



Signature stemmed from two transcription factor families determines histological fate and regulates immune infiltration in patients with lung cancer

Qian Hong^{1^}, Shuang Liang², Rui Han¹, Chenguang Zhao¹, Juwei Mu^{1^}

¹Department of Thoracic Surgery, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China; ²State Key Laboratory of Molecular Oncology, Department of Clinical Laboratory, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

Contributions: (I) Conception and design: J Mu; (II) Administrative support: Q Hong; (III) Provision of study materials or patients: J Mu; (IV) Collection and assembly of data: Q Hong, S Liang, R Han; (V) Data analysis and interpretation: Q Hong, S Liang, C Zhao; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Juwei Mu, MD. Professor, Department of Thoracic Surgery, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, No. 17 Panjiayuan South Lane, Chaoyang District, Beijing 100021, China. Email: mujuwei@cicams.ac.cn.

Background: Earlier research has reported that transcription factors play a crucial role in the anti-tumorigenic immune response of lung cancer patients. The aim of this study is to determine the relationship between post-translational modifications of transcription factors and histological fate and patient prognosis.

Methods: Based on the information of 293 lung cancer patients in the Gene Expression Omnibus (GEO) database, differentially expressed genes (DEGs) related to the interferon regulatory factor (IRF) and signal transducer and activator of transcription (STAT) families between patients experiencing early death and those with long-term survival were identified and characterized. A survival prediction model was established by incorporating 7 STAT genes and 9 IRF genes into the least absolute shrinkage and selection operator (LASSO) algorithm. Gene Ontology (GO) enrichment analysis indicated that these two transcription factor families can govern lung cancer tissue differentiation and predict patient prognosis. Moreover, the Cox proportional hazards regression model was applied to select the genes with the highest predictive capability to construct a gene-based signature. Lastly, the data of 1,803 and 784 lung cancer patients from the Kaplan-Meier plotter (KMPLLOT) and The Cancer Genome Atlas (TCGA) databases were used to evaluate the accuracy and sensitivity of the model.

Results: Based on the minimum criterion, TRIM28, IRF3, and STAT3 were employed to generate the prognostic model. The 1-, 3-, and 5-year area under the curve (AUC) values of the three-gene-based signature showed consistent results, signifying that the model had excellent accuracy and sensitivity in predicting overall survival (OS) for patients with lung cancer. Finally, the three-gene signature and tumor-node-metastasis (TNM) staging system were combined to construct a nomogram for evaluating the OS of lung cancer patients. TRIM28 may affect the stability of IRF3. Encouragingly, the predicted OS was highly consistent with the observed OS in multiple cohorts.

Conclusions: Taken together, these findings implied that the predictive model based on the three-gene signature showed robust discriminatory performance.

Keywords: Lung cancer; biomarkers; transcriptomics; immunology

[^] ORCID: Qian Hong, 0000-0001-7218-2681; Juwei Mu, 0000-0002-1972-8646.

Submitted May 06, 2024. Accepted for publication Jul 26, 2024. Published online Sep 13, 2024.

doi: 10.21037/jtd-24-733

View this article at: <https://dx.doi.org/10.21037/jtd-24-733>

Introduction

As well documented, lung cancer is one of the leading cancer types both in terms of incidence and mortality (1). Indeed, its 5-year survival rate is less than 30% (2). The non-small cell lung cancer (NSCLC) histological subtype accounts for approximately 85% of lung cancer patients, of which lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC) are the most prevalent (1). Ascribed to the malignant nature and considerable inter-/intra-tumor heterogeneity of NSCLC, its treatment and management have extensively garnered the attention of clinicians and researchers. Despite advancements in treatment strategies such as the use of cytotoxic drugs, small-molecule tyrosine kinase inhibitors, and the introduction of immunotherapy, having enhanced survival benefits in selected patients, the survival and overall cure rates for NSCLC remain largely unsatisfactory. Immune checkpoint blockades (ICB) have been established as an effective therapeutic strategy for NSCLC but are still associated with challenges such as innate and acquired resistance. Previous researchers have dedicated tremendous efforts and resources to unraveling the mechanisms underlying resistance and identifying

solutions, such as combining immunotherapy and other adjuvant therapies. Nevertheless, there remains an urgent need for a more extensive characterization of immune therapy resistance.

Inflammation is elicited as a defense mechanism to combat invading pathogens, which are recognized by specific receptors that reside in the plasma membrane or the endosome. Specifically, the receptors sense danger-associated molecular patterns or pathogen-associated molecular patterns. Stimulation of the receptors leads to the induction of pro-inflammatory signaling cascades and transcription factors such as signal transducer and activator of transcription (STAT) and interferon regulatory factor (IRF), which in turn mediate the synthesis and release of inflammatory cytokines such as chemokines and macrophage migration inhibitory factors, interleukins, and tumor necrosis factor (TNF) (3). Cyclic guanosine 5'-monophosphate-adenosine monophosphate (GMP-AMP) synthase (cGAS) is a double-stranded DNA sensor that catalyzes the synthesis of cyclic GMP-AMP (cGAMP), which stimulates type I interferons through the stimulator of interferon response cGAMP interactor 1 (STING)-TANK binding kinase 1 (TBK1)-IRF3 signaling axis. STING oligomerizes after the binding of cGAMP, leading to the recruitment and activation of the TBK1 kinase. The IRF3 transcription factor is then recruited to the signaling complex and activated by TBK1 (4). Therefore, the IRF-STAT family is believed to play a vital role in the anti-tumorigenic immune response of lung cancer patients.

In this study, pivotal genes of the IRF-STAT family were selected to delineate three primary clusters of lung cancer patients by leveraging publicly available databases. Next, the gene expression pattern of the two clusters was analyzed to identify differentially expressed immune-related pathways and genes. Lastly, a prognostic model for survival based on these genes was constructed and validated in the Gene Expression Omnibus (GEO) and The Cancer Genome Atlas (TCGA) databases. We present this article in accordance with the TRIPOD reporting checklist (available at <https://jtd.amegroups.com/article/view/10.21037/jtd-24-733/rc>).

Highlight box

Key findings

- High expression of TRIM28 led to higher levels of IRF3, which was associated with poorer histological differentiation and poor prognosis in lung cancer patients.

What is known and what is new?

- Interferon regulatory factor (IRF) and STAT families have been considered as principal regulators of type I interferons response. However, the relationship between these transcription factors and the prognosis of lung cancer patients requires further exploration.
- Ubiquitination, as its common regulatory pathway, plays a more critical role in tumor progression. This research evaluated the prognostic performance and potential therapeutic value of ubiquitin ligase TRIM28.

What is the implication, and what should change now?

- For lung cancer patients, the signature established by TRIM28 combined with IRF3 and STAT3 can be a new prognostic marker.

Methods

Publicly available messenger RNA (mRNA) data

This study retrieved data from three publicly available databases. GSE30219, the most comprehensive lung cancer dataset from the GEO database (<http://www.ncbi.nlm.nih.gov/geo/>), and TCGA data of samples from lung cancer patients (5) were acquired from the UCSC Xena (<https://tcga.xenahubs.net>) to serve as one of the individual validation sets. Additional gene expression and prognostic datasets of lung cancer patients were downloaded from the Kaplan-Meier plotter (KMPLLOT) database (<http://kmplot.com/analysis/>) (6,7). After log₂ transformation and quantile normalization, mRNA expression data detected using multiple probes were calculated by computing the mean expression values. Patients with incomplete information were subsequently excluded from the analysis.

Cell lines and cell culture

Human lung cancer cell lines, NCI-H358 and NCI-H520, were cultured in RPMI-1640 media (10-040-CV; Corning, NY, USA) containing 10% fetal bovine serum (FBS) and antibiotic mixture, including 100 U/mL penicillin and 100 µg/mL streptomycin (15140-122; Gibco, NY, USA). 293T cells were cultured in Dulbecco's modified eagle medium (DMEM) media (10-013-CVR; Corning) containing 10% FBS and antibiotic mixture.

Western blotting and cell counting kit-8 (CCK-8) assay

Protein was extracted from cell lines with Pierce RIPA Buffer (C1053; Applygen Technologies, Beijing, China) and quantified by Pierce BCA Protein Assay Kit (23225 Thermo Fisher Scientific, NY, USA). Protein extracts were subjected to polyacrylamide gel electrophoresis using the 4–20% Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) gel system (AQ-129; Aoqing Biotechnology, Beijing, China), transferred to polyvinylidene fluoride (PVDF) membranes (Millipore, Darmstadt, Germany). Transferred protein was immunoblotted using antibodies against KAP1 (ab109287; Abcam, MA, USA), IRF3 (ab68481; Abcam), and GAPDH (ab8245; Abcam), incubation at 4 °C overnight. The membranes with horseradish peroxidase (HRP)-conjugated secondary antibodies at room temperature for 1 h, and visualized the protein bands using enhanced chemiluminescence.

A CCK-8 assay was used to analyse cell proliferation.

Cells were seeded into 96-well plates at 6,000 cells per well. After 24, 48, 72, 96, and 120 h, Cells were treated with 10 µL CCK-8 reagent (Dojindo, Kumamoto, Japan) for 2 h, and the absorbance values at 450 nm were recorded.

IRF-STAT-related clustering for lung cancer

Principal component analysis (PCA) and hierarchical cluster were applied to IRF-STAT family genes to identify molecular subtypes of lung cancer. These algorithms were unsupervised class discovery algorithms to classify patients by applying specific clustering techniques to the random subsets of the data.

Functional enrichment analysis and gene set enrichment analysis (GSEA)

Gene Ontology (GO) functional enrichment analysis was performed to identify significantly enriched pathways associated with differentially expressed genes (DEGs) correlated with the signature, using the R package “clusterProfiler” (8). Biological pathways with $P < 0.05$ were considered significant, using functional annotation chart options with the whole human genome as a background. Meanwhile, GSEA was also performed between different risk subgroups via “javaGSEA” to obtain GSEA results (9).

Prognostic model establishment and validation

The least absolute shrinkage and selection operator (LASSO) algorithm was employed to further screen prognosis-specific IRF-STAT family genes (10). The linear model for microarray data (LIMMA) package was employed to identify DEGs and targeted metabolic enzymes in patients with different molecular subtypes. The Cox proportional hazards regression model was used to identify significant prognostic genes and to determine correlation estimated coefficients, which were employed to develop a risk score incorporating the expression levels of optimized genes. Patients were divided into high-risk and low-risk groups according to the signature.

Following this, time-dependent receiver operating characteristic (ROC) analysis was utilized to determine the cut-off value and to calculate the area under the curve (AUC) for 1-, 3-, and 5-year overall survival (OS) and relapse-free survival in order to verify the performance of the signature using the ‘survivalROC’ R package (11). Kaplan-Meier (K-M) survival curve analyses and log-rank tests were used

to evaluate the prognostic significance of the risk score formula. Besides, the relationship between the distribution of patients' risk scores and survival and recurrence status was investigated. Using the "ComplexHeatmap" R package, a heatmap with cluster analysis based on differences in gene expression was generated (12). Variables with $P < 0.05$ in the univariate model were used to construct the nomogram. ROC and K-M survival analyses were used to evaluate the performance of the developed nomogram.

For clinical assessment, human tissue microarray (TMA) chips that contain 90 cases of paired LUAD and paracancerous normal specimens were purchased (HLugA180Su11, ShGnghGi Outdo Biotech Company, Shanghai, China). We performed immunohistochemistry (IHC) of several markers (IRF3, STAT3 and TRIM28) on the TMA. Two independent pathologists without prior knowledge of our research evaluated the IHC staining results. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

Results

Identification and characterization of DEGs between patients experiencing early death and those with long-term survival

The data of 293 lung cancer patients with complete clinical and transcriptional information were downloaded from the GSE30219 dataset. When categorizing patients who died within 2 years as early death cases and those who survived for more than 4 years as long-term survival cases, a total of 128 patients exhibited significant differences in prognosis. Then, GSEA was performed on the GO gene sets, and the findings exposed that the immune response to cytokine production and the JAK-STAT signaling pathway were markedly enriched (Figure 1A). Thereafter, DEGs associated with the cGAS-STING pathway were selected through a thorough literature review, and the results demonstrated that the IRF family and STAT family were two important transcription factor families related to the cGAS-STING pathway and interferon response. Thus, 7 STAT genes and 9 IRF genes were introduced into unsupervised cluster analysis. Importantly, PCA determined significant heterogeneity in the expression levels of these two transcription factor families in lung cancer patients (Figure 1B). At the same time, hierarchical clustering accurately characterized these three patient clusters (Figure 1C). The expression patterns of IRF and STAT genes in these three clusters are displayed

in the heatmap (Figure 1D). It is worth noting that patients in groups A and C were predominantly composed of LUAD and LUSC patients, while those in group B were largely poorly differentiated lung cancer patients. Additionally, K-M survival analysis demonstrated significant differences in prognosis between unsupervised lung cancer subgroups based on the expression levels of IRF and STAT, with group A patients manifesting a better prognosis (Figure 1E). Then, all 16 genes were further inputted into the LASSO algorithm to establish a survival prediction model. According to the minimum criterion, IRF3 and STAT3 were selected for the construction of the risk signature formula (Figure 1F).

Furthermore, GO enrichment analysis indicated that immune pathways played an instrumental role in regulating lung cancer and influencing patient prognosis.

DEGs were identified using LIMMA for both groups A versus B and groups A versus C, with a fold change ≥ 2.5 and P value < 0.01 . Of note, the expression level of 43 genes was differentially upregulated in group C compared to group A, whereas that of 21 genes was differentially downregulated. Similarly, the expression level of 23 genes was differentially upregulated in group B compared to group A, whilst that of 44 genes was downregulated. We speculated that these two transcription factor families could regulate the expression levels of key genes involved in type II alveolar epithelial development (such as SFTPB, SFTPD, SFTA2, Figure 2A-2G), and keratinization-related genes, which were upregulated in group C (KRT family, Figure 2G-2I). In addition, neuroendocrine-related genes, such as INSM1, CHGA, and CHGB, were significantly upregulated in group B and, therefore, may also be regulated by these two transcription factor families (Figure 2D-2F). Based on these results, the complement pathway and B cell-mediated immune response were identified as key biological mechanisms mediating the impact of IRF and STAT on the prognosis of lung cancer patients (Figure 2B, 2C).

Target identification and establishment of prognostic models

Attributed to challenges in targeting transcription factors with bioactive molecules, an attempt was made to identify their upstream regulatory factors from metabolic enzymes. A list of 3,667 metabolic enzymes was downloaded from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database, focusing on those that were positively correlated with IRF3 and negatively correlated with STAT3 while also significantly correlated with the prognosis of lung cancer

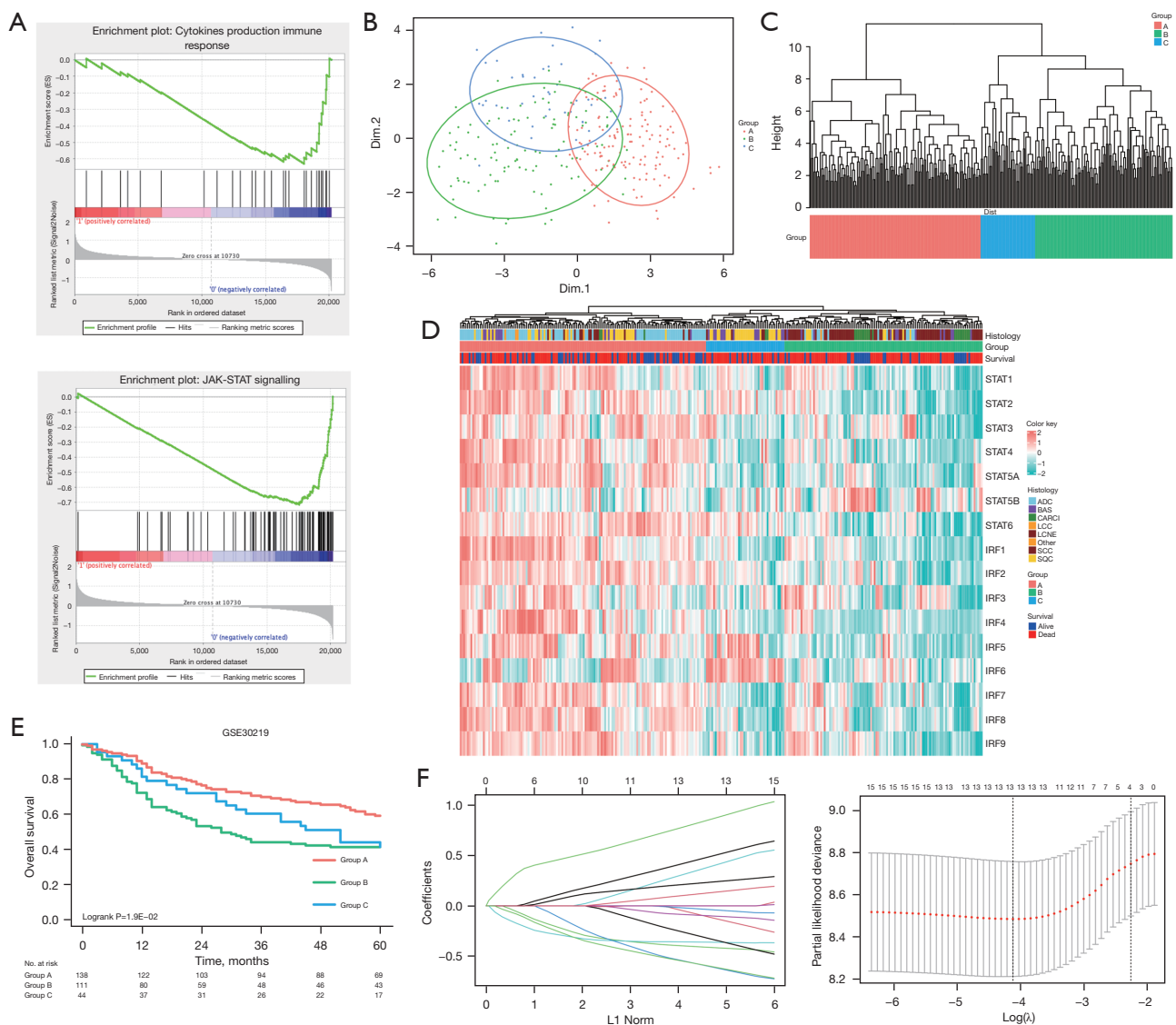


Figure 1 Concentration on IRF and STAT transcription factor family through unsupervised learning in GSE30219. (A) Lung cancer patients with different prognoses enriched into the IRF and STAT families through GSEA. (B) PCA based on two transcription factor families. (C) Hierarchical cluster on account of 16 IRF and STAT genes. (D) The expression pattern of IRF and STAT families and clinicopathological features of patients from GSE30219. (E) Kaplan-Meier survival curves of OS for three unsupervised groups in the GSE30219 cohort. (F) LASSO and partial likelihood deviance coefficient profiles of the selected genes. Dim, dimension; STAT, signal transducer and activator of transcription; IRF, interferon regulatory factor; ADC, adenocarcinoma; BAS, basaloid squamous cell carcinoma; CARCI, carcinoids; LCC, large cell carcinoma; LCNE, large cell neuroendocrine carcinoma; SCC, small cell carcinoma; SQC, squamous cell carcinoma; GSE, gene series; GSEA, gene set enrichment analysis; PCA, principal component analysis; OS, overall survival; LASSO, least absolute shrinkage and selection operator.

patients. Noteworthy, TRIM28, an enzyme that may regulate protein ubiquitination, was significantly enriched in IRF3-expressing patients (*Figure 3A*). A Cox proportional hazards regression model was applied to select the most

predictive genes, which yielded three genes, comprising TRIM28, IRF3, and STAT3 ($P < 0.05$). Upregulating those genes was associated with worse survival outcomes in lung cancer patients. The risk score was calculated as

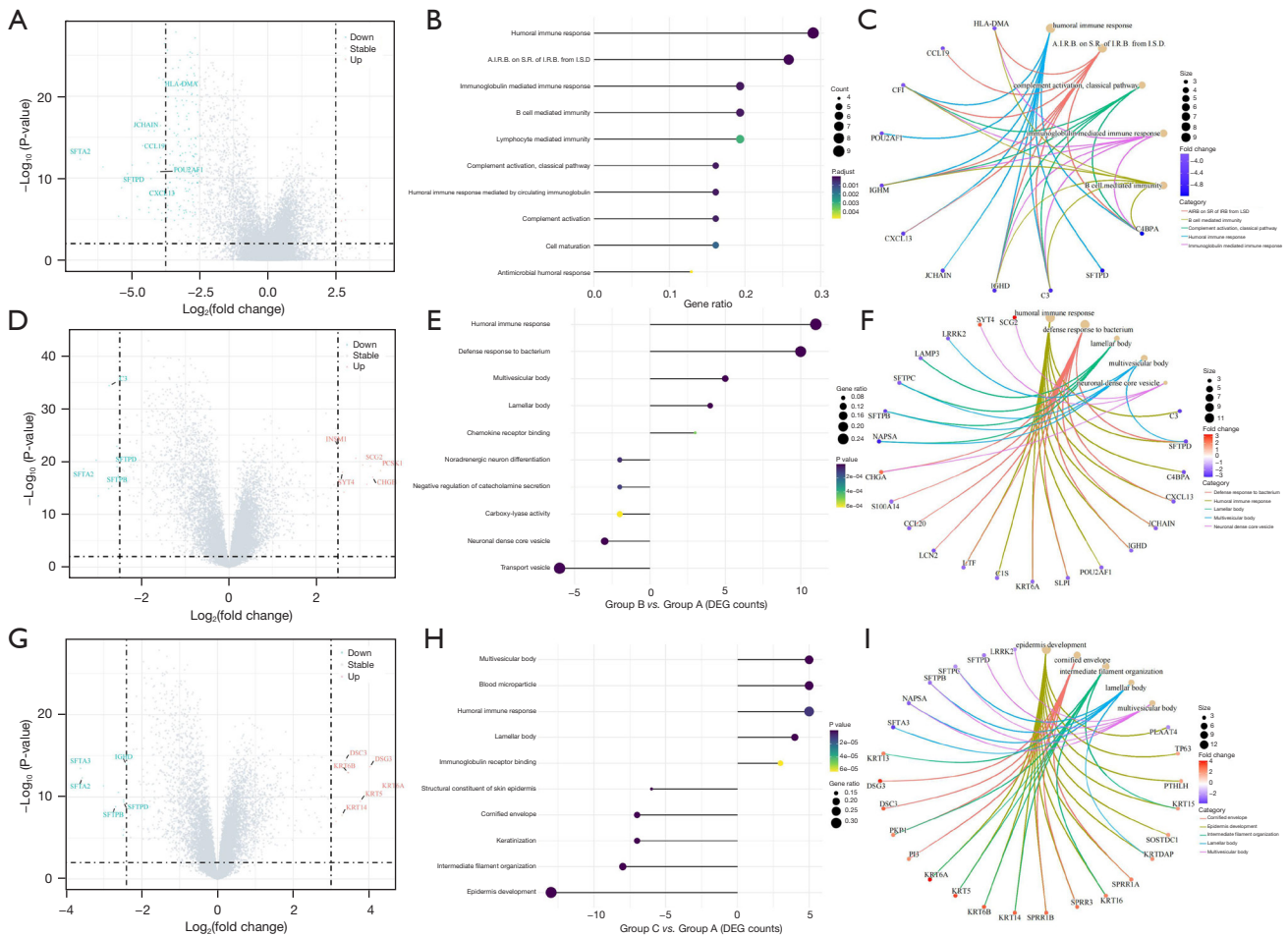


Figure 2 Differential expression and GO pathway enrichment analysis between three unsupervised groups. (A) DEGs in group A relative to group B and group C. (B) Significantly upregulated pathways in group A, which are associated with better prognosis in lung cancer patients. (C) Cnetplot of upregulated genes in group A. (D) DEGs in group A relative to group B. (E) Significantly upregulated pathways in group A and group B. (F) Cnetplot of upregulated genes in group A and group B. (G) DEGs in group A relative to group C. (H) Significantly upregulated pathways in group A and group C. (I) Cnetplot of upregulated genes in group A and group C. AIRB, adaptive immune response based; SR, somatic recombination; IRB, immune receptors built; ISD, immunoglobulin superfamily domains; DEGs, differentially expressed genes; GO, Gene Ontology.

follows: risk value = $(0.4652 \times \text{TRIM28 expression}) + (0.2910 \times \text{IRF3 expression}) - (0.2776 \times \text{STAT3 expression})$. This formula was subsequently used to calculate the risk score for each patient from the GSE30219, TCGA, and KMPLLOT databases. As anticipated, K-M survival analyses illustrated that the OS of patients in the low-risk group was significantly longer than that in the high-risk group (Figure 3B). The distributions of patient survival and risk scores calculated based on the signature are displayed in Figure 3C. Moreover, time-dependent ROC analyses were conducted to assess the prognostic performance of the

three-gene-based classifier, with AUC of 0.714, 0.687, and 0.691 at the 1, 3, and 5 years, respectively (Figure 3D). Next, the interactions between TRIM28, IRF3, and STAT3, as well as downstream molecules that regulate neuroendocrine differentiation, were predicted using the STRING database (Figure 3E).

Validation of the prognostic model for OS in larger patient cohorts

The clinical information of 1,803 and 784 lung cancer

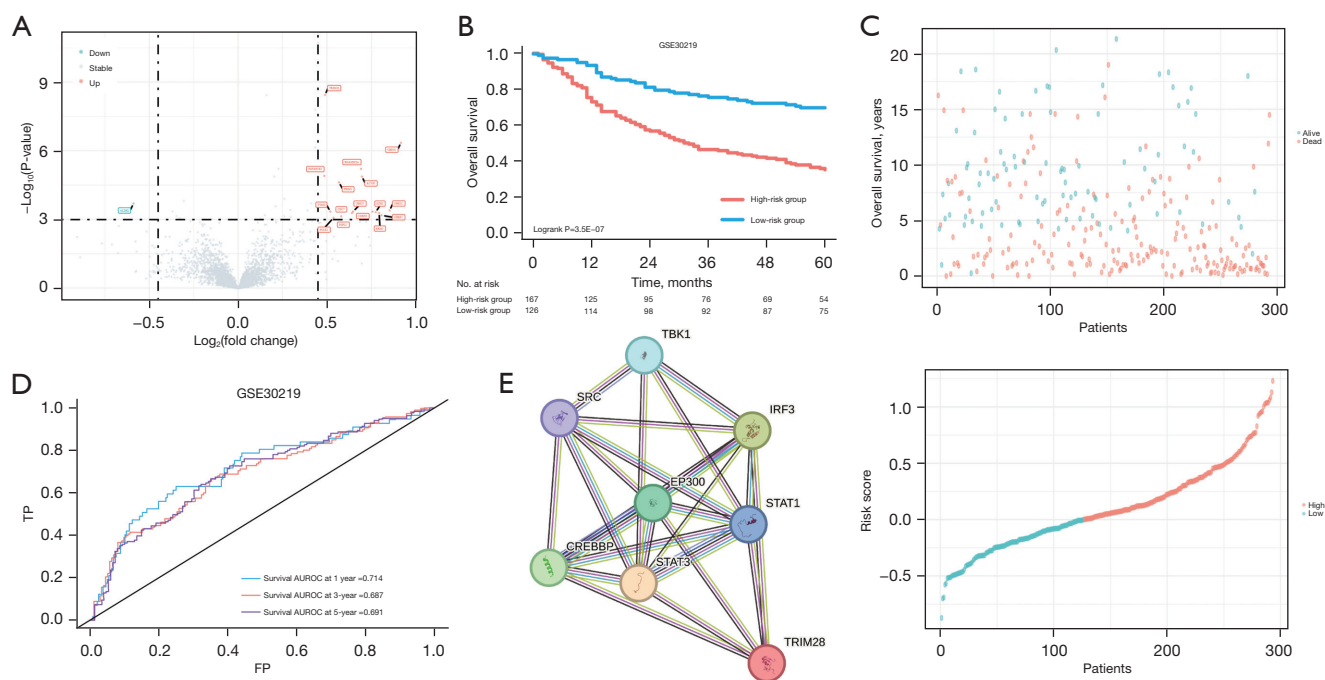


Figure 3 Identification and establishment of three-gene signature combined with performance and functional analysis. (A) TRIM28, the most significant differential gene. (B) Kaplan-Meier survival curves of OS between high-risk and low-risk patients in the GSE30219 cohort. (C) The distributions of the risk score and survival status of lung cancer patients. (D) AUC values of ROC predicted 1-, 3- and 5-year OS of the signature in the GSE30219 cohort. (E) Predicting protein-protein interactions based on STRING database. TP, true positive; FP, false positive; OS, overall survival; GSE, gene series; AUC, area under the curve; ROC, receiver operating characteristic.

patients was retrieved from KMPLLOT and TCGA databases, respectively, with pathological types composed of LUAD and squamous cell carcinoma. The 1-, 3-, and 5-year AUC values of the three-gene-based signature were consistent with the aforementioned results, corroborating the superior accuracy and sensitivity of the model to predict the OS of lung cancer patients (Figure 4A,4B). Notably, K-M survival analyses displayed that the OS in the high-risk group was significantly shorter (Figure 4C,4D). What's more, the survival distribution of high-risk cancer patients was significantly worse in both datasets (Figure 4E,4F).

After a comprehensive analysis of the results of univariable Cox regression, four variables, including the signature, T stage, N stage, and M stage, were identified as independent predictive parameters. Consequently, they were utilized to develop a nomogram in the GSE30219 cohort (Figure 5A). Calibration curves were used to examine the reliability and predictive ability of the nomogram (Figure 5B-5D). Time-dependent ROC analyses were employed to assess the predictive performance of the nomogram. The AUC were 0.723, 0.714, and 0.751 for 1-,

3-, and 5-year OS, respectively (Figure 5E). Lastly, K-M survival analyses revealed that patients in the low-risk group had significantly longer OS (Figure 5F, $P < 0.0001$).

We evaluated the IHC expression level of IRF3, STAT3, and TRIM28 in our cohort and capitalized on time-dependent ROC analyses to verify the operation of three-gene signature. On account of the result of TMA, the model predicted that the AUC for survival of 1, 3, and 5 years were 0.946, 0.931, and 0.858, respectively (Figure 6A). Significantly, the OS of patients in the high-risk group appeared to be lower than that in the low-risk group (Figure 6B). High-risk patients were concentrated in dying within 3 years (Figure 6C). Compared to normal tissue, IRF3 was notably upregulated in tumor tissue (Figure 6D). The nomogram constructed based on IHC and clinical features showed good consistency, and the prognostic effect of T stage was higher than grade in the cohort of LUAD patients (Figure 6E,6F). The representative IHC images of patient A and patient B selected from the high-risk group in Figure 6C are shown in the upper and lower parts of Figure 6G, respectively. Besides, we deployed

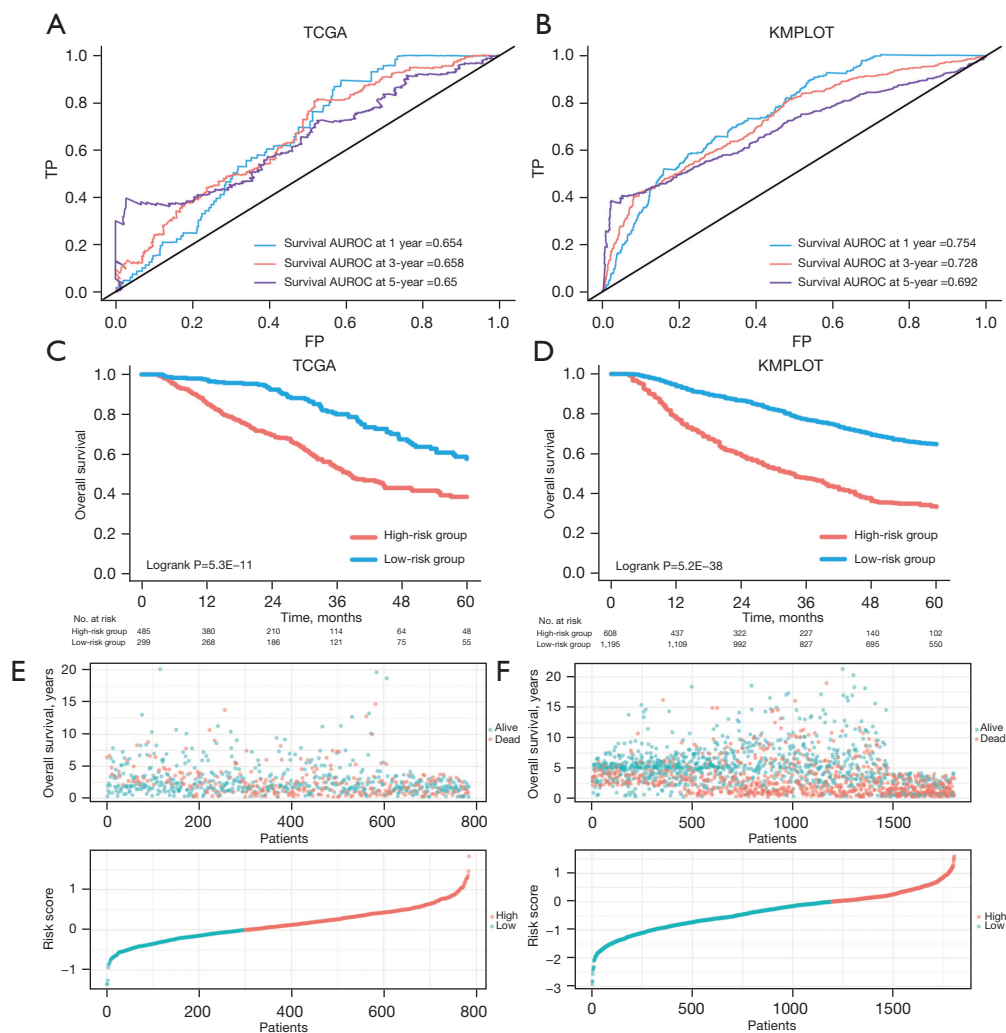


Figure 4 Validation of the predictive performance of the three-gene signature in the TCGA and KMPLLOT cohorts. (A) Time-dependent ROC curves at the OS time of 1, 3, and 5 years in the TCGA cohort. (B) Time-dependent ROC curves at the OS time of 1, 3, and 5 years in the KMPLLOT cohort. (C) Kaplan-Meier survival curves of OS between high-risk and low-risk patients in the TCGA cohort. (D) Kaplan-Meier survival curves of OS between high-risk and low-risk patients in the KMPLLOT cohort. (E) The distributions of the risk score and survival time and status in the TCGA cohort. (F) The distributions of the risk score and survival time and status in the KMPLLOT cohort. TCGA, The Cancer Genome Atlas; KMPLLOT, Kaplan-Meier plotter; TP, true-positive; FP, false-positive; AUROC, area under the ROC curve; ROC, receiver operating characteristic; OS, overall survival.

small hairpin RNA (shRNA)-mediated TRIM28 knocked down. We observed that knocking down TRIM28 led to a decrease in the expression of IRF3 in two lung cancer cell lines, H358 and H520, indicating a regulatory relationship between TRIM28 and IRF3 (Figure 6H). Through the CCK-8 experiment, it was found that compared to negative control cell lines, reducing TRIM28 expression led to decreased cell proliferation (Figure 6I).

Discussion

Owing to inter-/intra-tumor heterogeneity, the prognosis of lung cancer patients is different, even among those with localized carcinoma. In this study, the tumor-node-metastasis (TNM) staging system, in conjunction with the three-gene signature, was applied to predict the OS of lung cancer patients and to design a nomogram. ROC and K-M survival curves indicated that the three-gene signature had

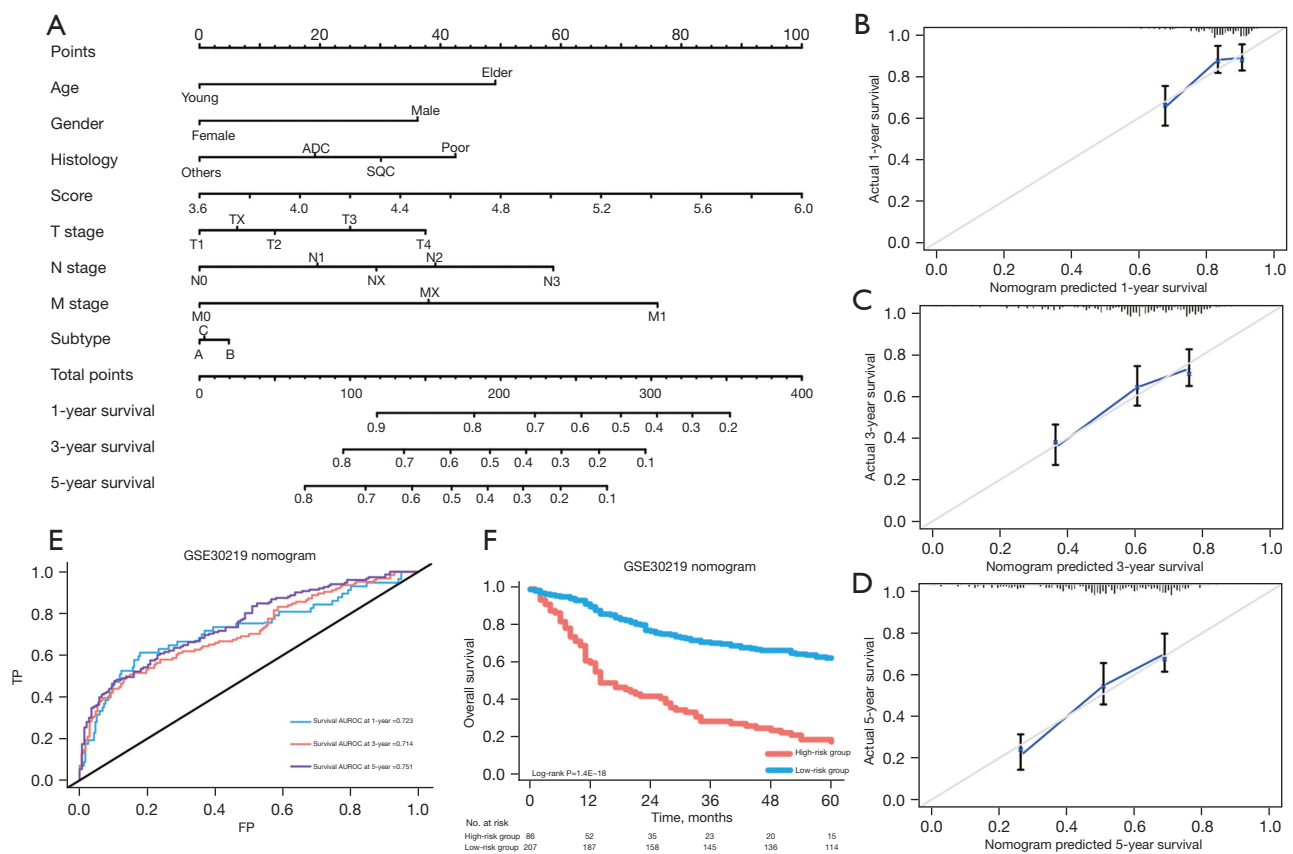


Figure 5 Establishing a prognostic prediction model by combining clinical and transcriptome information. (A) Nomograms convey the results of prognostic models using the three-gene signature and TNM staging system to predict OS of patients with lung cancer. (B) The calibration curve for predicting patients' OS at 1-year. (C) The calibration curve for predicting patients' OS at 3-year. (D) The calibration curve for predicting patients' OS at 5-year. (E) Time-dependent ROC analysis to assess the predictive function of the nomogram at 1, 3, and 5 years. (F) Kaplan-Meier survival curves of OS between high-risk and low-risk patients classified by nomogram in the GSE30219 cohort. ADC, adenocarcinoma; SQC, squamous cell carcinoma; GSE, gene series; AUROC, area under the ROC curve; ROC, receiver operating characteristic; TP, true-positive; FP, false-positive; OS, overall survival; TNM, tumor-node-metastasis.

robust discriminative performance. The predictive value of the nomogram was validated via calibration curve analysis, which demonstrated a high degree of concordance between the predicted OS and the actual observed outcomes. Herein, a new mechanism for histological determination of lung cancer has been proposed, and an innovative method for drug target discovery has been developed.

Previous studies have attempted to discover molecular markers for the prediction of OS and the detection of response to immunotherapy in lung cancer patients. Based on the expression levels of transcription factor families, numerous markers have been found to be correlated with histological classification and prognosis of lung cancer. For instance, the BACH1 transcription factor has been shown

to be associated with an unfavorable prognosis in lung cancer patients (13). Likewise, ASCL1 is a determining transcription factor in small-cell lung cancer and is predictive of a poor prognosis in lung cancer patients (14,15). Prior investigations have described a correlation between the IRF and STAT families and inter-associated epigenetic memory (16). In addition, significant differences in their expression levels between primary and metastatic tumors have been documented, which may not only impact the malignant biological behavior of tumors but also play a decisive role in regulating immune responses (17). This also confirmed our research findings that gene signature based on the IRF and STAT transcription factor families can distinguish lung cancer with different degrees of

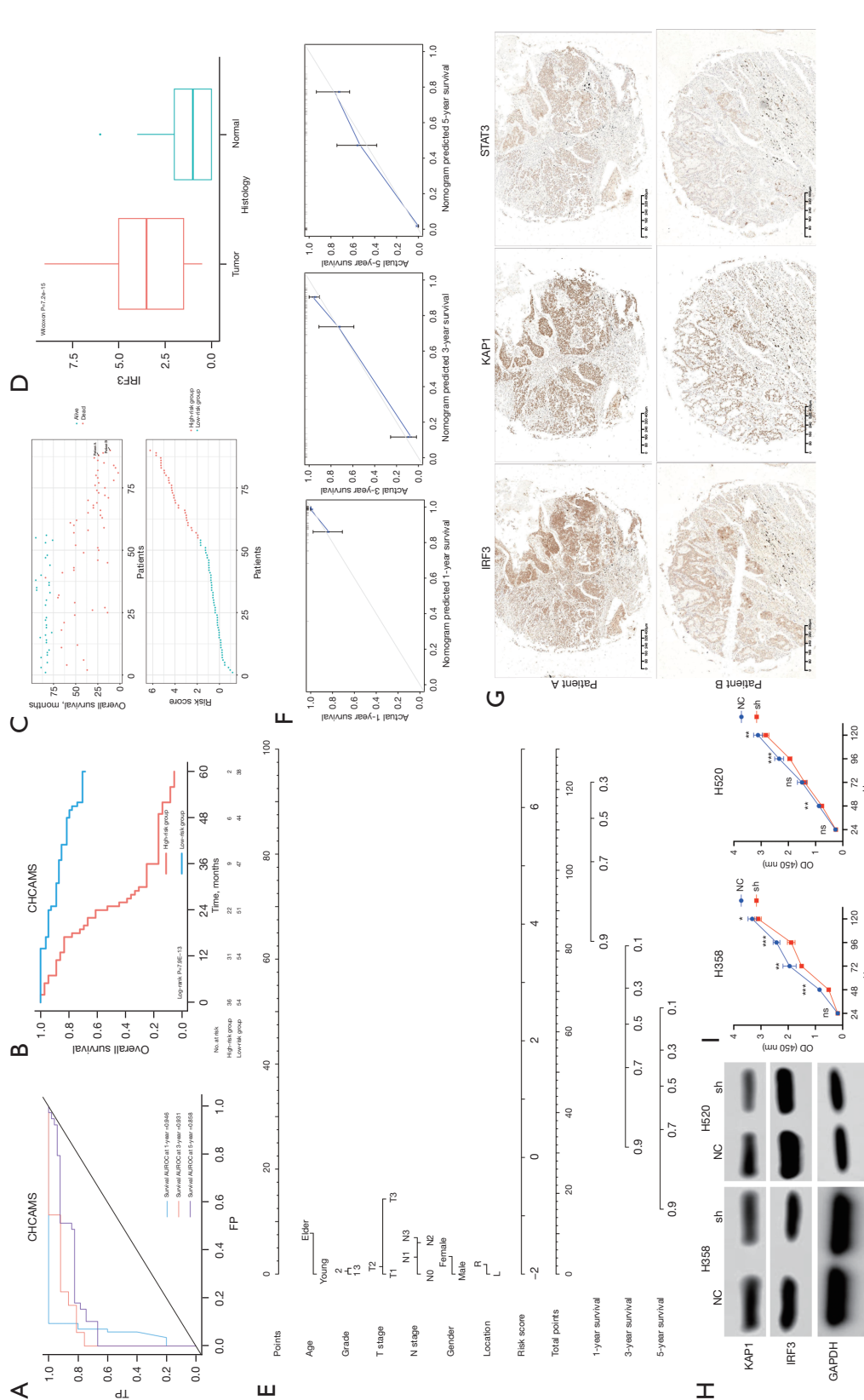


Figure 6 Validation of the predictive performance and molecular biology functions of the three-gene signature in the CHCAMS cohorts. (A) Time-dependent ROC curves of prognostic models constructed by IHC scores at 1, 3, and 5 years. (B) Kaplan-Meier survival curves of OS between high-risk and low-risk patients using TMA. (C) The distributions of the risk score and survival time and status in our cohort. (D) The boxplot of IRF3 expression in cancer and adjacent tissues. (E) Nomogram prediction of OS in lung adenocarcinoma patients based on three-gene signature and clinical features. (F) The calibration curve for predicting patients' OS at 1, 3, and 5 years. (G) Representative IHC images of high-risk group patient A and B. (H) Western blotting analysis of IRF3 and TRIM28 (KAP1) protein levels in two NSCLC cell lines. (I) CCK-8 assay to analyze the proliferative capacity of different cell lines. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$. ns, no significance; CHCAMS, Cancer Hospital Chinese Academy of Medical Sciences; TP, true-positive; FP, false-positive; AUROC, area under the ROC curve; ROC, receiver operating characteristic; IRF, interferon regulatory factor; IHC, immunohistochemistry; OS, overall survival; TMA, tissue microarray; NC, negative control; sh, small hairpin RNA; CCK-8, cell counting kit-8; NSCLC, non-small cell lung cancer; OD, optical density.

malignancy and predict patient prognosis. In addition, an earlier study concluded that TRIM28-induced changes in mouse prostates contributed to invasive prostate carcinoma progression, leading to a shorter OS (18), implying that TRIM28 may potentially play a similar role in lung cancer.

This research has some limitations that merit acknowledgment. To begin, this was a retrospective study; therefore, the possibility of inherent selection bias cannot be excluded. Secondly, further cytological research is warranted to investigate the mechanism by which the expression of TRIM28, IRF3, and STAT3 affect lung cancer differentiation. Lastly, the model necessitates further prospective assessment to corroborate its clinical effectiveness.

Conclusions

In summary, a three-gene signature was developed and validated for predicting the OS of lung cancer patients. The novel nomogram developed based on this signature exhibited sufficient discrimination and calibration capabilities and could serve as a valuable tool for guiding treatment strategies.

Acknowledgments

All authors would like to thank the research groups of TCGA, GEO, and K-M plotter database for providing data for this collection.

Funding: This work was supported by the National Key R&D Plan, the Ministry of Science and Technology of the People's Republic of China (No. 2017YFC1308700).

Footnote

Reporting Checklist: The authors have completed the TRIPOD reporting checklist. Available at <https://jtd.amegroups.com/article/view/10.21037/jtd-24-733/rc>

Data Sharing Statement: Available at <https://jtd.amegroups.com/article/view/10.21037/jtd-24-733/dss>

Peer Review File: Available at <https://jtd.amegroups.com/article/view/10.21037/jtd-24-733/prf>

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at <https://jtd.amegroups.com/article/view/10.21037/jtd-24-733/coif>). The authors

have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

1. Dejima H, Hu X, Chen R, et al. Immune evolution from preneoplasia to invasive lung adenocarcinomas and underlying molecular features. *Nat Commun* 2021;12:2722.
2. Allemani C, Matsuda T, Di Carlo V, et al. Global surveillance of trends in cancer survival 2000-14 (CONCORD-3): analysis of individual records for 37 513 025 patients diagnosed with one of 18 cancers from 322 population-based registries in 71 countries. *Lancet* 2018;391:1023-75.
3. O'Neill LA, Golenbock D, Bowie AG. The history of Toll-like receptors - redefining innate immunity. *Nat Rev Immunol* 2013;13:453-60.
4. Zhao B, Du F, Xu P, et al. A conserved PLPLRT/SD motif of STING mediates the recruitment and activation of TBK1. *Nature* 2019;569:718-22.
5. Danaher P, Warren S, Lu R, et al. Pan-cancer adaptive immune resistance as defined by the Tumor Inflammation Signature (TIS): results from The Cancer Genome Atlas (TCGA). *J Immunother Cancer* 2018;6:63.
6. Nagy Á, Munkácsy G, Győrffy B. Pancancer survival analysis of cancer hallmark genes. *Sci Rep* 2021;11:6047.
7. Győrffy B. Survival analysis across the entire transcriptome identifies biomarkers with the highest prognostic power in breast cancer. *Comput Struct Biotechnol J* 2021;19:4101-9.
8. Yu G, Wang LG, Han Y, et al. clusterProfiler: an R

- package for comparing biological themes among gene clusters. *OMICS* 2012;16:284-7.
9. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 2005;102:15545-50.
 10. Tibshirani R. The lasso method for variable selection in the Cox model. *Stat Med* 1997;16:385-95.
 11. Heagerty PJ, Zheng Y. Survival model predictive accuracy and ROC curves. *Biometrics* 2005;61:92-105.
 12. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* 2016;32:2847-9.
 13. Wiel C, Le Gal K, Ibrahim MX, et al. BACH1 Stabilization by Antioxidants Stimulates Lung Cancer Metastasis. *Cell* 2019;178:330-345.e22.
 14. Gay CM, Stewart CA, Park EM, et al. Patterns of transcription factor programs and immune pathway activation define four major subtypes of SCLC with distinct therapeutic vulnerabilities. *Cancer Cell* 2021;39:346-360.e7.
 15. Megyesfalvi Z, Barany N, Lantos A, et al. Expression patterns and prognostic relevance of subtype-specific transcription factors in surgically resected small-cell lung cancer: an international multicenter study. *J Pathol* 2022;257:674-86.
 16. Qiu J, Xu B, Ye D, et al. Cancer cells resistant to immune checkpoint blockade acquire interferon-associated epigenetic memory to sustain T cell dysfunction. *Nat Cancer* 2023;4:43-61.
 17. Szekely B, Bossuyt V, Li X, et al. Immunological differences between primary and metastatic breast cancer. *Ann Oncol* 2018;29:2232-9.
 18. Yende AS, Williams EC, Pletcher A, et al. TRIM28 promotes luminal cell plasticity in a mouse model of prostate cancer. *Oncogene* 2023;42:1347-59.

Cite this article as: Hong Q, Liang S, Han R, Zhao C, Mu J. Signature stemmed from two transcription factor families determines histological fate and regulates immune infiltration in patients with lung cancer. *J Thorac Dis* 2024;16(9):5663-5674. doi: 10.21037/jtd-24-733