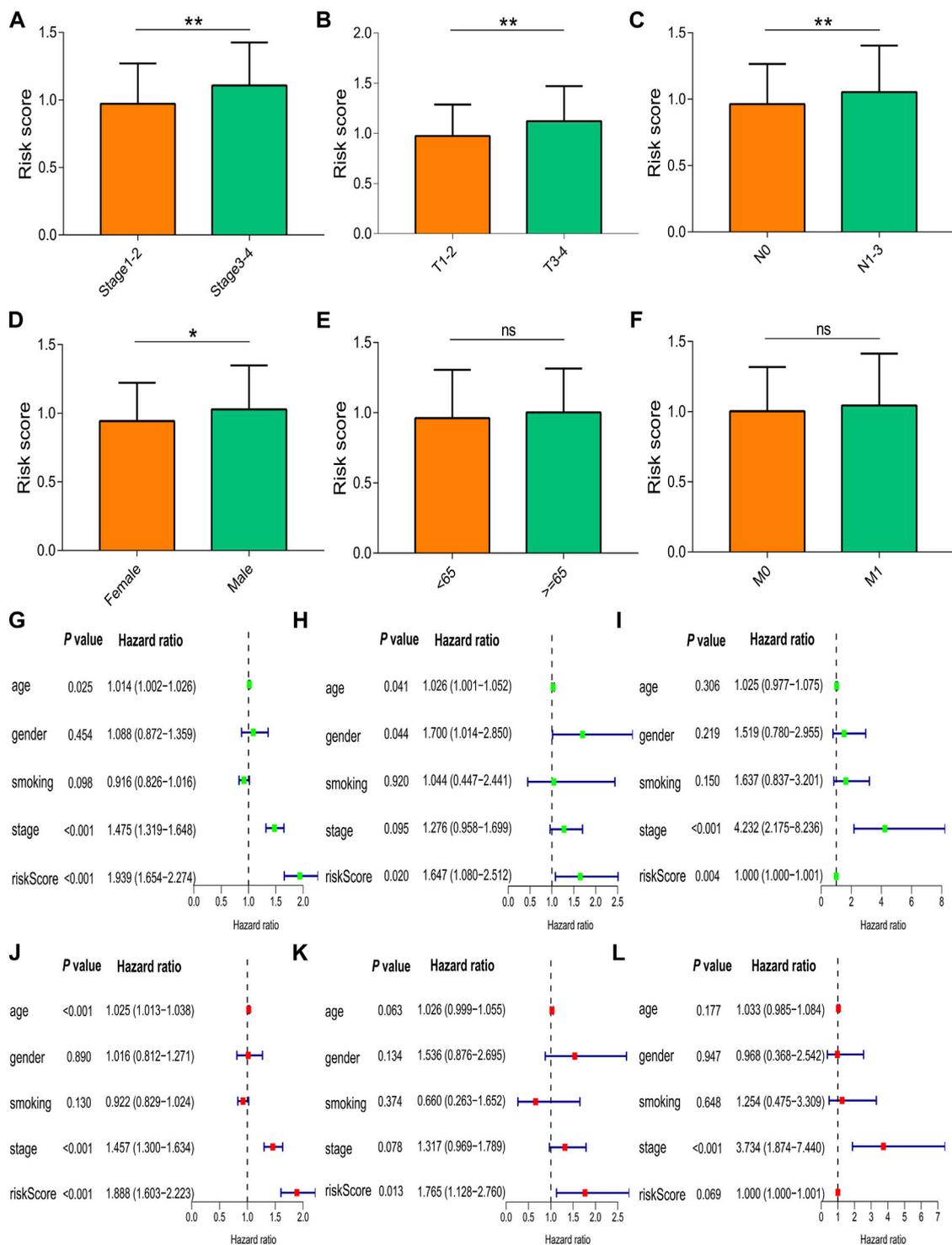
## Sensitivity of chemotherapy drugs

To further explore the clinical value of the prognostic signature, the sensitivity of nine chemotherapy drugs was analyzed and compared between the two risk groups in NSCLC. The results showed that the IC<sub>50</sub> values of dasatinib, docetaxel, erlotinib, gemcitabine, paclitaxel, parthenolide, and

shikonin were lower in the high-risk group than those in the low-risk group ( $p < 0.001$ , Figures 10A–G), whereas the IC<sub>50</sub> values of axitinib and metformin were lower in the low-risk group than those in the high-risk group ( $p < 0.001$ , Figures 10H,I).

## Quantitative PCR validation

The mRNA expression levels of *FASTKD3* ( $p = 0.013$ ), *IGF2BP1* ( $p = 0.026$ ), *MRPL15* ( $p < 0.001$ ), *SNRPB* ( $p = 0.005$ ), and *INTS7* ( $p = 0.027$ ) were higher in tumor tissues than in control tissues (Figures 11A–E); however, the mRNA expression levels of *MBNL2* ( $p = 0.015$ ), *SMAD9* ( $p < 0.001$ ), and



**FIGURE 5**

Correlation of the risk score with clinical characteristics. **(A)** UICC stages. **(B)** T stages. **(C)** N stages. **(D)** Gender groups. **(E)** Age groups. **(F)** M stages. Univariate and multivariate Cox regression analyses in **(G,J)** TCGA, **(H,K)** GSE26939, and **(I,L)** GSE31210. Data are presented as the median (interquartile range). TCGA, The Cancer Genome Atlas; UICC, Union for International Cancer Control; ns, no significance; \* $p < 0.05$ ; \*\* $p < 0.01$ ; \*\*\* $p < 0.001$ .

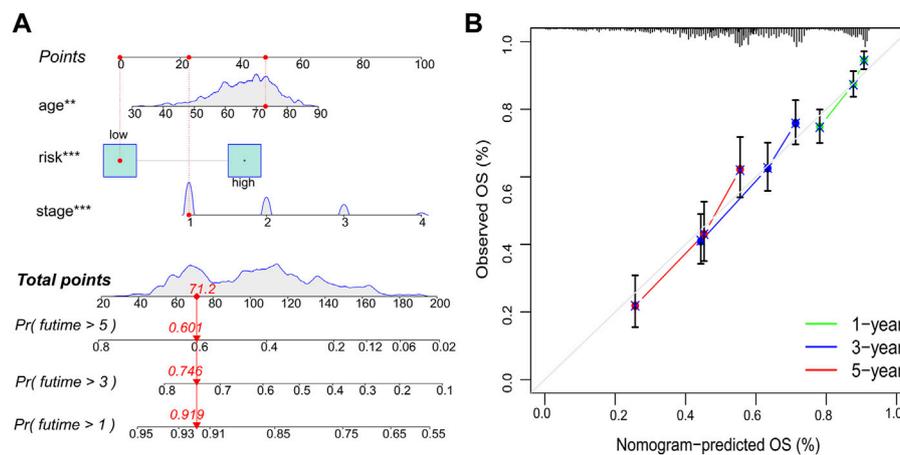


FIGURE 6

Nomogram and calibration plots. (A) Nomogram to predict OS probability at 1, 3, and 5 years. (B) Calibration plots of the nomogram. OS, overall survival; \*\* $p < 0.01$ ; \*\*\* $p < 0.001$ .

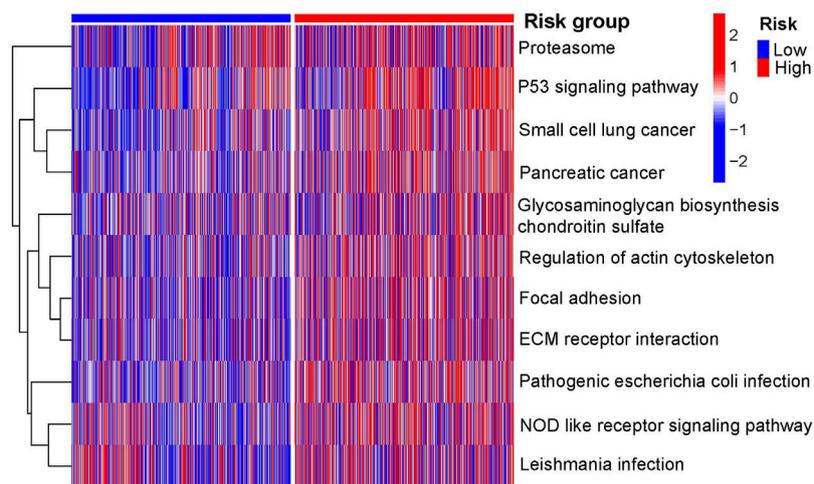


FIGURE 7

Primary KEGG pathways between the two risk groups according to GSEA. GSEA, gene set enrichment analysis; KEGG, Kyoto Encyclopedia of Genes and Genomes.

*ZC3H12C* ( $p < 0.001$ ) were decreased in tumor tissues compared to control tissues (Figures 11F–H).

## Discussion

In the present study, an RBP-related signature was successfully constructed to predict prognosis. The increased risk score was associated with more advanced tumor stages and lower OS probability. The risk score was an independent

prognostic factor for NSCLC when adjusted for age and UICC stage. Moreover, the constructed nomogram better predicted prognosis. In addition, the high-risk group had increased immune infiltration, upregulated relative expression levels of *PD1* and *CTLA4*, higher TMB, and lower IC50 of chemotherapy drugs than the low-risk group.

KEGG pathway analysis was performed to explore the underlying molecular mechanisms of RBPs in NSCLC. Previous studies have demonstrated that the mRNA surveillance pathway is involved in the initiation and

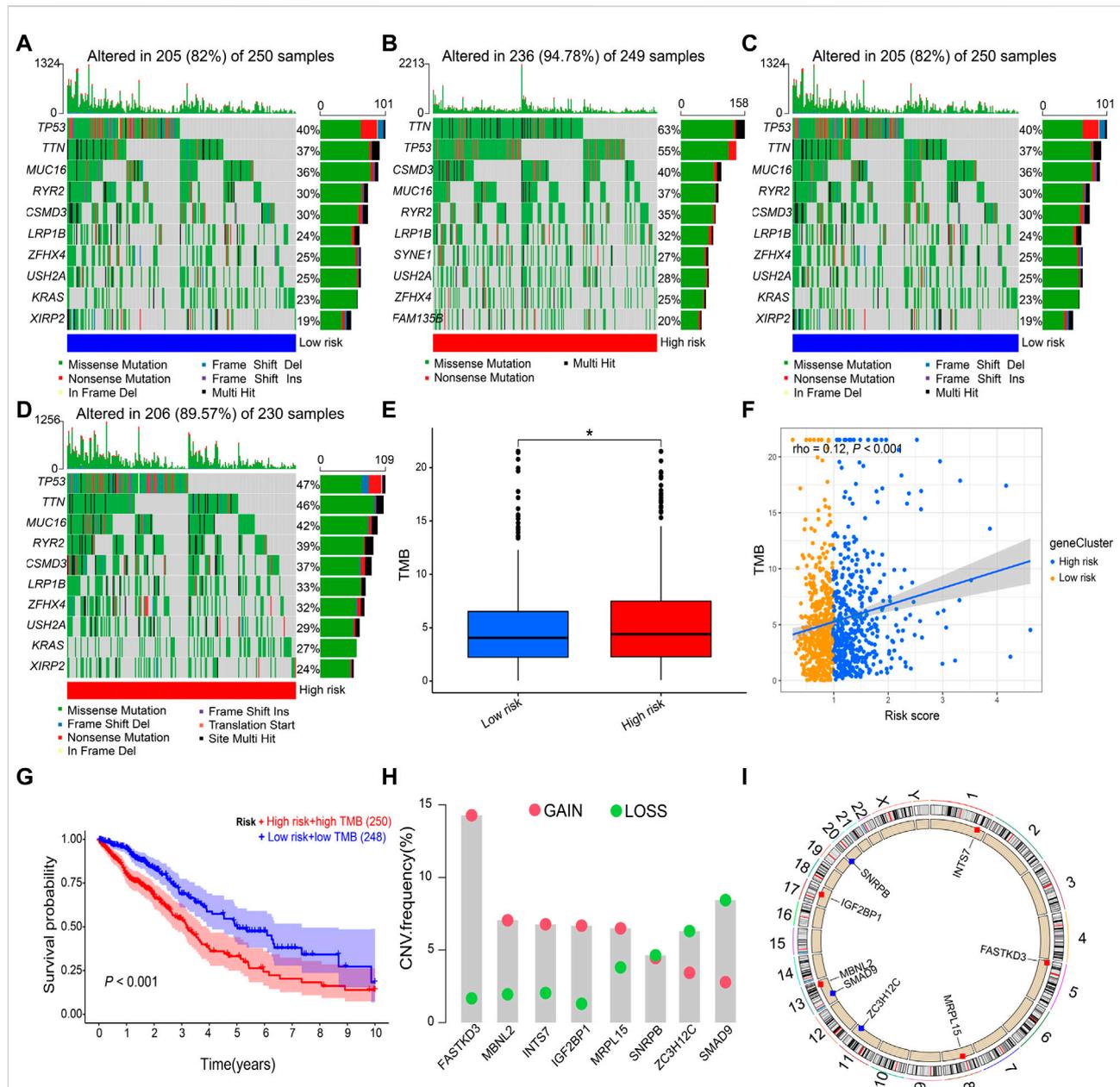
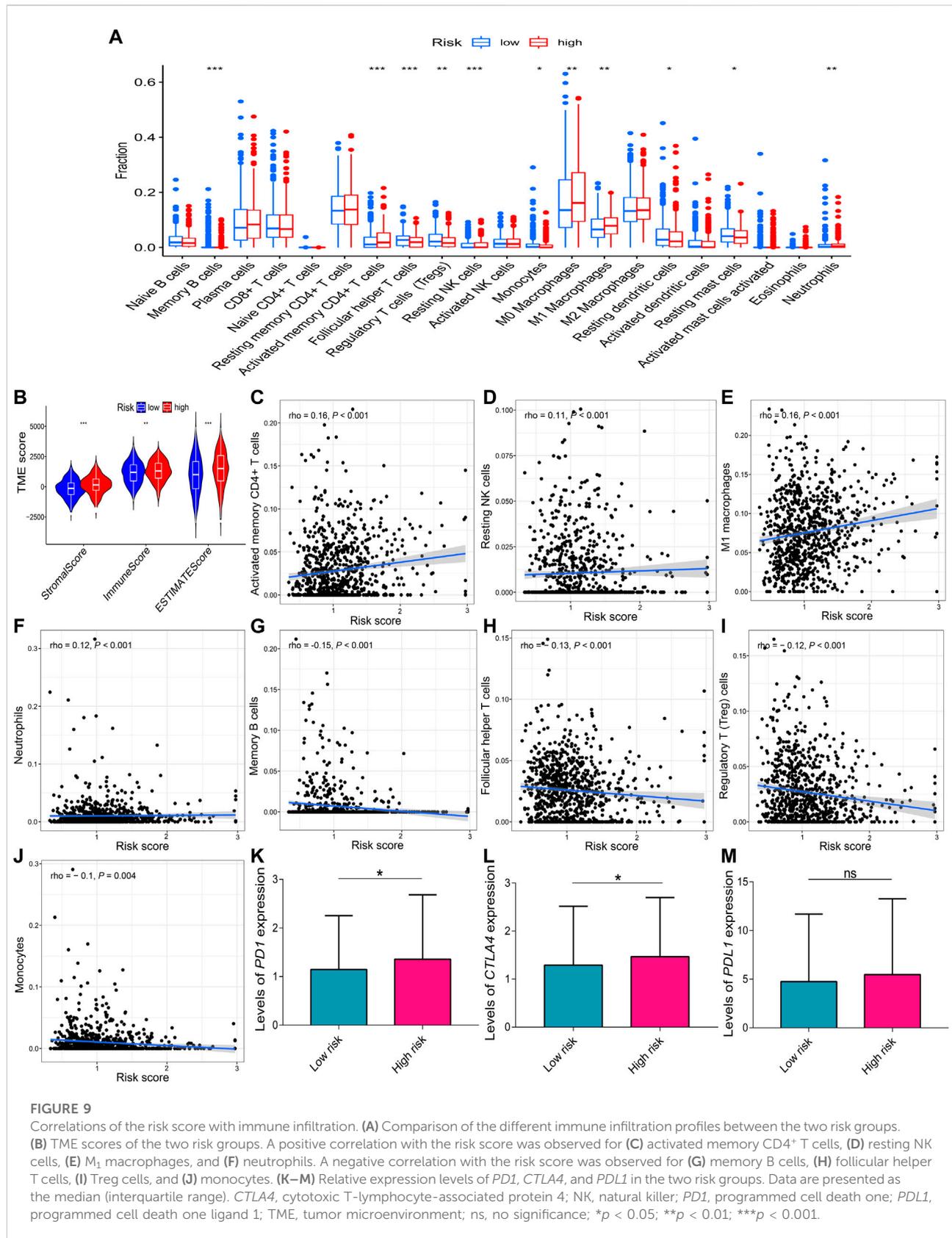


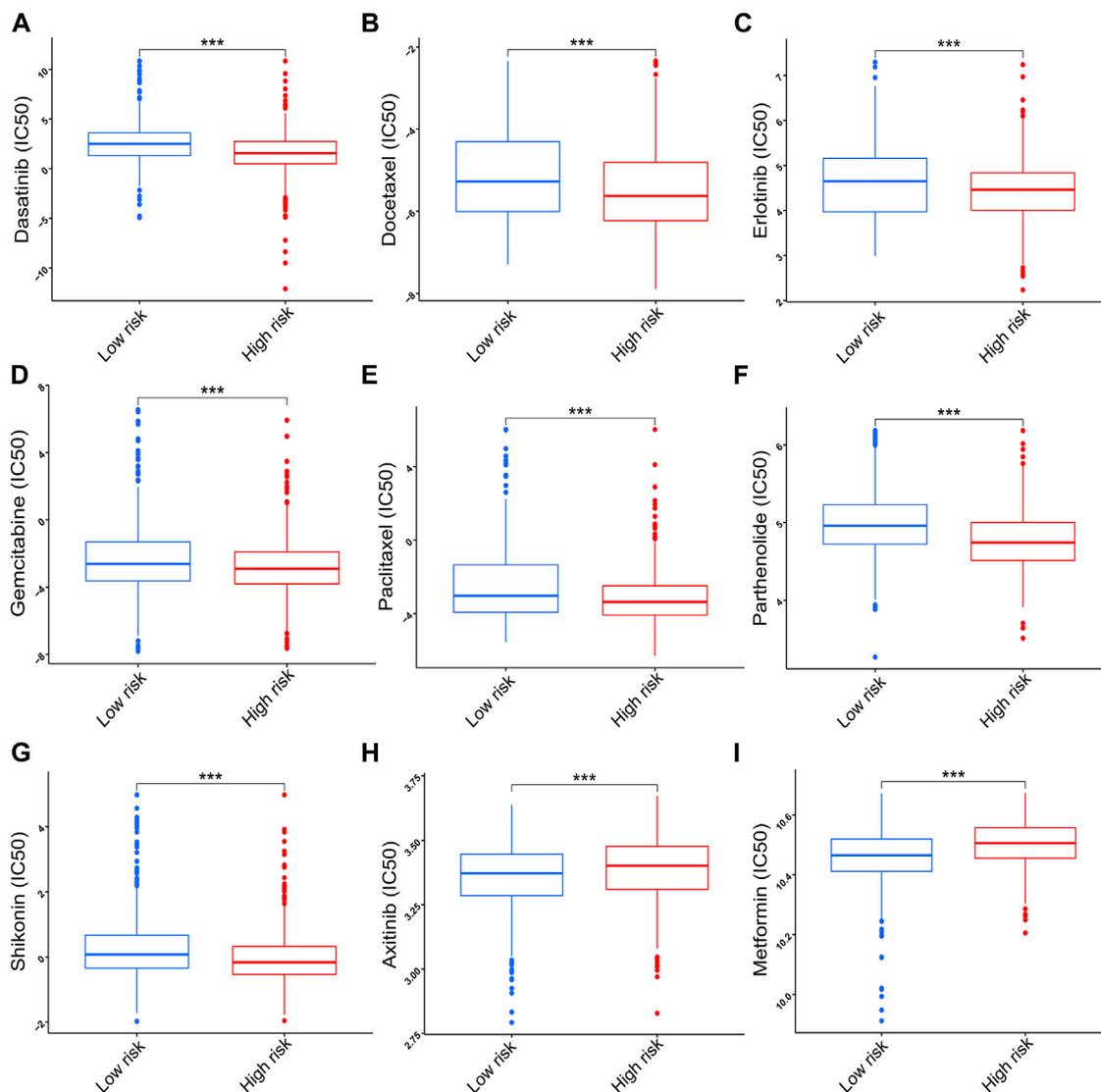
FIGURE 8

Gene mutation and copy number variation. (A,B) Gene mutation frequencies in the low-risk and high-risk groups of LUAD. (C,D) Gene mutation frequencies in the low-risk and high-risk groups of LUSC. (E,F) Correlation of the risk score with TMB. (G) Kaplan–Meier survival curve between the high-risk score + high TMB group and the low-risk score + low TMB group. (H) Copy number variations of prognostic genes. (I) Positions of prognostic genes in the chromosome. LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma; TMB, tumor mutational burden. Data are presented as the median (interquartile range). \* $p < 0.05$ .

progression of cancer (Long et al., 2017; Popp and Maquat, 2018; Zhang et al., 2019a), which was confirmed in NSCLC in the present study. In addition, the present study showed that ribosome biogenesis in eukaryotes played an important role in NSCLC. RPL15, a large ribosomal subunit protein, is significantly upregulated in human cancer tissues and cultured cell lines, and it is closely correlated with

clinicopathological characteristics (Dong et al., 2019). Interestingly, the present study demonstrated that influenza A was correlated with NSCLC. A previous study has reported that exposure to the influenza virus is associated with an increased risk of lung cancer and that the risk increases with cumulative exposure to influenza (Weng et al., 2019). Thus, the annual influenza vaccination administration may reduce





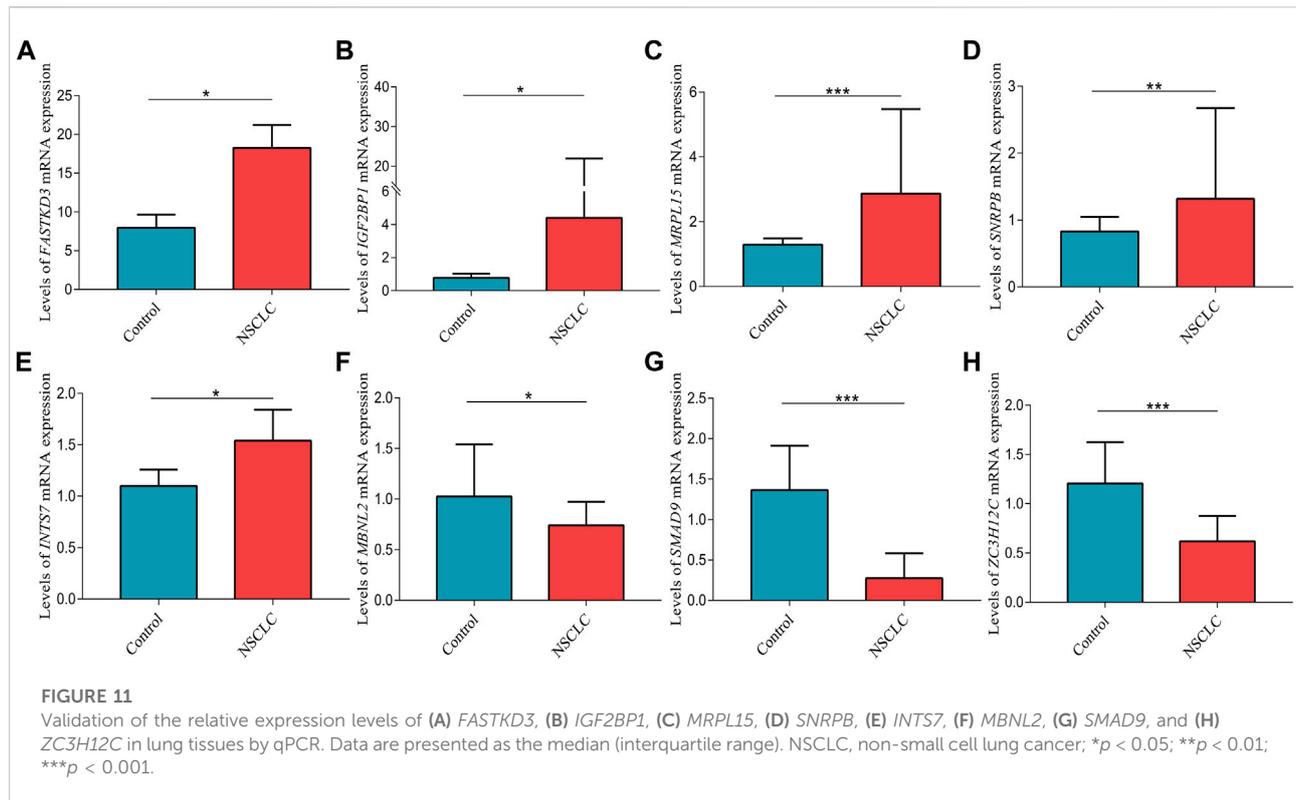
**FIGURE 10**

Correlation of the risk score with chemotherapy sensitivity. (A) Dasatinib. (B) Docetaxel. (C) Erlotinib. (D) Gemcitabine. (E) Paclitaxel. (F) Parthenolide. (G) Shikonin. (H) Axitinib. (I) Metformin. Data are presented as the median (interquartile range). IC50, half inhibitory concentration; \*\*\* $p < 0.001$ .

the incidence of lung cancer in patients with COPD (Chen et al., 2019). However, the precise molecular mechanisms need to be further explored in future studies.

The molecular mechanisms through which these RBPs contribute to the pathogenesis of cancer remain poorly understood. *IGF2BP1* has been shown to function as an oncogene in multiple cancers (Ohdaira et al., 2012; Huang et al., 2020a; Glaß et al., 2020). Low *IGF2BP1* expression inhibits cell proliferation and migration but induces cell cycle arrest and apoptosis in NSCLC, and it correlates with a good prognosis (Huang et al., 2019; Huang et al., 2020b; Zhang et al., 2020). *MRPL15* is associated with the progression of LUAD

(Deng et al., 2020). *SMAD9* is involved in the progression of NSCLC by miR-744 delivered by cancer-derived extracellular vesicles (Gao et al., 2021). Moreover, *SMAD9* mRNA expression is decreased in LUAD, which is correlated with worse OS (Dai et al., 2020). *SNRPB* promotes tumor cell proliferation and stemness by mediating RNA splicing (Zhan et al., 2020). In NSCLC, *SNRPB* facilitates tumorigenesis *via* regulation of *RAB26* expression, and it is correlated with prognosis (Liu et al., 2019). *MBNL2* inhibits tumor growth and metastasis (Lee et al., 2016; Zhang et al., 2019b). *MBNL2* controls lung cancer cell responses to hypoxia by regulating the expression and alternative splicing of hypoxia-induced genes (Fischer et al., 2020). However, there



are few reports on *FASTKD3* and *ZC3H12C* involved in cancer. The results of the present study showed that the relative expression levels of *FASTKD3*, *IGF2BP1*, *MRPL15*, *SNRPB*, and *INTS7* were higher in NSCLC tissues than those in control tissues but that the relative expression levels of *MBNL2*, *SMAD9*, and *ZC3H12C* were reduced in NSCLC tissues compared to those in control tissues.

The prognostic RBP-related signatures have been constructed in LUAD (Meng et al., 2020; Yang et al., 2021) and LUSC (Li et al., 2020a; Zhao et al., 2021). However, the constructed RBP-related signatures for LUAD and LUSC only use simple bioinformatics analysis and lack experimental validation, and each signature is suitable for only one type of NSCLC. Thus, a prognostic RBP-related signature should be generated and validated for NSCLC. In the present study, a prognostic RBP-related signature was successfully constructed using LASSO regression analysis rather than multivariable Cox regression analysis based on eight RBPs that were validated using qPCR, which minimized the risk of overfitting among the signature and increased the reliability of the signature. Thus, the present signature was superior to the previously reported prognostic signatures for LUAD (Meng et al., 2020; Yang et al., 2021) and LUSC (Li et al., 2020a; Zhao et al., 2021). Kaplan–Meier survival analysis showed that the present signature predicted prognosis and discriminated against different risk groups. The risk score was positively correlated

with tumor progression, including UICC stages, T stages, and N stages. However, there was no correlation between the risk score and M stages, which may have been due to an insufficient sample size in the M1 stage. Moreover, the risk score was an independent prognostic factor for NSCLC, according to multivariate Cox regression analysis. Thus, the risk score was associated with the progression and prognosis of NSCLC. Finally, the prognostic RBP-related signature was validated using four GEO datasets, which demonstrated that the prognostic signature was not restricted by different sequencing techniques and platforms. Previous studies have shown that a nomogram better predicts disease prognosis due to its multidimensional parameters (Huang et al., 2022; Luo et al., 2022). Thus, a nomogram was constructed in the present study to predict the 1-, 3-, and 5-year OS probability in NSCLC, and calibration plots of the nomogram showed high predictive accuracy.

The underlying molecular mechanisms between the two risk groups were investigated using GSEA. The pathways were mainly enriched in the *P53* signaling pathway and NOD-like receptor signaling pathway. As a canonical tumor suppressor, *P53* plays an important role in cancer (Yoshida and Miki, 2010; Timofeev et al., 2020). Mutant and wild-type *P53* may exert different functions on cancer (Yoshida and Miki, 2010; Muller and Vousden, 2014; Timofeev et al., 2020). *P53* is a frequently mutated gene in lung cancer (Oduah and Grossman, 2020),

which was also confirmed using mutation analysis in the present study. The mutant *P53* promotes tumor progression by binding to and upregulating chromatin regulatory genes, such as *MLL1* and *MLL2*, leading to genome-wide increases in histone methylation and acetylation (Zhu et al., 2015). Moreover, the mutant *P53* accelerates the recycling of integrin beta1 and *EGFR* to exert its oncogenic function (Muller et al., 2009). Dysregulated TMB is correlated with the prognosis of cancer (Cao et al., 2022; Zhou and Gao, 2022). In the present study, the high-risk group had a higher TMB, and patients in the high-risk group with high TMB had a worse probability. In addition, activation of the NLR family pyrin domain containing 3 (*NLRP3*) inflammasome enhances the proliferation and migration of A549 cells (Wang et al., 2016). Tumor-derived exosomal *TRIM59* induces the tumor-promoting function of macrophages to activate the *NLRP3* inflammasome signaling pathway, thereby promoting lung cancer progression (Liang et al., 2020b). Thus, inhibiting *NLRP3* inflammasome activation may suppress cancer cell proliferation and metastasis in NSCLC (Zou et al., 2018). Nucleotide-binding and oligomerization domain-containing protein 2 (*NOD2*) deficiency confer a protumorigenic macrophage phenotype to promote LUAD progression (Wang et al., 2021). Therefore, *P53* mutation and the NOD-like receptor signaling pathway may play a critical role in disease progression and worse OS probability in the high-risk group.

Immune cells are an important part of the TME, and they play a critical role in tumor development (Bindea et al., 2013). In lung cancer, macrophages stimulate tumor angiogenesis and promote cancer cell invasion, migration, and intravasation (Qian and Pollard, 2010). Tumor-associated macrophages are significantly associated with angiogenesis and a poor prognosis in NSCLC (Montuenga and Pio, 2007; Li et al., 2020e). Neutrophils in peripheral blood are effective diagnostic biomarkers for lung cancer (Zhu et al., 2020). Increased neutrophils are associated with a worse prognosis in bronchoalveolar carcinoma (Bellocq et al., 1998). A high percentage of CD4<sup>+</sup> tumor-infiltrating lymphocytes in the tumor stroma is correlated with a worse prognosis (Giatromanolaki et al., 2021). In the present study, increased M<sub>0</sub> macrophages, M<sub>1</sub> macrophages, neutrophils, and activated memory CD4<sup>+</sup> T cells were found in the high-risk group. The worse OS probability in the high-risk group may be attributed to increased immune infiltration, indicating that the high-risk group may have a better immunotherapy response. Therefore, the correlations of risk scores with the relative expression levels of immune checkpoint inhibitors (*PDI*, *PDL1*, and *CTLA4*) were evaluated to further explore their association with immunotherapy because immune checkpoint inhibitors (ICI) are becoming standard in the first-line treatment of advanced NSCLC (Garon et al., 2015; Reck et al., 2016; Gandhi et al., 2018). The results showed that the high-risk group had higher relative expression levels of *PDI* and *CTLA4*, suggesting that patients in the high-risk group may benefit more from ICIs

against *PDI* and *CTLA4*. In addition to immunotherapy, the correlations of risk scores with chemotherapeutic drugs were also explored. The results demonstrated that the high-risk group was more sensitive to seven chemotherapy drugs and that the low-risk group was more sensitive to two chemotherapy drugs. Thus, the prognostic signature may be applied to guide individualized chemotherapy choices.

The present study had several advantages. First, the signature was constructed using LASSO regression analysis, which minimized the risk of overfitting among the signature. Second, the present study established a promising prognostic RBP-related signature to evaluate patient prognosis, and we performed comprehensive bioinformatics analysis, including correlation of the risk score with clinical characteristics, methylation levels, TMB, copy number variation, immune infiltration, and chemotherapy response, as well as GSEA between the two risk groups, which have not been performed in previous RBP-related signatures for LUAD (Meng et al., 2020; Yang et al., 2021) and LUSC (Li et al., 2020a; Zhao et al., 2021). Third, the previously constructed RBP-related signatures lack experimental validation, but the dysregulated genes in the present prognostic signature were validated using qPCR in another independent cohort. Fourth, the prognostic RBP-related signature was validated using four GEO databases. Last, the present RBP-related signature was generated for all NSCLC cohorts rather than one subgroup of NSCLC. Nevertheless, the present study had several limitations. First, the results of the present study were based on bioinformatics analyses of public databases, which need to be validated in multicentric, prospective clinical studies. Second, Kaplan–Meier survival curve analysis could not be performed using our samples due to insufficient sample size and lack of follow-up data. Last, the present study did not perform *in vivo* and *in vitro* experiments to explore the function of RBPs and prognostic signatures. Thus, further studies are required to clarify the molecular mechanism of RBPs in NSCLC.

## Conclusion

A prognostic RBP-related signature was successfully constructed based on eight RBPs using LASSO regression analysis. The risk score was associated with progression of disease and OS probability, and it was an independent prognostic factor for NSCLC. Moreover, the high-risk group had increased immune infiltration, upregulated relative expression levels of *PDI* and *CTLA4*, higher gene mutation frequency, higher TMB, and better chemotherapy response. Therefore, an RBP-related signature was successfully constructed to predict prognosis in NSCLC, which may function as a reference for individualized therapy, including immunotherapy and chemotherapy.

## Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material; further inquiries can be directed to the corresponding author.

## Ethics statement

The studies involving human participants were reviewed and approved by the Clinical Trial and Biomedical Ethics Committee of West China Hospital of Sichuan University (No. 2016-120). The patients/participants provided their written informed consent to participate in this study.

## Author contributions

T-WM, L-YD and F-YC collected and analyzed data and drafted the manuscript. WX collected data and performed quantitative PCR. J-JF designed the experiment and revised the manuscript.

## Funding

This study is funded by the National Natural Science Foundation of China (No.81870014, 82174139) and 1.3.5 project for disciplines of excellence, West China Hospital, Sichuan University (No. 2018-119).

## References

- Arrieta, O., Zatarain-Barrón, Z. L., Turcott, J. G., Barrón, F., Yendamuri, S., Cardona, A. F., et al. (2022). Association of bmi with benefit of metformin plus epidermal growth factor receptor-tyrosine kinase inhibitors in patients with advanced lung adenocarcinoma: A secondary analysis of a phase 2 randomized clinical trial. *JAMA Oncol.* 8 (3), 477–479. doi:10.1001/jamaoncol.2021.7015
- Belloq, A., Antoine, M., Flahault, A., Philippe, C., Crestani, B., Bernaudin, J. F., et al. (1998). Neutrophil alveolitis in bronchioloalveolar carcinoma: Induction by tumor-derived interleukin-8 and relation to clinical outcome. *Am. J. Pathol.* 152 (1), 83–92.
- Bindea, G., Mlecnik, B., Tosolini, M., Kirilovsky, A., Waldner, M., Obenaus, A. C., et al. (2013). Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. *Immunity* 39 (4), 782–795. doi:10.1016/j.immuni.2013.10.003
- Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., and Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *Ca. Cancer J. Clin.* 68 (6), 394–424. doi:10.3322/caac.21492
- Cao, J., Yang, X., Chen, S., Wang, J., Fan, X., Fu, S., et al. (2022). The predictive efficacy of tumor mutation burden in immunotherapy across multiple cancer types: A meta-analysis and bioinformatics analysis. *Transl. Oncol.* 20, 101375. doi:10.1016/j.tranon.2022.101375
- Chen, K. Y., Wu, S. M., Liu, J. C., and Lee, K. Y. (2019). Effect of annual influenza vaccination on reducing lung cancer in patients with chronic obstructive pulmonary disease from a population-based cohort study. *Medicine* 98 (47), e18035. doi:10.1097/md.00000000000018035
- Cook, K. B., Kazan, H., Zuberi, K., Morris, Q., and Hughes, T. R. (2011). Rbpdb: A database of RNA-binding specificities. *Nucleic Acids Res.* 39, D301–D308. doi:10.1093/nar/gkq1069
- Dai, Z. T., Wang, J., Zhao, K., Xiang, Y., Li, J. P., Zhang, H. M., et al. (2020). Integrated TCGA and GEO analysis showed that SMAD7 is an independent prognostic factor for lung adenocarcinoma. *Medicine* 99 (44), e22861. doi:10.1097/md.00000000000022861
- Deng, F., Shen, L., Wang, H., and Zhang, L. (2020). Classify multicategory outcome in patients with lung adenocarcinoma using clinical, transcriptomic and clinico-transcriptomic data: Machine learning versus multinomial models. *Am. J. Cancer Res.* 10 (12), 4624–4639.
- Dong, Z., Jiang, H., Liang, S., Wang, Y., Jiang, W., and Zhu, C. (2019). Ribosomal protein L15 is involved in colon carcinogenesis. *Int. J. Med. Sci.* 16 (8), 1132–1141. doi:10.7150/ijms.34386
- Facchinetti, F., Marabelle, A., Rossi, G., Soria, J. C., Besse, B., and Tiseo, M. (2016). Moving immune checkpoint blockade in thoracic tumors beyond NSCLC. *J. Thorac. Oncol.* 11 (11), 1819–1836. official publication of the International Association for the Study of Lung Cancer. doi:10.1016/j.jtho.2016.05.027
- Fischer, S., Di Liddo, A., Taylor, K., Gerhardus, J. S., Sobczak, K., Zarnack, K., et al. (2020), 26. New York, NY), 648–663. doi:10.1261/rna.073353.119Muscleblind-like 2 controls the hypoxia response of cancer cellsRNA5
- Gandhi, L., Rodriguez-Abreu, D., Gadgeel, S., Esteban, E., Felip, E., De Angelis, F., et al. (2018). Pembrolizumab plus chemotherapy in metastatic non-small-cell lung cancer. *N. Engl. J. Med.* 378 (22), 2078–2092. doi:10.1056/NEJMoa1801005

## Acknowledgments

The authors thank staff in West China Biobanks, Department of Clinical Research Management, for providing specimens of lung tissue for this study.

## Conflict of interest

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.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2022.930826/full#supplementary-material>

- Gao, L., Tian, Q., Wu, T., Shi, S., Yin, X., Liu, L., et al. (2021). Reduction of miR-744 delivered by NSCLC cell-derived extracellular vesicles upregulates SUV39H1 to promote NSCLC progression via activation of the Smad9/BMP9 axis. *J. Transl. Med.* 19 (1), 37. doi:10.1186/s12967-020-02654-9
- Garon, E. B., Rizvi, N. A., Hui, R., Leigh, N., Balmanoukian, A. S., Eder, J. P., et al. (2015). Pembrolizumab for the treatment of non-small-cell lung cancer. *N. Engl. J. Med.* 372 (21), 2018–2028. doi:10.1056/NEJMoa1501824
- Gerstberger, S., Hafner, M., and Tuschl, T. (2014). A census of human RNA-binding proteins. *Nat. Rev. Genet.* 15 (12), 829–845. doi:10.1038/nrg3813
- Giatromanolaki, A., Anestopoulos, I., Panayiotidis, M. I., Mitrakas, A., Pappa, A., and Koukourakis, M. I. (2021). Prognostic relevance of the relative presence of CD4, CD8 and CD20 expressing tumor infiltrating lymphocytes in operable non-small cell lung cancer patients. *Anticancer Res.* 41 (8), 3989–3995. doi:10.21873/anticancer.15196
- Glaß, M., Michl, P., and Hüttelmaier, A. S. (2020). RNA binding proteins as drivers and therapeutic target candidates in pancreatic ductal adenocarcinoma. *Int. J. Mol. Sci.* 21 (11), E4190. doi:10.3390/ijms21114190
- Goldstraw, P., Chansky, K., Crowley, J., Rami-Porta, R., Asamura, H., Eberhardt, W. E., et al. (2016). The IASLC lung cancer staging project: Proposals for revision of the TNM stage groupings in the forthcoming (eighth) edition of the TNM classification for lung cancer. *J. Thorac. Oncol.* 11 (1), 39–51. official publication of the International Association for the Study of Lung Cancer. doi:10.1016/j.jtho.2015.09.009
- Guo, B., Wei, J., Wang, J., Sun, Y., Yuan, J., Zhong, Z., et al. (2022). CD44-targeting hydrophobic phosphorylated gemcitabine prodrug nanotherapeutics augment lung cancer therapy. *Acta Biomater.* 145, 200–209. doi:10.1016/j.actbio.2022.04.016
- Hong, S. (2017). RNA binding protein as an emerging therapeutic target for cancer prevention and treatment. *J. Cancer Prev.* 22 (4), 203–210. doi:10.15430/jcp.2017.22.4.203
- Huang, H., Wang, D., Guo, W., Zhuang, X., and He, Y. (2019). Correlated low IGF2BP1 and FOXM1 expression predicts a good prognosis in lung adenocarcinoma. *Pathol. Res. Pract.* 215 (7), 152433. doi:10.1016/j.prp.2019.152433
- Huang, L., Lin, W., Xie, D., Yu, Y., Cao, H., Liao, G., et al. (2022). Development and validation of a preoperative CT-based radiomic nomogram to predict pathology invasiveness in patients with a solitary pulmonary nodule: A machine learning approach, multicenter, diagnostic study. *Eur. Radiol.* 32 (3), 1983–1996. doi:10.1007/s00330-021-08268-z
- Huang, Q., Guo, H., Wang, S., Ma, Y., Chen, H., Li, H., et al. (2020). A novel circular RNA, circXPO1, promotes lung adenocarcinoma progression by interacting with IGF2BP1. *Cell Death Dis.* 11 (12), 1031. doi:10.1038/s41419-020-03237-8
- Huang, W., Dong, S., Cha, Y., and Yuan, X. (2020). SNHG11 promotes cell proliferation in colorectal cancer by forming a positive regulatory loop with c-Myc. *Biochem. Biophys. Res. Commun.* 527 (4), 985–992. doi:10.1016/j.bbrc.2020.05.002
- Kang, C., Jia, X., and Liu, H. (2020). Development and validation of a RNA binding protein gene pair-associated prognostic signature for prediction of overall survival in hepatocellular carcinoma. *Biomed. Eng. Online* 19 (1), 68. doi:10.1186/s12938-020-00812-0
- Kim, C., Liu, S. V., Crawford, J., Torres, T., Chen, V., Thompson, J., et al. (2021). A phase I trial of dasatinib and osimertinib in TKI naïve patients with advanced EGFR-mutant non-small-cell lung cancer. *Front. Oncol.* 11, 728155. doi:10.3389/fonc.2021.728155
- Lang, Y., Kong, X., He, C., Wang, F., Liu, B., Zhang, S., et al. (2017). Musashi1 promotes non-small cell lung carcinoma malignancy and chemoresistance via activating the akt signaling pathway. *Cell. Physiol. Biochem.* 44 (2), 455–466. doi:10.1159/000485012
- Lee, Y. H., Jhuang, Y. L., Chen, Y. L., Jeng, Y. M., and Yuan, R. H. (2016). Paradoxical overexpression of MBNL2 in hepatocellular carcinoma inhibits tumor growth and invasion. *Oncotarget* 7 (40), 65589–65601. doi:10.18632/oncotarget.11577
- Li, W., Gao, L. N., Song, P. P., and You, C. G. (2020). Development and validation of a RNA binding protein-associated prognostic model for lung adenocarcinoma. *Aging* 12 (4), 3558–3573. doi:10.18632/aging.102828
- Li, W., Li, N., Gao, L., and You, C. (2020). Integrated analysis of the roles and prognostic value of RNA binding proteins in lung adenocarcinoma. *PeerJ* 8, e8509. doi:10.7717/peerj.8509
- Li, W., Li, X., Gao, L. N., and You, C. G. (2020). Integrated analysis of the functions and prognostic values of RNA binding proteins in lung squamous cell carcinoma. *Front. Genet.* 11, 185. doi:10.3389/fgene.2020.00185
- Li, X., Huang, R., Li, M., Zhu, Z., Chen, Z., Cui, L., et al. (2020). Parthenolide inhibits the growth of non-small cell lung cancer by targeting epidermal growth factor receptor. *Cancer Cell Int.* 20 (1), 561. doi:10.1186/s12935-020-01658-1
- Li, Y., You, M. J., Yang, Y., Hu, D., and Tian, C. (2020). The role of tumor-associated macrophages in leukemia. *Acta Haematol.* 143 (2), 112–117. doi:10.1159/000500315
- Liang, G., Meng, W., Huang, X., Zhu, W., Yin, C., Wang, C., et al. (2020). miR-196b-5p-mediated downregulation of TSPAN12 and GATA6 promotes tumor progression in non-small cell lung cancer. *Proc. Natl. Acad. Sci. U. S. A.* 117 (8), 4347–4357. doi:10.1073/pnas.1917531117
- Liang, M., Chen, X., Wang, L., Qin, L., Wang, H., Sun, Z., et al. (2020). Cancer-derived exosomal TRIM59 regulates macrophage NLRP3 inflammasome activation to promote lung cancer progression. *J. Exp. Clin. Cancer Res.* 39 (1), 176. doi:10.1186/s13046-020-01688-7
- Liu, N., Wu, Z., Chen, A., Wang, Y., Cai, D., Zheng, J., et al. (2019). SNRPB promotes the tumorigenic potential of NSCLC in part by regulating RAB26. *Cell Death Dis.* 10 (9), 667. doi:10.1038/s41419-019-1929-y
- Long, N. P., Jung, K. H., Yoon, S. J., Anh, N. H., Nghi, T. D., Kang, Y. P., et al. (2017). Systematic assessment of cervical cancer initiation and progression uncovers genetic panels for deep learning-based early diagnosis and proposes novel diagnostic and prognostic biomarkers. *Oncotarget* 8 (65), 109436–109456. doi:10.18632/oncotarget.22689
- Luo, L. M., Wang, Y., Lin, P. X., Su, C. H., and Huang, B. T. (2022). The clinical outcomes, prognostic factors and nomogram models for primary lung cancer patients treated with stereotactic body radiation therapy. *Front. Oncol.* 12, 863502. doi:10.3389/fonc.2022.863502
- Meng, L., He, X., Zhang, X., Zhang, X., Wei, Y., Wu, B., et al. (2020). Predicting the clinical outcome of lung adenocarcinoma using a novel gene pair signature related to RNA-binding protein. *Biomed. Res. Int.* 2020, 8896511. doi:10.1155/2020/8896511
- Montuenga, L. M., and Pio, R. (2007). Tumour-associated macrophages in nonsmall cell lung cancer: The role of interleukin-10. *Eur. Respir. J.* 30 (4), 608–610. doi:10.1183/09031936.00091707
- Muller, P. A., Caswell, P. T., Doyle, B., Iwanicki, M. P., Tan, E. H., Karim, S., et al. (2009). Mutant p53 drives invasion by promoting integrin recycling. *Cell* 139 (7), 1327–1341. doi:10.1016/j.cell.2009.11.026
- Muller, P. A., and Vousden, K. H. (2014). Mutant p53 in cancer: New functions and therapeutic opportunities. *Cancer Cell* 25 (3), 304–317. doi:10.1016/j.ccr.2014.01.021
- Müller-McNicoll, M., and Neugebauer, K. M. (2013). How cells get the message: Dynamic assembly and function of mRNA-protein complexes. *Nat. Rev. Genet.* 14 (4), 275–287. doi:10.1038/nrg3434
- Ni, M., Liu, X., Wu, J., Zhang, D., Tian, J., Wang, T., et al. (2018). Identification of candidate biomarkers correlated with the pathogenesis and prognosis of non-small cell lung cancer via integrated bioinformatics analysis. *Front. Genet.* 9, 469. doi:10.3389/fgene.2018.00469
- Oduah, E. I., and Grossman, S. R. (2020). Harnessing the vulnerabilities of p53 mutants in lung cancer - focusing on the proteasome: A new trick for an old foe? *Cancer Biol. Ther.* 21 (4), 293–302. doi:10.1080/15384047.2019.1702403
- Ohdaira, H., Sekiguchi, M., Miyata, K., and Yoshida, K. (2012). MicroRNA-494 suppresses cell proliferation and induces senescence in A549 lung cancer cells. *Cell Prolif.* 45 (1), 32–38. doi:10.1111/j.1365-2184.2011.00798.x
- Pan, J., Li, M., Yu, F., Zhu, F., Wang, L., Ning, D., et al. (2021). Up-regulation of p53/miR-628-3p pathway, a novel mechanism of shikonin on inhibiting proliferation and inducing apoptosis of A549 and PC-9 non-small cell lung cancer cell lines. *Front. Pharmacol.* 12, 766165. doi:10.3389/fphar.2021.766165
- Popp, M. W., and Maquat, L. E. (2018). Nonsense-mediated mRNA decay and cancer. *Curr. Opin. Genet. Dev.* 48, 44–50. doi:10.1016/j.gde.2017.10.007
- Qian, B. Z., and Pollard, J. W. (2010). Macrophage diversity enhances tumor progression and metastasis. *Cell* 141 (1), 39–51. doi:10.1016/j.cell.2010.03.014
- Reck, M., Rodríguez-Abreu, D., Robinson, A. G., Hui, R., Csösz, T., Fülöp, A., et al. (2016). Pembrolizumab versus chemotherapy for PD-L1-positive non-small-cell lung cancer. *N. Engl. J. Med.* 375 (19), 1823–1833. doi:10.1056/NEJMoa1606774
- Saito, Y., Takekuma, Y., Shinagawa, N., and Sugawara, M. (2022). Evaluation of risk factors associated with carboplatin and nab-paclitaxel treatment suspension in patients with non-small cell lung cancer. *Support. Care Cancer* 30 (5), 4081–4088. doi:10.1007/s00520-021-06757-x
- Shen, D. J., Jiang, Y. H., Li, J. Q., Xu, L. W., and Tao, K. Y. (2020). The RNA-binding protein RBM47 inhibits non-small cell lung carcinoma metastasis through modulation of AXIN1 mRNA stability and Wnt/ $\beta$ -catenin signaling. *Surg. Oncol.* 34, 31–39. doi:10.1016/j.suronc.2020.02.011
- Solomon, B., Callejo, A., Bar, J., Berchem, G., Bazhenova, L., Saintigny, P., et al. (2022). A WIN Consortium phase I study exploring avelumab, palbociclib, and

- axitinib in advanced non-small cell lung cancer. *Cancer Med.* 11, 2790–2800. doi:10.1002/cam4.4635
- Song, G., Guo, G., Du, T., Li, X., Wang, J., Yan, Y., et al. (2020). RALY may cause an aggressive biological behavior and a dismal prognosis in non-small-cell lung cancer. *Exp. Cell Res.* 389 (2), 111884. doi:10.1016/j.yexcr.2020.111884
- Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., et al. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* 102 (43), 15545–15550. doi:10.1073/pnas.0506580102
- Sun, L., Yuan, W., Wen, G., Yu, B., Xu, F., Gan, X., et al. (2020). Parthenolide inhibits human lung cancer cell growth by modulating the IGF-1R/PI3K/Akt signaling pathway. *Oncol. Rep.* 44 (3), 1184–1193. doi:10.3892/or.2020.7649
- Timofeev, O., Koch, L., Niederau, C., Tscherne, A., Schneikert, J., Klimovich, M., et al. (2020). Phosphorylation control of p53 DNA-binding cooperativity balances tumorigenesis and aging. *Cancer Res.* 80 (23), 5231–5244. doi:10.1158/0008-5472.can-20-2002
- Wang, D., Zhou, J., Fang, W., Huang, C., Chen, Z., Fan, M., et al. (2022). A multifunctional nanotherapeutic agent potentiates erlotinib to EGFR wild-type non-small cell lung cancer. *Bioact. Mat.* 13, 312–323. doi:10.1016/j.bioactmat.2021.10.046
- Wang, Y., Kong, H., Zeng, X., Liu, W., Wang, Z., Yan, X., et al. (2016). Activation of NLRP3 inflammasome enhances the proliferation and migration of A549 lung cancer cells. *Oncol. Rep.* 35 (4), 2053–2064. doi:10.3892/or.2016.4569
- Wang, Y., Miao, Z., Qin, X., Li, B., and Han, Y. (2021). NOD2 deficiency confers a pro-tumorigenic macrophage phenotype to promote lung adenocarcinoma progression. *J. Cell. Mol. Med.* 25 (15), 7545–7558. doi:10.1111/jcmm.16790
- Wang, Z., Tang, W., Yuan, J., Qiang, B., Han, W., and Peng, X. (2020). Integrated analysis of RNA-binding proteins in glioma. *Cancers* 12 (4), E892. doi:10.3390/cancers12040892
- Weng, C. F., Chen, L. J., Lin, C. W., Chen, H. M., Lee, H. H., Ling, T. Y., et al. (2019). Association between the risk of lung cancer and influenza: A population-based nested case-control study. *Int. J. Infect. Dis. IJID* 88, 8–13. official publication of the International Society for Infectious Diseases. doi:10.1016/j.ijid.2019.07.030
- Xiao, D., Hu, X., and Zhang, J. (2022). Tumour targeted polymer nanoparticles co-loaded with docetaxel and siCCAT2 for combination therapy of lung cancer. *J. Drug Target.* 30, 534–543. doi:10.1080/1061186x.2021.2016773
- Yan, M., Sun, L., Li, J., Yu, H., Lin, H., Yu, T., et al. (2019). RNA-binding protein KHSRP promotes tumor growth and metastasis in non-small cell lung cancer. *J. Exp. Clin. Cancer Res.* 38 (1), 478. doi:10.1186/s13046-019-1479-2
- Yang, L., Zhang, R., Guo, G., Wang, G., Wen, Y., Lin, Y., et al. (2021). Development and validation of a prediction model for lung adenocarcinoma based on RNA-binding protein. *Ann. Transl. Med.* 9 (6), 474. doi:10.21037/atm-21-452
- Yoshida, K., and Miki, Y. (2010). The cell death machinery governed by the p53 tumor suppressor in response to DNA damage. *Cancer Sci.* 101 (4), 831–835. doi:10.1111/j.1349-7006.2010.01488.x
- Zhan, Y. T., Li, L., Zeng, T. T., Zhou, N. N., Guan, X. Y., and Li, Y. (2020). SNRPB-mediated RNA splicing drives tumor cell proliferation and stemness in hepatocellular carcinoma. *Aging* 13 (1), 537–554. doi:10.18632/aging.202164
- Zhang, J., Luo, W., Chi, X., Zhang, L., Ren, Q., Wang, H., et al. (2020). IGF2BP1 silencing inhibits proliferation and induces apoptosis of high glucose-induced non-small cell lung cancer cells by regulating Netrin-1. *Arch. Biochem. Biophys.* 693, 108581. doi:10.1016/j.abb.2020.108581
- Zhang, J., Zheng, Z., Wu, M., Zhang, L., Wang, J., Fu, W., et al. (2019). The natural compound neobactatin inhibits tumor metastasis by upregulating the RNA-binding-protein MBNL2. *Cell Death Dis.* 10 (8), 554. doi:10.1038/s41419-019-1789-5
- Zhang, L., Lu, D., Liu, M., Zhang, M., and Peng, Q. (2019). Identification and interaction analysis of key miRNAs in medullary thyroid carcinoma by bioinformatics analysis. *Mol. Med. Rep.* 20 (3), 2316–2324. doi:10.3892/mmr.2019.10463
- Zhao, S., Liu, Q., Li, J., Hu, C., Cao, F., Ma, W., et al. (2021). Construction and validation of prognostic regulation network based on RNA-binding protein genes in lung squamous cell carcinoma. *DNA Cell Biol.* 40 (12), 1563–1583. doi:10.1089/dna.2021.0145
- Zhou, B., and Gao, S. (2022). Construction and validation of a novel immune and tumor mutation burden-based prognostic model in lung adenocarcinoma. *Cancer Immunol. Immunother.* 71 (5), 1183–1197. doi:10.1007/s00262-021-03066-4
- Zhu, J., Sammons, M. A., Donahue, G., Dou, Z., Vedadi, M., Getlik, M., et al. (2015). Gain-of-function p53 mutants co-opt chromatin pathways to drive cancer growth. *Nature* 525 (7568), 206–211. doi:10.1038/nature15251
- Zhu, X., Chen, Y., and Cui, Y. (2020). Absolute neutrophil count and mean platelet volume in the blood as biomarkers to detect lung cancer. *Dis. markers* 2020, 1371964. doi:10.1155/2020/1371964
- Zou, J., Yang, Y., Yang, Y., and Liu, X. (2018). Polydatin suppresses proliferation and metastasis of non-small cell lung cancer cells by inhibiting NLRP3 inflammasome activation via NF- $\kappa$ B pathway. *Biomed. Pharmacother. = Biomedicine Pharmacother.* 108, 130–136. doi:10.1016/j.biopha.2018.09.051