Heliyon 10 (2024) e26974

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

# Identification of immune-related gene signature for non-small cell lung cancer patients with immune checkpoint inhibitors

Li Wang [a], Chaonan Han [a], Chenlei Cai [a], Jing Wu [a], Jianing Chen [a], Chunxia Su [a,b,*]

[a] Department of Medical Oncology, Shanghai Pulmonary Hospital & Thoracic Cancer Institute, Tongji University School of Medicine, Shanghai, 200433, PR China

[b] Department of Clinical Research Center, Shanghai Pulmonary Hospital & Thoracic Cancer Institute, Tongji University School of Medicine, Shanghai, 200433, PR China

ARTICLE INFO

*Keywords:*
Immune-related gene
NSCLC
Immune checkpoint inhibitors
Immune infiltration
Gene signature

ABSTRACT

*Background:* The utilization of immune checkpoint inhibitors (ICIs) has become the established protocol for treating advanced non-small cell lung cancer (NSCLC). This work aimed to identify the immune-related gene signature that can predict the prognosis of NSCLC patients receiving ICI treatment.

*Methods:* The ImmPort database was queried to obtain a list of immune-related genes (IRGs). Differentially expressed IRGs in NSCLC patients were identified using the TCGA database. RNA-seq data and clinical information from NSCLC patients receiving immunotherapy were obtained from the GEO database (GSE93157 and ////). A gene signature was generated through multivariate Cox and LASSO regression analyses. The prognostic value and function of this gene signature were thoroughly investigated using comprehensive bioinformatics analyses.

*Results:* A total of 6 prognostic-related genes were identified from 617 differentially expressed genes, and two prognostic-related differentially expressed genes (CAMP and IL17A) were determined to construct gene signature. Our gene signature demonstrated superior performance compared to other clinicopathological parameters in predicting the prognosis of NSCLC patients receiving immunotherapy, with an area under the ROC curve (AUC) of 0.812. Furthermore, immune infiltration analysis indicated that the high-risk group was enriched with resting CD4 T cell memory, while the low-risk group showed a "hot" tumor microenvironment that promotes anti-tumor immunity in NSCLC patients.

*Conclusion:* Gene signatures based on immune-related genes exhibited excellent indicator performance of prognosis and immune infiltration, which has the potential to be an effective biomarker for NSCLC with ICI treatment.

## 1. Introduction

Nowadays, lung cancer is the second most incidence and the leading cause of death of malignant tumors worldwide among various cancers [1]. At present, ICIs have demonstrated compelling clinical benefits, substantially extending the survival of advanced NSCLC [2–5]. However, the response rate of NSCLC with ICIs is approximately 20%, indicating that there are still most NSCLC who failed to

* Corresponding author. Department of Medical Oncology and Clinical Research Center, Shanghai Pulmonary Hospital & Thoracic Cancer Institute, Tongji University School of Medicine, Shanghai, 200433, PR China.
*E-mail address:* susu_mail@126.com (C. Su).

https://doi.org/10.1016/j.heliyon.2024.e26974
Received 2 June 2023; Received in revised form 31 January 2024; Accepted 22 February 2024
Available online 27 February 2024

respond to ICIs [6]. The identification of potential biomarkers for predicting the efficacy of ICI treatment and identifying the appropriate patient population represents a significant challenge in clinical practice.

Currently, several biomarkers associated with ICIs have gained approval through numerous clinical trials. These include PD-L1 expression, tumor mutational burden (TMB), and microsatellite instability-high (MSI-H) [7,8]. Most clinical trials have revealed heightened response rates associated with increased PD-L1 expression in NSCLC. Nonetheless, the augmentation of responses is not fully comprehensive [8,9]. Relying exclusively on PD-L1 expression proves inadequate to fulfill the precision criteria for clinical applications. The spatial and temporal heterogeneity of PD-L1 expression introduces potential variations between tumor sites, encompassing both primary and metastatic lesions. Moreover, the continuous distribution of PD-L1 expression complicates the establishment of a dependable binary threshold [10]. TMB is acknowledged as an independent biomarker for the response to ICIs in various cancers [11–13]. High TMB is linked to improved survival among patients receiving ICIs. Nevertheless, a universally accepted definition for high TMB is currently lacking [14]. The major challenge of TMB lies in its ability to effectively predict PFS but not OS in the clinical application [15]. TMB faces other challenges such as high detection expenses and relatively long testing durations. Furthermore, specific clinical indicators can also function as predictive factors for the prognosis of immunotherapy in NSCLC. ICI have shown notably lower efficacy in never-smokers, encompassing those with EGFR mutations, ALK rearrangements, and other less common oncogenic drivers of NSCLC [16,17]. In summary, there is still a lack of official biomarkers to guide the ICIs treatment.

With the rapid development of various omics technologies, the acquisition and analysis methods of omics data have gradually become more refined and mature. Various omics data related to cancer, including transcriptomic, proteomic, and metabolomic data, are accumulating rapidly [18]. Among these, transcriptomics has generated a large amount of sequencing data, which is used for studies related to the elucidation of tumor mechanisms and the discovery of novel anti-cancer drugs [19–21]. A gene signature denotes a distinctive arrangement of gene expressions within a cell, occurring either individually or as a coordinated group of genes [22,23]. Utilizing a gene signature allows for a more precise classification of NSCLC, offering valuable guidance for advanced NSCLC with ICIs. Besides that, many components of the immune system are a determining factor during cancer initiation and progression. Evading immune destruction has been recognized as an emerging hallmark of cancer. The prognostic potential of molecular features describing the interplay between tumors and the immune system in NSCLC treated with ICIs has not been comprehensively explored to date [24]. Therefore, it is essential to explore an immune-relate gene signature to predict the prognosis of NSCLC with ICI treatment.

The ImmPort database consists of four elements including private data, shared data, and resources. These components serve for the storage, distribution, analysis, and reutilization of research data. Many studies have been made freely available through the Shared Data portal, facilitating the repurposing of research data to expedite the transformation of novel insights into discoveries [25]. In this study, we aim to identify potential immune-related genes in NSCLC, provide gene signatures to predict the prognosis of NSCLC with ICI treatment, and provide some potentially effective drugs for NSCLC patients with different risk scores. Immune-related gene lists were obtained from the ImmPort database.

## 2. Materials and methods

### 2.1. Data download and pre-processing

GSE93157 included IRGs or signatures of 35 NSCLC patients treated with anti-PD-1 [26]. GSE135222 was included RNA-seq for NSCLC patients with treated ICIs [27,28]. The “Limma” R package is an R software package developed for the analysis of microarray data and RNA-seq data. It provides a set of statistical models and tools for identifying differentially expressed genes, making it one of the commonly used methods for differential gene expression analysis. First, the raw data undergo quality control and preprocessing steps, which include checking data quality, background noise removal, and standardization. Then, the raw data is log-transformed, and the accurate measure of differential expression for each gene is estimated based on the differences between samples. The voom function combines these measures of differences and the change in variance of genes to obtain differential expression data. Finally, visualization of sample distances is achieved. And Kaplan-Meier analysis was performed to obtain IRGs significantly related to the prognosis of NSCLC with ICIs. In this work, we analyzed the differential expression of IRGs between NSCLC and normal tissues by the “Limma” R package, with thresholds of false discovery rate (FDR) $< 0.05$ and a $|\log2\text{-fold change}| > 1$ in the TCGA cohort. By inputting the genes matrix into R software, and running the "clusterProfiler" R package, the corresponding DEGs list and heatmap can be obtained.

### 2.2. Enrichment analysis for survival-related IRGs

Gene Ontology (GO) and KEGG (Kyoto Encyclopedia of Genes and Genomes) databases serve as widely utilized methodologies for delving into the biological processes, cellular components, and molecular functions intricately associated with genetic data [29,30]. The “clusterProfiler” is an R Package for Comparing Biological Themes Among Gene Clusters. The “clusterProfiler” package includes groupGO, enrichGO, and enrichKEGG for conducting classification and enrichment analysis. It classifies genes based on their projection at specific levels of GO and KEGG and performs enrichment analysis using hypergeometric distribution for both GO and KEGG. Additionally, the “clusterProfiler” package provides a function called compareCluster for automatic calculation of enriched functional categories for each gene cluster and offers various visualization methods. By inputting the genes list into R software, and leveraging the "clusterProfiler" R package, corresponding GO and KEGG diagrams can be generated. In this study, we performed GO and KEGG enrichment analyses to obtain functional annotations associated with survival-related IRGs using the "clusterProfiler" R package. A p-value$< 0.05$ was considered as an indicator of statistical significance.

### 2.3. Consensus clustering for survival related IRGs

Based on the survival-related IRGs expression, NSCLC patients with ICI treatment were categorized into different clusters by the "ConsensusClusterPlus" R package. "ConsensusClusterPlus" is an R package designed specifically for performing consensus clustering analysis. After preparing the gene expression matrix, utilizing this R package in the R software allows for the generation of corresponding delta area plots, consensus cumulative distribution functions, and a heatmap of the consensus matrix. We investigated potential differences in survival time among distinct clusters by analyzing Kaplan-Meier curves.

### 2.4. Construction and validation of immune related gene signature

In this work, GSE93157 was divided into a training cohort and a test cohort. The training cohort was used for the construction of the gene signature and the test cohort was used for the internal validation of the gene signature. Then, we employed LASSO regression analysis to remove highly correlated survival-related IRGs. The LASSO regression analysis was performed by "glmnet" R package. "glmnet" is an R package that is designed for fitting various forms of generalized linear models and generalized additive models. It employs an advanced coordinate descent algorithm for fitting linear models with penalty forms such as Lasso or Elastic Net. This algorithm can efficiently handle high-dimensional data, mitigating issues of overfitting. Common applications include prediction, classification, feature selection, and signal discretization, among others. By inputting the gene expression list into R software, and leveraging the " glmnet " R package, corresponding results can be generated. Next, we performed multivariate Cox regression analysis to identify prognostic candidate genes and their regression coefficients, on which gene signatures were constructed. We further generated calibration curves and calculated the AUC to assess the accuracy of our gene signature. After conducting this evaluation, we proceeded to externally validate the gene signature using the independent cohort GSE135222. The risk score for NSCLC patients in the GSE135222 was calculated using the same formula that was derived from the GSE93157.

### 2.5. Independent prognostic analysis and GSEA of gene signature

Independent prognostic analysis was carried out to determine whether the gene signature was an independent risk factor for NSCLC patients. By conducting gene set enrichment analysis (GSEA), we obtained varying levels of entire gene sets through a comparison of risk groups.

### 2.6. Estimation of tumor immune infiltration and drug sensitivity analysis

CIBERSORT, a widely used computational method in tumor immunology research, is capable of quantitatively analyzing the proportions of different immune cell subtypes in tumor tissues using gene expression data. Based on the principle of linear support vector regression, CIBERSORT can estimate the relative abundance of various immune cell subtypes in mixed cell samples by training on the known gene expression profiles of characteristic genes for immune cells [31]. To elucidate the composition of immune cells in NSCLC patients, we employed the "CIBERSORT" R package to convert gene expression data into proportions of immune cell types. This approach allowed us to estimate the infiltration status of immune and stromal cells in each NSCLC sample, providing valuable insights into the immune microenvironment of NSCLC tumors [32]. By running the "CIBERSORT" R package in the R software using the corresponding gene expression matrix that was downloaded from the GEO database, it can generate the corresponding heatmap of immune cell abundance, bar plots, and box plots of immune cells.

"pRRophetic" is an R Package for prediction of clinical response from tumor gene expression [33]. The approach involved constructing statistical models using gene expression and drug sensitivity information gathered from an extensive array of cancer cell lines. Subsequently, these models were applied to analyze gene expression data obtained from primary tumor biopsies. The "pRRophetic" R package utilizes expression data matrices and drug treatment information from the Cancer Genome Project (CGP) database, which encompasses 138 anticancer drugs tested against 727 cell lines. Despite the lack of biomarkers for predicting resistance to ICIs in NSCLC and drug sensitivity, we employed the "pRRophetic" R package to predict the clinical response to chemotherapeutic treatment based on tumor gene expression. In this study, we aim to gain insights into potential predictive factors for therapeutic outcomes in NSCLC patients undergoing ICI treatment.

### 2.7. Genetic alteration and expression analyses of IRGs in gene signature

The cBioPortal for cancer genomics database was employed to detect the mutation frequencies of IRGs in gene signatures. Kaplan–Meier plotter website were used to assess the prognostic value of IRGs. The Human Protein Atlas (HPA) was applied to determine the protein expression level of IRGs. In this study, protein expression analysis of CAMP and IL17A was conducted using the HPA database.

### 2.8. Statistical analyses

Statistical analyses were conducted using GraphPad Prism 9.0 and R 4.2.2 software. Survival analyses were performed employing the Kaplan-Meier plot method and compared using the log-rank test, where a P-value $<0.05$ was deemed statistically significant.

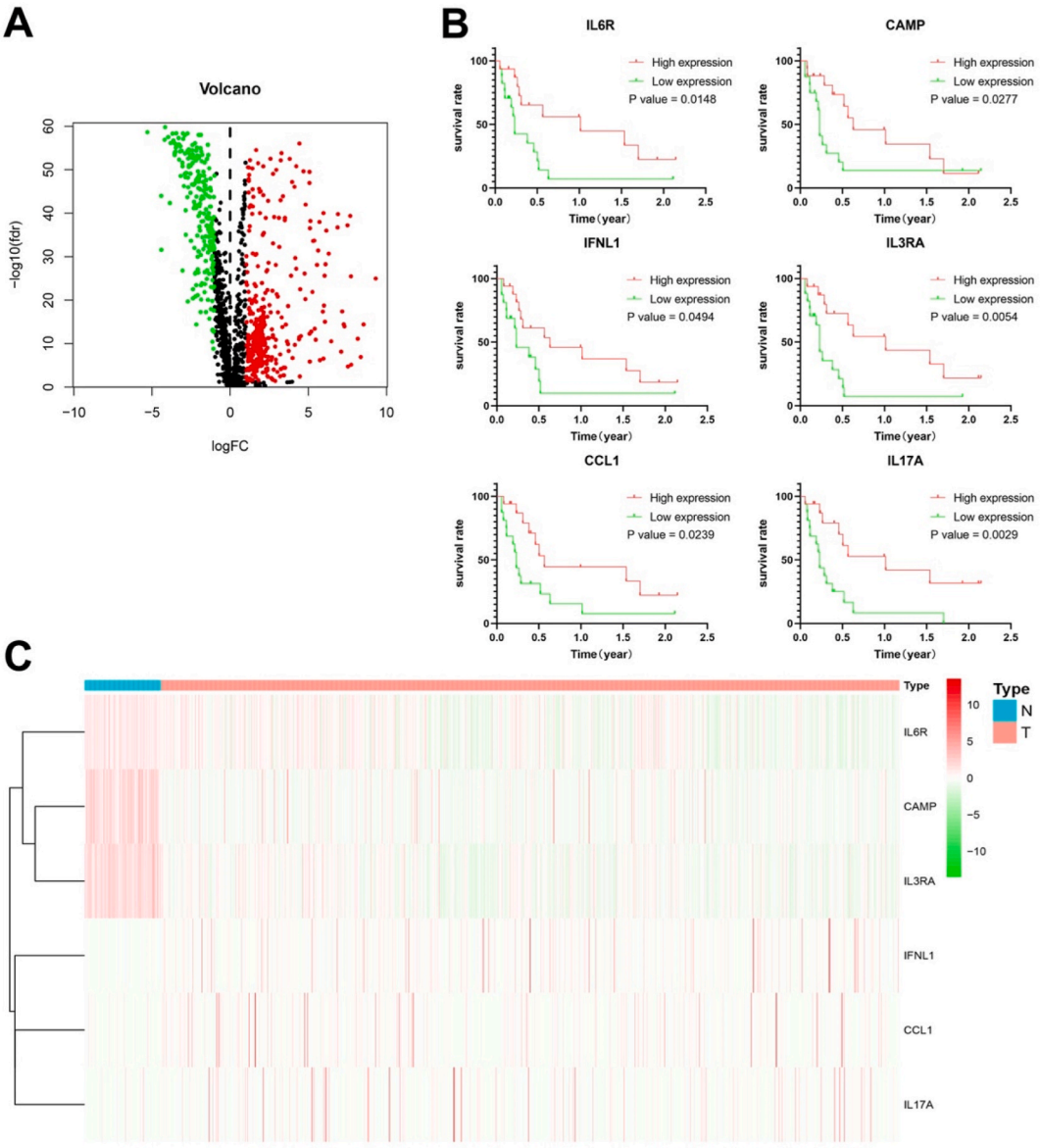


**Fig. 1.** Identification of different expressions of survival-related IRGs. (A) Volcano plots of different expression IRGs in TCGA cohort. (B) IL6R, CAMP, IFNL1, IL3RA, CCL1, and IL17A were positively associated with prognosis in NSCLC patients with ICIs. (C) Heatmap of IL6R, CAMP, IFNL1, IL3RA, CCL1, and IL17A expression in TCGA cohort. Red, higher expression; green, lower expression. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

## 3. Result

### *3.1. Identification of different expression of survival related IRGs*

Total of 2483 Immune-relate genes were downloaded from ImmPort database. Then, 616 different expression IRGs were identified from the TCGA cohort (Extended Data Fig. 1). We also visualized the expression pattern of different expression IRGs in Fig. 1A. Among 616 differently expression IRGs, high expression of IL6R (P = 0.0148), CAMP (P = 0.0277), IFNL1 (P = 0.0494), IL3RA (P = 0.0054), CCL1 (P = 0.0239), and IL17A (P = 0.0029), were positively associated with prognosis in NSCLC patients with ICI **in** Fig. 1B. In addition, as depicted in Fig. 1C**,** IL6R, CAMP and IL3RA exhibited upregulation in normal tissues, while IFNL1, CCL1, IL17A were upregulated in lung cancer tissue. These results suggest that IL6R, CAMP, IFNL1, IL3RA, CCL1, and IL17A may serve as potential biomarkers for advanced NSCLC with ICIs.

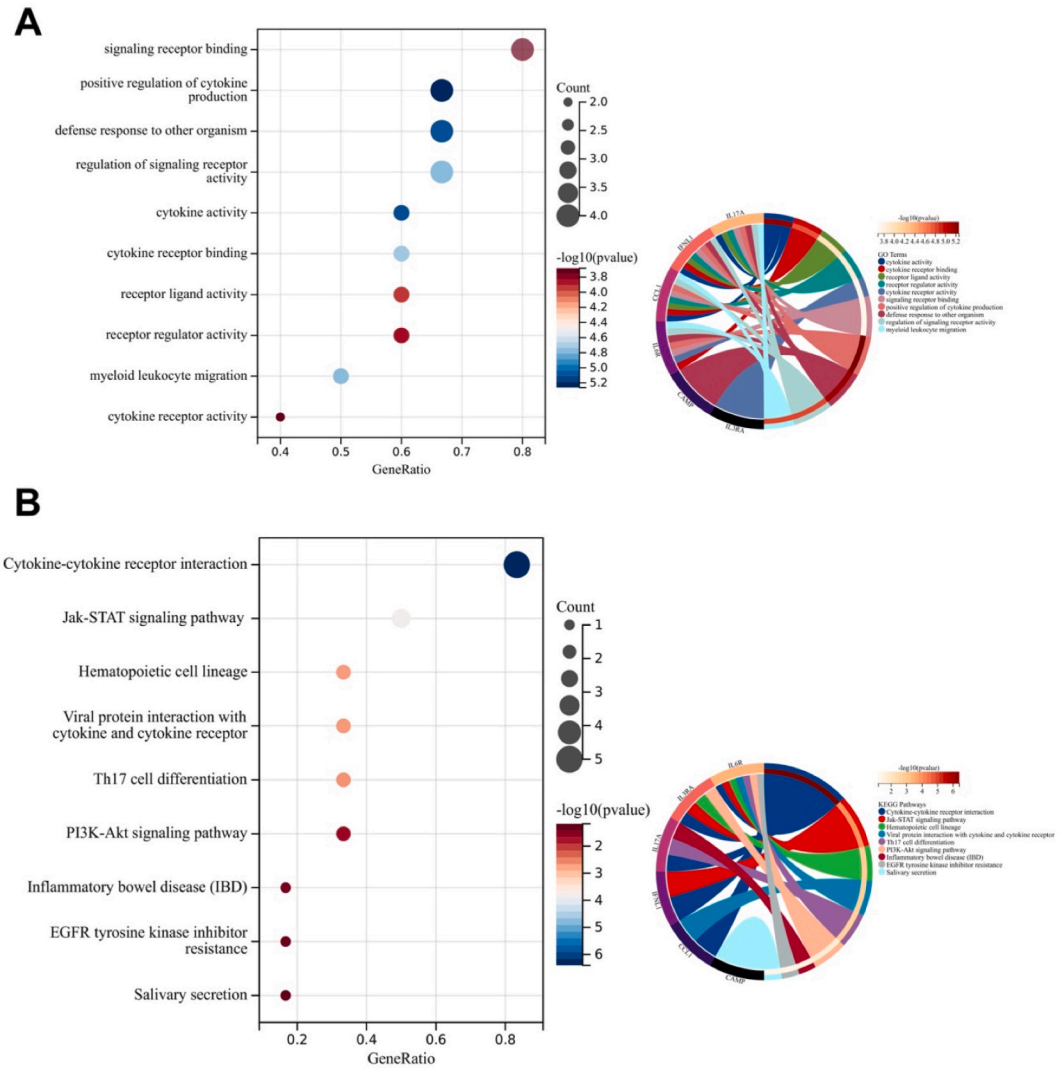


**Fig. 2.** Enrichment analysis. (A) GO analysis. (B) KEGG analysis. The y-axis shows significantly enriched pathways, and the x-axis shows the Rich factor. Rich factor stands for the ratio of the number of target genes belonging to a pathway to the number of all the annotated genes located in the pathway.

### *3.2. Enrichment analysis of survival related IRGs*

The GO analysis results indicated significant enrichment of survival-related IRGs, particularly in processes related to signaling receptor binding, positive regulation of cytokine production, defense response to another organism, and regulation of signaling receptor activity, as illustrated in Fig. 2A. These processes are crucial for coordinating immune responses and is often activated in response to infections, inflammation, or other immune-related challenges. They play a vital role in upholding the body's immune defense mechanisms.

There are four pathways that were particularly enriched in cytokine-cytokine receptor interaction, Jak-STAT signaling pathway, Viral protein interaction with cytokine and cytokine receptor, and TH17 cell differentiation in Fig. 2B. Among these, cytokine-cytokine receptor interaction involves the vital communication between immune cells through signaling molecules called cytokines and their corresponding receptors [25]. These interactions are crucial for coordinating immune responses, regulating inflammation, and maintaining immune system balance. Besides that, the Jak-STAT pathway is a fundamental signaling mechanism used by cytokines and growth factors to activate genes involved in immune responses, cell growth, and differentiation [34]. Viral protein interaction with cytokine and cytokine receptor pathway explores the intricate interactions between viral proteins and components of the cytokine-cytokine receptor system. Understanding these interactions is essential for unraveling viral strategies for replication and immune evasion. TH17 cell differentiation refers to the specialized development of T-helper cells known as TH17 cells [35]. These cells produce the cytokine IL-17 and play a critical role in immune responses against certain pathogens, particularly fungi and bacteria.

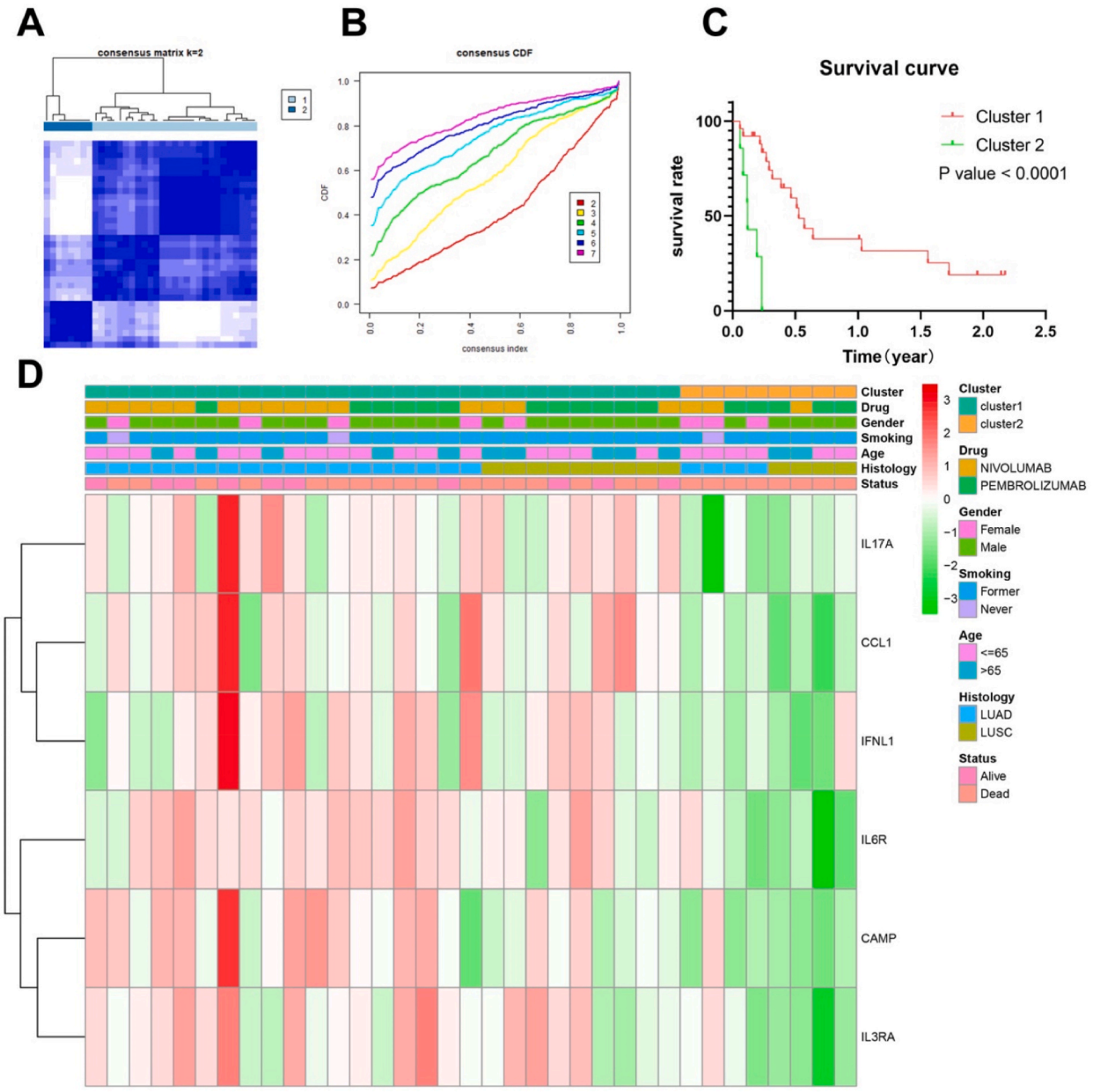


**Fig. 3.** IRG clusters and relevant clinical features. (A, B) Consensus matrices of IRGs in NSCLC for k = 2. (C) Kaplan-Meier curve of the two IRG clusters for NSCLC with ICI treatment. (D) The heatmap of clinical characteristics and gene expression of two clusters.

### *3.3. Identification of IRGs cluster in NSCLC*

Consensus clustering analysis was employed to explore the interactions between IRGs and NSCLC. Subsequently, NSCLC patients were stratified into distinct clusters using the consensus clustering algorithm. Considering the clustering stability and sample size, all NSCLC patients were divided into Cluster 1 and Cluster 2 (Fig. 3AandB). This evidence suggests that our classification approach relying on IRGs is suitable.

To visually depict the intuitive relationship between tumor subtypes and clinical characteristics, we conducted a survival analysis and generated a heatmap illustrating the clinical correlations. The Kaplan-Meier curves of two IRG clusters showed that cluster 1 had the most significant survival advantage in Fig. 3C. Moreover, we also explored the clinicopathological features and expression of IRGs in NSCLC. Among the six genes of the IRGs, IL17A, CCL1, IFNL1, IL6R, CAMP, and IL3RA are highly expressed in cluster 1 (Fig. 3D).

**Table 1**
Clinical characteristics of NSCLC patients in train and test cohorts.

| Clinical characteristics | Train cohorts | Test cohorts | P value |
|---|---|---|---|
| **Total cases** | 23 | 12 | |
| **Histology** | | | 0.4806 |
| LUAD | 13 (56.52%) | 9 (75%) | |
| LUSC | 10 (43.48%) | 3 (25%) | |
| **Age** | | | 0.2985 |
| <65 | 17 (73.91%) | 6 (50%) | |
| ≥65 | 6 (26.09%) | 6 (50%) | |
| **Gender** | | | 1 |
| Male | 18 (78.26%) | 9 (75%) | |
| Female | 5 (21.74%) | 3 (25%) | |
| **Smoking** | | | 0.5487 |
| Former | 22 (95.65%) | 10 (16.67%) | |
| Never | 1 (4.35%) | 2 (83.33%) | |
| **Drug** | | | 0.0971 |
| Nivolumab | 9 (39.13%) | 9 (75%) | |
| Pembrolizumab | 14 (60.87%) | 3 (25%) | |

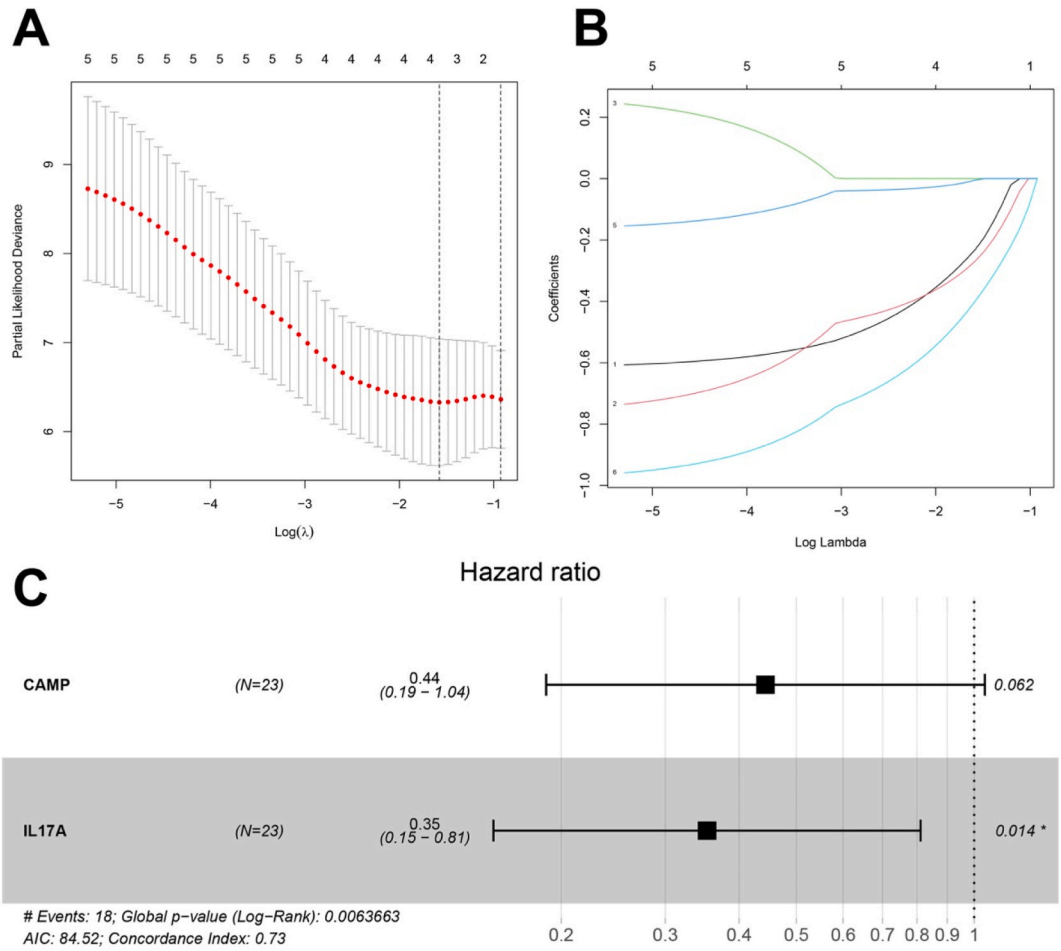


**Fig. 4.** The LASSO regression analysis applied to screening IRGs that are optimally used for the construction of the gene signature. (A) Screening of optimal parameter (lambda) at which the vertical lines were drawn. (B) LASSO coefficient profiles of the 4 IRGs with non-zero coefficients determined by the optimal lambda. (C) Multivariate Cox regression of 2 candidate IRGs genes.

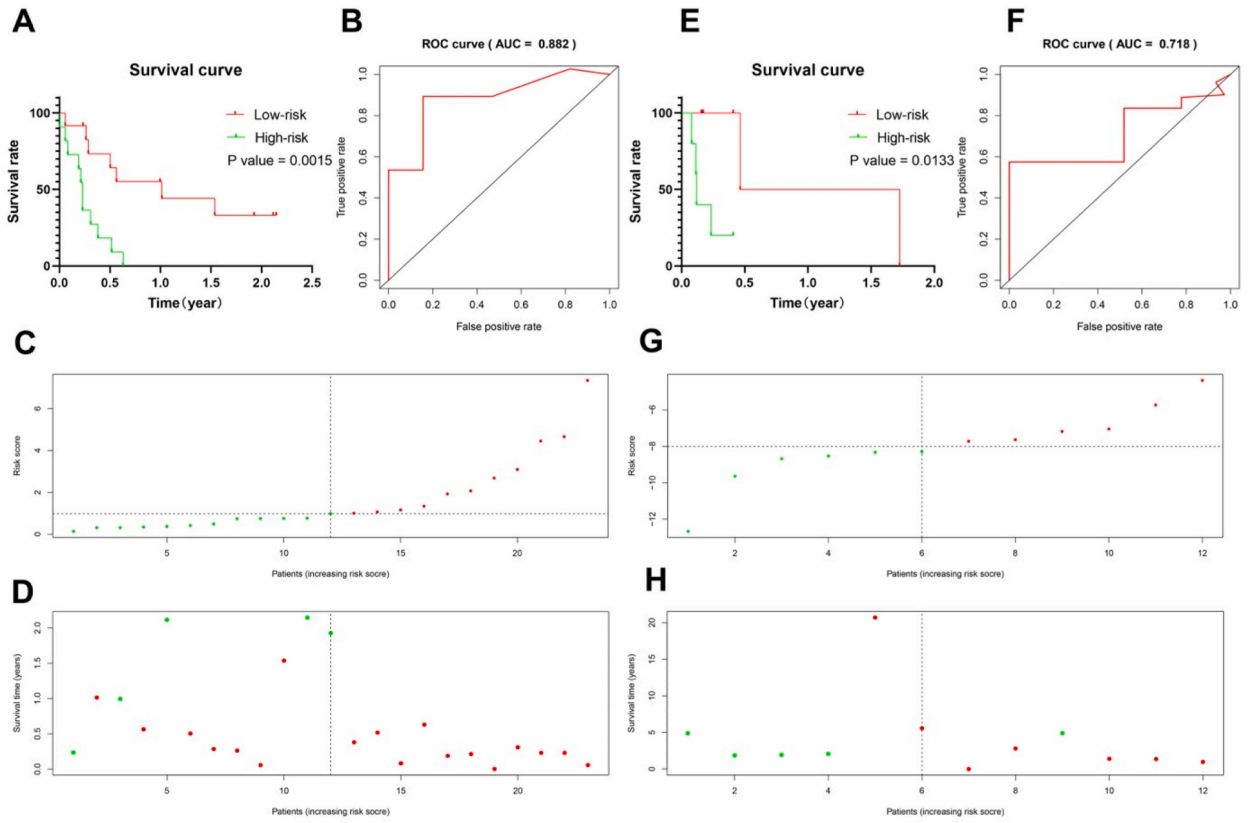


**Fig. 5.** Construction and validation of gene signature. (A) Kaplan–Meier curve of high-risk and low-risk NSCLC patients in the training cohort. (B) ROC curve of OS-related gene signature. (C) Risk score distribution of NSCLC patients with different risks. (D) Scatterplots of NSCLC patients with different survival status. (E–H) Validation of gene signature by test cohort.

### *3.4. Construction and validation of gene signature*

The clinical characteristics of the train cohort and test cohort are summarized in Table 1. In the training cohort, co-expressed IRGs were filtered out using LASSO regression analysis to prevent data overfitting (Fig. 4A - B**)**. Finally, a total of 4 IRGs were selected for inclusion in the multivariate Cox proportional hazards model, and 2 IRGs (CAMP and IL17A) were identified to construct the gene signature (Fig. 4C). The risk score was calculated using the following formula: risk score = CAMP * −0.8134 + PCDH7 * −1.0413.

NSCLC patients were stratified into low-risk and high-risk groups based on the median of risk score. Kaplan-Meier survival analysis revealed that the low-risk group had a significantly better survival rate compared to the high-risk group (P = 0.0015) in Fig. 5A. Besides that, the ROC curve of risk score showed that our gene signature predicts prognosis with satisfactory accuracy (AUC = 0.882) in Fig. 5B. Furthermore, it is apparent from Fig. 5C and D that as the patient's risk score increases, their prognosis deteriorates. In the test cohort, The Kaplan–Meier survival curves also verified the prognostic value of our gene signature (P = 0.0133) in Fig. 5E. We can observe that the results of the test cohort are similar to those of the training cohort from Fig. 5F–H. Additionally, we performed an independent external validation of our gene signature by calculating the risk score for each NSCLC patient in the GSE135222 cohort. The Kaplan–Meier survival curves verified the prognostic value of our gene signature in GSE135222 (P = 0.0216) in Fig. 6A. The validation results confirmed the ability of our gene signature to predict the prognosis of NSCLC with ICI treatment in Fig. 6B–D.

### *3.5. Functional Analyses of gene signature*

We conducted univariate and multivariate independent prognostic analyses to further evaluate the predictive value of our gene signature. As shown in Fig. 7AandB, the uniform findings from the independent prognostic analysis confirm the prognostic relevance of our gene signature for NSCLC with ICIs, independent of other clinical factors. Compared with other clinical information, the ROC curve of risk score showed that our gene signature predicts prognosis with satisfactory accuracy (AUC = 0.812) in Fig. 7C. In addition, the GSEA enrichment analysis performed with the GSE135222 cohort showed that Notch signaling pathway, endocytosis, and MAPK signaling pathway were significantly enriched in a high-risk group (Fig. 7D). The Notch signaling pathway is a highly conserved cellular communication system present in the majority of multicellular organisms [36]. It regulates various cellular processes, including cell fate determination, differentiation, and apoptosis. It plays a crucial role in development, tissue homeostasis, and disease.

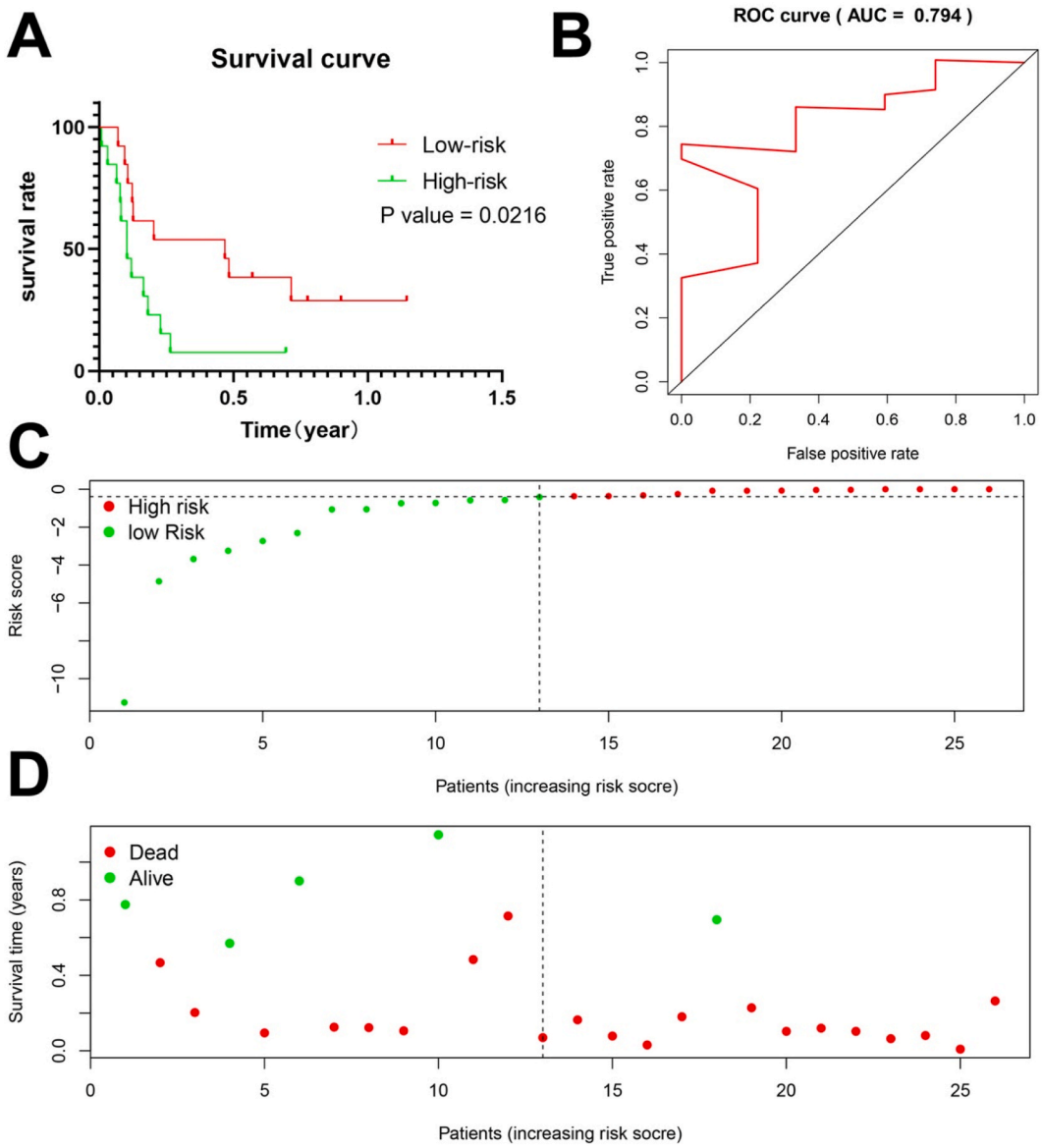


**Fig. 6.** Externally validated of gene signature using independent cohort GSE135222. (A) Kaplan–Meier curve of high-risk and low-risk NSCLC patients in GSE135222 cohort. (B) ROC curve of OS-related gene signature. (C) Risk score distribution of NSCLC patients with different risks. (D) Scatterplots of NSCLC patients with different survival status.

It involves the formation of vesicles that enclose the material and transport it into the cell. Endocytosis is essential for nutrient uptake, receptor recycling, and the regulation of signaling pathways. The MAPK signaling pathway forms a complex network of proteins that transmit signals from the cell surface to the nucleus, coordinating a variety of cellular processes, such as cell proliferation, differentiation, and survival [37]. It is activated in response to external stimuli, including growth factors, stress, and cytokines, and plays a critical role in cell responses to environmental changes.

### *3.6. Immune infiltration and drug sensitivity analysis of gene signature*

We evaluated the correlation between gene signature and tumor-infiltrating immune cells in NSCLC using the CIBERSORT algorithm. As shown in Fig. 8A, we found that resting CD4 T cell memory was more enriched in a high-risk group. We also found that individual genes of gene signature were also strongly associated with immune infiltration in Fig. 8B. For example, CD8 T cell and Neutrophils were negatively associated with risk score in Fig. 8C and D. In particular, we found that resting CD4 T cell memory was positively associated with risk score in Fig. 8E. As the risk score increases, the abundance of resting CD4 T cell memory also increases. Besides, we assessed the immunity status of NSCLC patients in different risk groups. As shown in Fig. 8F, the immune score was

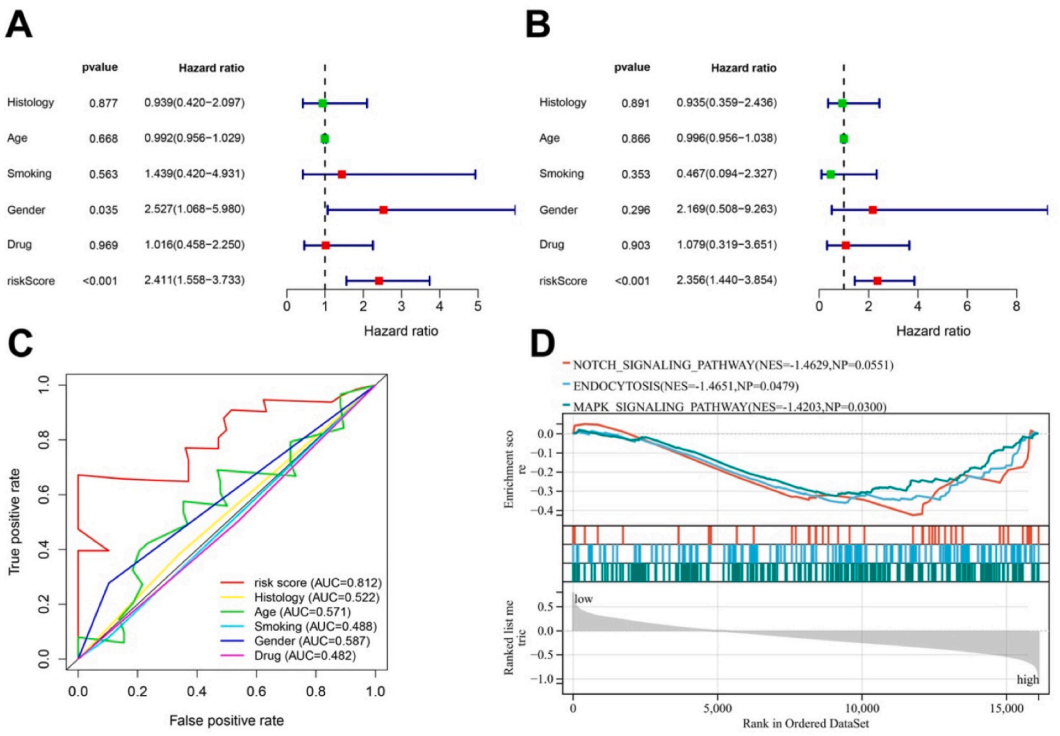


**Fig. 7.** Functional Analyses of gene signature. (A) Univariate factor independent prognostic analysis. (B) Multivariate factor independent prognostic analysis. (C) ROC curve of gene signature. (D) GSEA enrichment analysis of gene signature.

considerably greater in the low-risk group. The low-risk group of NSCLC patients exhibited higher immune scores and better responses to immunotherapy, likely due to increased enrichment of CD8 T cells and neutrophils in the tumor microenvironment (TME). These findings suggest that patients in the low-risk group may have a "hot" TME characterized by enhanced anti-tumor immunity in NSCLC patients.

In summary, our findings indicate that NSCLC patients with high-risk groups may not respond to ICI treatment. To know which treatment could be the personalized treatment of high-risk group NSCLC patients, we measure the sensitivity of drugs between two risk groups by drug sensitivity analysis. Interestingly, NSCLC patients with high-risk had higher sensitivity responses with PF-4708671 ($P = 0.0091$), and AZD8186 ($P = 0.0091$). And low-risk group is more sensitive to UMI-77 ($P = 0.00066$) in Fig. 8G–I. Those results indicated that PF-4708671 and AZD8186 may be potential treatment options for NSCLC patients with high-risk groups.

### *3.7. Genetic alteration, expression, and prognosis value of IRGs*

Next, we investigate the genetic alteration, expression profiles, and prognostic value of key IRGs in other databases. As shown in Fig. 9A, the mutation frequencies of CAMP and IL17A were 0.7% and 2.7% in the cBioPortal for cancer genomics database. The GEPIA database box plots results showed that the expression of CAMP was significantly higher expression in normal samples compared with cancer tissue in Fig. 9B. However, there is no significant difference in IL17A expression between cancer and normal tissue in Fig. 9C. Besides, these genes were not associated with the prognosis of NSCLC in Fig. 9DandE. In addition, we further verify the histological level of CAMP and IL17A by the HPA database in Fig. 9F and G. The results showed that the protein level of CAMP is upregulated in lung cancer tissues and downregulated in normal tissue. And IL17A is no significant difference in protein expression between cancer and normal tissue.

## 4. Discussion

Currently, both PD-L1 expression and TMB have not exhibited sufficient efficacy in NSCLC who are likely to respond to ICIs. In addition to PD-L1 expression and TMB, certain gene mutations, including KRAS/TP53, STK11, EGFR, and EPHA, have been acknowledged for their influence on the effectiveness of ICIs by modulating the tumor microenvironment [38–41]. For example, Sun et al. identified ARID1A mutation as a potential predictive biomarker for the prognosis of ICIs [42]. Additionally, several studies have indicated that mutations in ERBB4 and FGFR4 may function as potential biomarkers for predicting the prognosis of NSCLC with immunotherapy [43,44]. Nevertheless, the validation of these biomarkers necessitates large-scale, multicenter prospective clinical studies. Depending on a single biomarker may not completely satisfy current clinical requirements. In summary, there is a need for

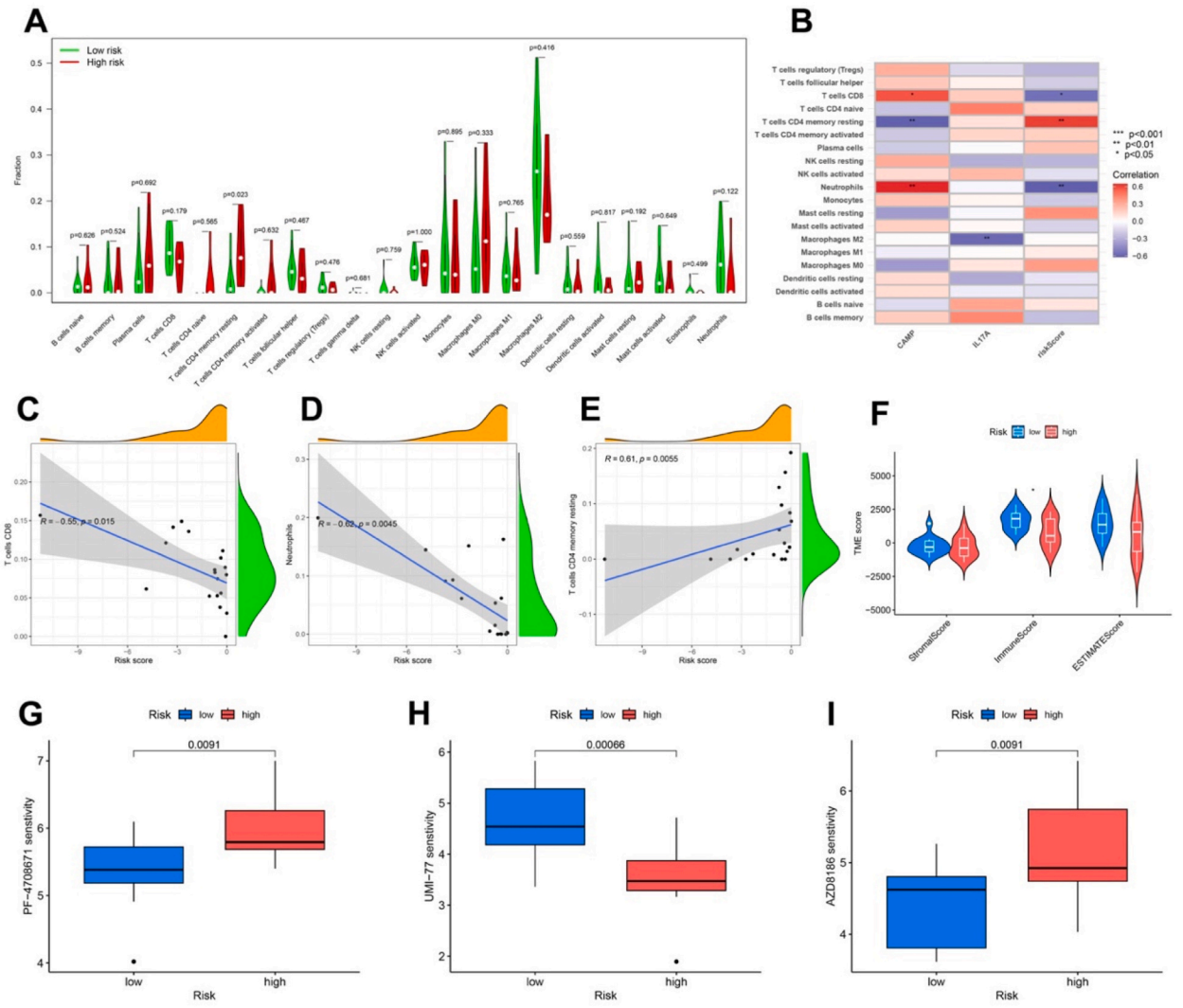


**Fig. 8.** Immune infiltration and Drug sensitivity analysis of gene signature. (A) The composition of 22 immune cell types in the high-risk group and the low-risk group. (B) The relation between IRGs and immune infiltration. As the risk score increases, the abundance of CD8 T cells (C) and neutrophils (D) decreases, while the abundance of resting CD4 T cell memory (E) increases. (F) Immunity status of NSCLC patients in different risk groups. (G) The estimated IC50 of PF-4708671 in the low-risk group was lower than in the high-risk group. (H) The estimated IC50 of UMI-77 in the high-risk group was lower than in the low-risk group. (I) The estimated IC50 of AZD8186 in the low-risk group was lower than in the high-risk group. *, $p < 0.05$; ***, $p < 0.001$.

ongoing improvement in the accuracy of existing biomarkers. The development of gene signatures for tumor prediction and prognosis allows for a comprehensive exploration of the molecular characteristics of tumors, providing substantial support for the advancement of precision medicine. Therefore, it is crucial to explore gene signatures that can predict immunotherapy response with higher precision.

In this study, we screened out IRGs that are associated with the prognosis of NSCLC patients with ICI treatment from GSE93157 and GSE135222. Then, we constructed gene signatures by CAMP and IL17A to predict the prognosis of NSCLC with ICI treatment. The gene signature was verified by the test cohort and GSE135222 cohort with the same risk score formula. Compared with other studies gene signatures, our gene signature performed with satisfactory accuracy with an AUC of 0.812. CAMP and IL17A could serve as potential biomarkers for the sensitivity to ICI treatment and prognosis of NSCLC. Therefore, we reasonably deduced that our gene signature has good predictive value in clinical practice for NSCLC with ICI treatment.

To explain why low-risk groups were associated with favorable ICI prognosis, we evaluated the role of gene signature in immune infiltration. We found that resting CD4 T cell memory was more enriched in the high-risk group. Memory $CD4^+$ T cells are distributed throughout the body, providing protection against reinfection and cancer. They are also involved in processes related to allergies, autoimmune disorders, transplant rejection, and chronic inflammation [45]. Moreover, CD8 T cell and Neutrophils were negatively associated with risk score. Resting CD4 T cell memory was positively associated with risk score. Similarly, the immune score was

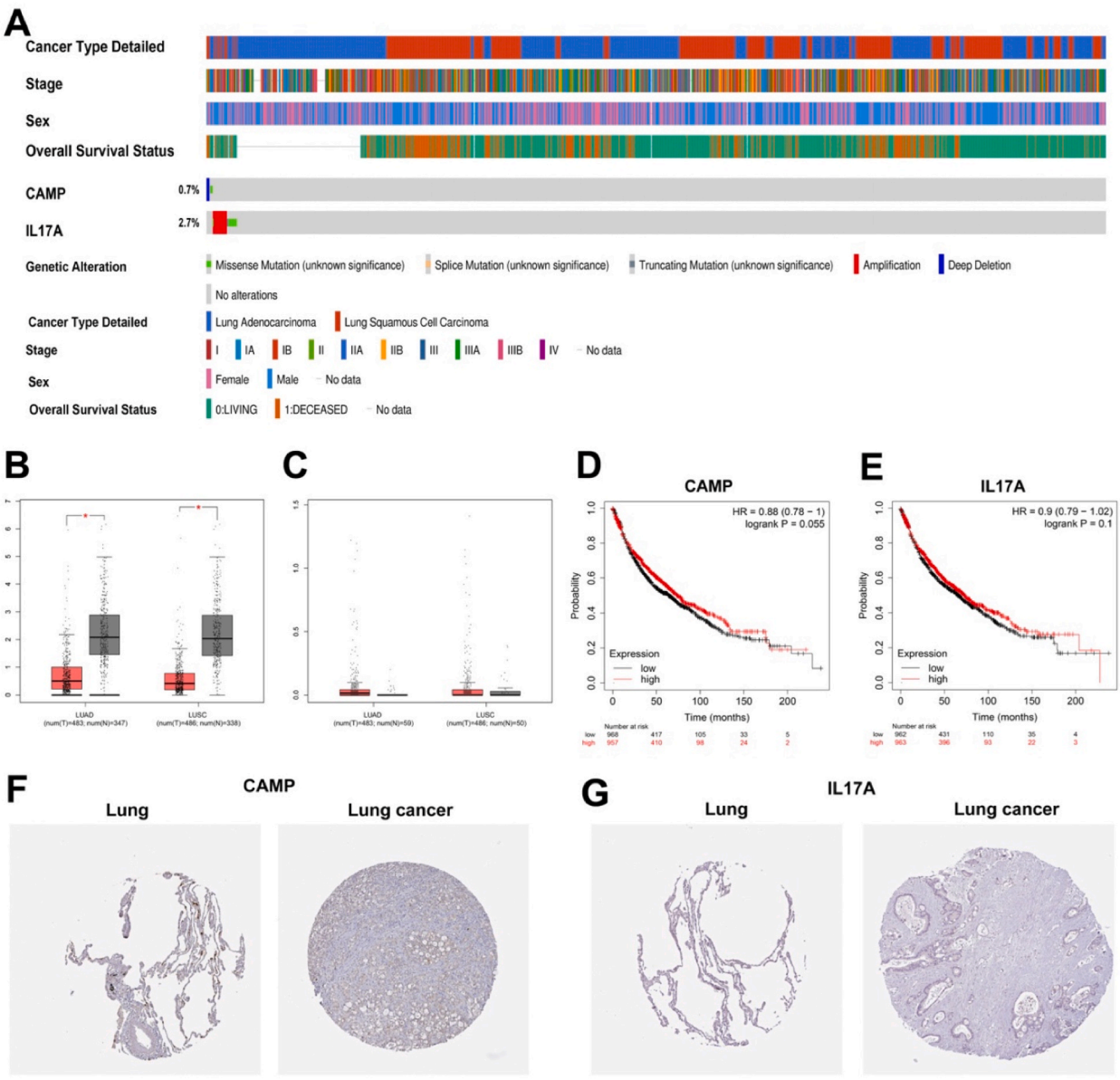


**Fig. 9.** Genetic alteration, expression, and prognosis value of the key gene. (A) The frequencies of gene alterations of CAMP and IL17A. (B, C) The expression level of CAMP and IL17A. (D, E) The prognosis value of CAMP and IL17A. (F, G) The protein level of CAMP and IL17A among normal lung tissues and NSCLC tissues.

considerably greater in patients with low-risk scores, suggesting that patients with low-risk scores may respond better to ICIs. Due to more CD8 T cells and Neutrophils enriched in TME, NSCLC patients in the low-risk group had higher immune scores and better responses to immunotherapy. Therefore, we reasonably deduction suggests that CAMP and IL17A may exert an influence on resting CD4 T cell memory, CD8 T cells, and neutrophils, potentially fostering the development of a "hot" TME characterized by heightened anti-tumor immunity in NSCLC.

Due to NSCLC patients in high-risk groups not responding to ICI treatment, to understand which treatment could be personalized treatment for high-risk groups, we measure the sensitivity of several drugs between two risk groups. Interestingly, NSCLC patients with high risk had higher sensitivity response with PF-4708671 ($P = 0.0091$), and AZD8186 ($P = 0.0091$). PF-4708671, an effective S6K1 inhibitor, has been shown to significantly impede the proliferation and invasion of A549, SK-MES-1, and NCI–H460 cells in vitro [46]. AZD8186, a potent inhibitor targeting both PIK3CB and PIK3CD, is presently undergoing Phase I clinical trials for the treatment of advanced solid tumors, including NSCLC [47]. Therefore, our findings suggest that PF-4708671 and AZD8186 could serve as promising therapeutic alternatives for NSCLC patients in the high-risk group.

The current literature indicates that the mechanism of CAMP and IL17A still requires further exploration in NSCLC. This study

represents the first attempt to elucidate the roles of CAMP and IL17A as potential biomarkers influencing the sensitivity to ICI treatment and prognosis in NSCLC. CAMP and IL17A have the potential to serve as biomarkers involved in modulating the TME and correlating with the prognosis in advanced NSCLC treated with ICI. This study represents the inaugural utilization of an extensive database to formulate gene signatures for prognostic predictions in NSCLC patients undergoing ICI. Undoubtedly, this presents a novel clinical strategy for managing NSCLC patients. However, there are still certain limitations in our study. First, the clinical sample sizes of GSE93157 and GSE13522 were insufficient. Secondly, given the relatively limited number of gene markers, there is a potential risk of overfitting associated with the gene signature. Moreover, this study still lacks real-world validation, including validation in clinical cohorts and animal experiments. Given the retrospective nature of the analysis, there exists the potential for selection bias within the utilized patient cohort. In the future, we will further enhance our gene signatures, encompassing prospective validation of these markers in clinical trials and investigating their predictive potential for other ICIs.

## 5. Conclusion

To conclude, our study demonstrates that our gene signatures based on immune-related genes exhibited excellent indicator performance of prognosis and immune infiltration, which has the potential to be an effective biomarker for NSCLC with ICI treatment.

## Funding

This study was supported by National Natural Science Foundation of China (grant numbers: 82072568, 82373320), the Excellent Academic Leader of Shanghai "Science and Technology Innovation Action Plan" (grant number: 22XD1402500), the Shanghai Shenkang Hospital Development Center (grant number: SHDC12020110, and Shanghai Shenkang development research physician project (grant number: SHDC2022CRD048)

## Ethics approval and consent to participate

Not applicable.

## Consent for publication

Not applicable.

## Data availability

All data included in this study are available including GEO database (GSE93157 and GSE135222), TCGA database, and cBioPortal for cancer genomics database.

## CRediT authorship contribution statement

**Li Wang:** Writing – original draft, Validation, Software, Investigation. **Chaonan Han:** Writing – original draft, Formal analysis. **Chenlei Cai:** Visualization, Resources, Conceptualization. **Jing Wu:** Formal analysis, Data curation. **Jianing Chen:** Validation, Software, Data curation. **Chunxia Su:** Writing – review & editing, Supervision, Funding acquisition.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

Not applicable.

## References

[1] K. Miller, et al., Cancer treatment and survivorship statistics, 2019, CA A Cancer J. Clin. (2019) 69.
[2] T.S.K. Mok, et al., Pembrolizumab versus chemotherapy for previously untreated, PD-L1-expressing, locally advanced or metastatic non-small-cell lung cancer (KEYNOTE-042): a randomised, open-label, controlled, phase 3 trial, Lancet 393 (10183) (2019) 1819–1830.
[3] M. Reck, et al., Five-year outcomes with Pembrolizumab versus chemotherapy for metastatic non-small-cell lung cancer with PD-L1 tumor proportion score ≥ 50, J. Clin. Oncol. 39 (21) (2021) 2339–2349.
[4] S. Gadgeel, et al., Updated analysis from KEYNOTE-189: pembrolizumab or placebo plus pemetrexed and platinum for previously untreated metastatic nonsquamous non-small-cell lung cancer, J. Clin. Oncol. 38 (14) (2020) 1505–1517.
[5] S. Novello, et al., Pembrolizumab plus chemotherapy in squamous non-small-cell lung cancer: 5-year update of the phase III KEYNOTE-407 study, J. Clin. Oncol. (2023) Jco2201990.

[6] Y.J. Park, D.S. Kuen, Y. Chung, Future prospects of immune checkpoint blockade in cancer: from response prediction to overcoming resistance, Exp. Mol. Med. 50 (8) (2018) 1–13.
[7] D.T. Le, et al., Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade, Science 357 (6349) (2017) 409–413.
[8] M. Reck, et al., Pembrolizumab versus chemotherapy for PD-L1-positive non-small-cell lung cancer, N. Engl. J. Med. 375 (19) (2016) 1823–1833.
[9] E.B. Garon, et al., Pembrolizumab for the treatment of non-small-cell lung cancer, N. Engl. J. Med. 372 (21) (2015) 2018–2028.
[10] M. Mino-Kenudson, et al., Predictive biomarkers for immunotherapy in lung cancer: perspective from the international association for the study of lung cancer pathology committee, J. Thorac. Oncol. 17 (12) (2022) 1335–1354.
[11] A.M. Goodman, et al., Tumor mutational burden as an independent predictor of response to immunotherapy in diverse cancers, Mol Cancer Ther 16 (11) (2017) 2598–2608.
[12] N.A. Rizvi, et al., Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer, Science 348 (6230) (2015) 124–128.
[13] R.M. Samstein, et al., Tumor mutational load predicts survival after immunotherapy across multiple cancer types, Nat. Genet. 51 (2) (2019) 202–206.
[14] D.L. Jardim, et al., The challenges of tumor mutational burden as an immunotherapy biomarker, Cancer Cell 39 (2) (2021) 154–173.
[15] A. Addeo, et al., TMB or not TMB as a biomarker: that is the question, Crit. Rev. Oncol. Hematol. 163 (2021) 103374.
[16] T.A. Chan, et al., Development of tumor mutation burden as an immunotherapy biomarker: utility for the oncology clinic, Ann. Oncol. 30 (1) (2019) 44–56.
[17] U. Dafni, et al., Immune checkpoint inhibitors, alone or in combination with chemotherapy, as first-line treatment for advanced non-small cell lung cancer. A systematic review and network meta-analysis, Lung Cancer 134 (2019) 127–140.
[18] R. Shen, A.B. Olshen, M. Ladanyi, Integrative clustering of multiple genomic data types using a joint latent variable model with application to breast and lung cancer subtype analysis, Bioinformatics 25 (22) (2009) 2906–2912.
[19] J.H. Finger, et al., The mouse gene expression database: new features and how to use them effectively, Genesis 53 (8) (2015) 510–522.
[20] S.A. Forbes, et al., COSMIC: exploring the world's knowledge of somatic mutations in human cancer, Nucleic Acids Res. 43 (2015) D805–D811. Database issue.
[21] J. Barretina, et al., The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity, Nature 483 (7391) (2012) 603–607.
[22] H. Itadani, S. Mizuarai, H. Kotani, Can systems biology understand pathway activation? Gene expression signatures as surrogate markers for understanding the complexity of pathway activation, Curr Genomics 9 (5) (2008) 349–360.
[23] J. Liu, et al., Identification of a gene signature in cell cycle pathway for breast cancer prognosis using gene expression profiling data, BMC Med Genomics 1 (2008) 39.
[24] B. Li, et al., Development and validation of an individualized immune prognostic signature in early-stage nonsquamous non-small cell lung cancer, JAMA Oncol. 3 (11) (2017) 1529–1537.
[25] C. Plata-Salamán, N. Turrin, Cytokine interactions and cytokine balance in the brain: relevance to neurology and psychiatry, Mol Psychiatry 4 (4) (1999) 302–306.
[26] A. Prat, et al., Immune-related gene expression profiling after PD-1 blockade in non-small cell lung carcinoma, head and neck squamous cell carcinoma, and melanoma, Cancer Res. 77 (13) (2017) 3540–3550.
[27] H. Jung, et al., DNA methylation loss promotes immune evasion of tumours with high mutation and copy number load, Nat. Commun. 10 (1) (2019) 4278.
[28] J.Y. Kim, J.K. Choi, H. Jung, Genome-wide methylation patterns predict clinical benefit of immunotherapy in lung cancer, Clin Epigenetics 12 (1) (2020) 119.
[29] M. Ashburner, et al., Gene ontology: tool for the unification of biology. The Gene Ontology Consortium, Nat. Genet. 25 (1) (2000) 25–29.
[30] M. Kanehisa, et al., KEGG as a reference resource for gene and protein annotation, Nucleic Acids Res. 44 (D1) (2016) D457–D462.
[31] A.M. Newman, et al., Robust enumeration of cell subsets from tissue expression profiles, Nat. Methods 12 (5) (2015) 453–457.
[32] K. Yoshihara, et al., Inferring tumour purity and stromal and immune cell admixture from expression data, Nat. Commun. 4 (2013) 2612.
[33] P. Geeleher, N. Cox, R.S. Huang, pRRophetic: an R package for prediction of clinical chemotherapeutic response from tumor gene expression levels, PLoS One 9 (9) (2014) e107468.
[34] X. Hu, et al., The JAK/STAT signaling pathway: from bench to clinic, Signal Transduct Target Ther 6 (1) (2021) 402.
[35] K. Yasuda, Y. Takeuchi, K. Hirota, The pathogenicity of Th17 cells in autoimmune diseases, Semin. Immunopathol. 41 (3) (2019) 283–297.
[36] B. Zhou, et al., Notch signaling pathway: architecture, disease, and therapeutics, Signal Transduct Target Ther 7 (1) (2022) 95.
[37] L. Santarpia, S.M. Lippman, A.K. El-Naggar, Targeting the MAPK-RAS-RAF signaling pathway in cancer therapy, Expert Opin. Ther. Targets 16 (1) (2012) 103–119.
[38] Z.Y. Dong, et al., Potential predictive value of TP53 and KRAS mutation status for response to PD-1 blockade immunotherapy in lung adenocarcinoma, Clin. Cancer Res. 23 (12) (2017) 3012–3024.
[39] J. Biton, et al., TP53, STK11, and EGFR mutations predict tumor immune profile and the response to anti-PD-1 in lung adenocarcinoma, Clin. Cancer Res. 24 (22) (2018) 5710–5723.
[40] H. Bai, et al., EPHA mutation as a predictor of immunotherapeutic efficacy in lung adenocarcinoma, J Immunother Cancer 8 (2) (2020).
[41] K. Zhang, et al., Identification of deleterious NOTCH mutation as novel predictor to efficacious immunotherapy in NSCLC, Clin. Cancer Res. 26 (14) (2020) 3649–3661.
[42] D. Sun, et al., Subunits of ARID1 serve as novel biomarkers for the sensitivity to immune checkpoint inhibitors and prognosis of advanced non-small cell lung cancer, Mol Med 26 (1) (2020) 78.
[43] X. Hu, et al., The role of ERBB4 mutations in the prognosis of advanced non-small cell lung cancer treated with immune checkpoint inhibitors, Mol Med 27 (1) (2021) 126.
[44] L. Wang, et al., Development of nomogram based on immune-related gene FGFR4 for advanced non-small cell lung cancer patients with sensitivity to immune checkpoint inhibitors, J. Transl. Med. 19 (1) (2021) 22.
[45] M. Künzli, D. Masopust, CD4(+) T cell memory, Nat. Immunol. 24 (6) (2023) 903–914.
[46] Z.X. Qiu, et al., The p70S6K specific inhibitor PF-4708671 impedes non-small cell lung cancer growth, PLoS One 11 (1) (2016) e0147185.
[47] A.D. Choudhury, et al., A phase I study investigating AZD8186, a potent and selective inhibitor of PI3Kβ/δ, in patients with advanced solid tumors, Clin. Cancer Res. 28 (11) (2022) 2257–2269.