Seven interferon gamma response genes serve as a prognostic risk signature that correlates with immune infiltration in lung adenocarcinoma

Boyang Yao[1,2,*], Lixin Wang[4,*], Heyong Wang[2,3,*], Jinxia Bao[2,3,*], Qiwen Li[2,3], Fengzhi Yu[4], Wenjing Zhu[1,2], Li Zhang[3], Wang Li[5], Zhan Gu[4], Ke Fei[2], Peng Zhang[2], Fan Zhang[2,3], Xiaoying Huang[1]

[1]Division of Pulmonary Medicine, The First Affiliated Hospital of Wenzhou Medical University, Key Laboratory of Heart and Lung, Wenzhou 325000, Zhejiang, China
[2]Department of Thoracic Surgery, Shanghai Pulmonary Hospital, Tongji University School of Life Science and Technology, Shanghai 200433, China
[3]Clinical Translational Research Center, Shanghai Pulmonary Hospital, Tongji University School of Medicine, Shanghai 200433, China
[4]Department of Traditional Chinese Medicine, Shanghai Pulmonary Hospital, Tongji University School of Medicine, Shanghai 200433, China
[5]College of Biological and Environmental Engineering, Binzhou University, Binzhou 256600, Shandong, China
*Equal contribution

Correspondence to: Xiaoying Huang, Fan Zhang, Peng Zhang; **email:** huangxiaoying@wzhospital.cn, fzhang@tongji.edu.cn, zhangpeng1121@tongji.edu.cn

ABSTRACT

Interferon-gamma (IFN-γ) plays a complex role in modulating tumor microenvironment during lung adenocarcinoma (LUAD) development. In order to define the role of IFN-γ response genes in LUAD progression, we characterized the gene expression, mutation profile, protein-protein interaction of 24 IFN-γ response genes, which exhibited significant hazard ratio in overall survival. Two subgroups of LUAD from the TCGA cohort, which showed significant difference in the survival rate, were identified based on the expression of these genes. Furthermore, LASSO penalized cox regression model was used to derive a risk signature comprising seven IFN-γ response genes, including CD74, CSF2RB, PTPN6, MT2A, NMI, LATS2, and PFKP, which can serve as an independent prognostic predictor of LUAD. The risk signature was validated in an independent LUAD cohort. The high risk group is enriched with genes regulating cell cycle and DNA replication, as well as a high level of pro-tumor immune cells. In addition, the risk score is negatively correlated with the expression of immune metagenes, but positively correlated with DNA damage repair genes. Our findings reveal that seven-gene risk signature can be a valuable prognostic predictor for LUAD, and they are crucial participants in tumor microenvironment of LUAD.

INTRODUCTION

IFN-γ is a key cytokine produced by CD4 T helper cells, CD8 cytotoxic T cells [1, 2], natural killer (NK), natural killer T cells (NKT) cells [3], and, to a less extent, by B cells [4], and professional antigen-presenting cells (APCs) in the tumor microenvironment. It plays an important role in coordinating tumor related immune response (10). Its expression is induced by cytokines, including IL-2, IL-12 [5], IL-15 [6], IL-18 [7], and type I interferon [8].

Under physiological conditions, endogenous IFN-γ is vital to many biological processes, including regulation of immune cell functions [9, 10], maintenance of the hematopoietic stem cell niche [10], formation of bone [11], anti-viral host defense. IFN-γ upregulates the expression of MHC class I and genes required for antigen processing to enhance tumor immunogenicity [12]. Therefore, IFN-γ enhances tumor recognition by tumor-specific cytotoxic T lymphocytes, which promotes tumor rejection. Loss of IFN-γ sensitivity is associated with an increased tumor incidence in animals treated with the chemical carcinogen. In humans, functional loss of the tumor suppressor IRF-1, a critical mediator of IFN-γ signaling, is associated with leukemia or gastric cancer [13, 14].

IFN-γ can also have inhibitory immune-regulatory effects on autoimmune [15], or antitumor responses [16]. IFN-γ from lymphocytes enhances the expression of PD-L1 and PD-L2 in tumor and stromal cells, which bind PD-1 on tumor-infiltrating T cells to suppress the cytotoxic response and promotes cancer progression [17, 18]. IFN-γ can also drive the up-regulation of other negative regulatory immune checkpoint molecules such as IDO1 within the tumor microenvironment [19]. Tumor is adapted to take advantage of this positive or negative immune signaling feedback loop to develop and progress. In addition, interferon-stimulated genes are involved in cross-resistance to radiotherapy in tamoxifen-resistant breast cancer [20].

IFN-γ pleiotropic functions are mediated by cell-specific expression of about 200 IFN-γ-regulated genes that include inflammatory signaling molecules, programmed cell death or cell cycle regulators, proteins involved in antigen presentation, and transcriptional factors, such as, major histocompatibility complex (MHC) class I and class II molecules, IRFs, Fc-gamma receptor (FCGR), GBPs (guanylate-binding proteins) and antiviral proteins like PKR, OAS proteins [21, 22]. Characterizing cellular targets of IFN-γ is critical for its prognostic or therapeutic application, particularly in cancers where this cytokine can induce both anti- or pro-tumorigenic effects.

Lung cancer is the leading cause of cancer-related death around the world. Lung adenocarcinoma (LUAD), one of the most frequently observed lung cancer subtypes, has distinct cellular and mutational landscapes with complex immune contexture [23]. Emerging evidence supports that tumor microenvironment impacts LUAD progression and clinical outcome [24]. Nonetheless, no existing study has comprehensively analyzed and screened interferon gamma response genes as the risk signature for LUAD prognosis, or its correlation with various clinical and pathological features. In this study,

we systematically profiled the expression and mutation profile, protein-protein interaction of key interferon gamma response genes in LUAD using RNA-seq data from The Cancer Genome Atlas (TCGA) database. Using consensus clustering based on interferon gamma response genes expression profile, we identified and characterized 2 clinically and molecularly distinct LUAD subtypes; furthermore, using Lasso penalized cox regression analysis, a risk signature comprising seven interferon gamma response genes was constructed. The accuracy and sensitivity of the risk signature in prognosis was successfully validated by 2 independent LUAD cohorts from the GEO database. In addition, the high risk group contains immune repressed features, and is enriched with biological pathways including cell cycle, DNA replication, and DNA damage repair. It also has different immune checkpoint gene expression profiles, compared to the low risk group. Overall, our findings could be valuable in predicting clinical outcome and guiding immunotherapy of lung cancer.

RESULTS

Transcriptional and genetic alterations of 24 interferon gamma response genes in LUAD

The flow chart of this study was presented (Supplementary Figure 1). We chose the gene set, 'HALLMARK_INTERFERON_GAMMA_RESPONSE', for our analysis, which include 200 genes up-regulated in response to interferon gamma, from the molecular signature database of GSEA (https://www.gsea-msigdb.org/). We used TCGA LUAD data as our training set for the risk signature construction. The clinico-pathological information of TCGA LUAD patients is summarized (Supplementary Table 1A). Based on initial univariate cox analysis, 24 interferon gamma response genes, which had statistically significant hazard ratio (HR) related to patient overall survival, were used in the subsequent study.

To characterize these genes, we compared the expression profile of 24 interferon gamma response genes between tumor and normal samples, and presented the results in the heatmap (Figure 1A) and box plots (Figure 1B), which showed that, the expression of 15 interferon gamma response genes was markedly downregulated in tumor samples, including CD69, CD74, CD86, CDKN1A, CIITA, CSF2RB, IL10RA, IRF8, LATS2, LCP2, MT2A, NOD1, PTPN6, SELP, SOD2, while the expression of 9 genes was markedly upregulated in LUAD, including IRF4, ITGB7, NMI, OAS3, PFKP, PNP, RBCK1, RIPK2, and TRAFD1. The differential gene expression analysis results were summarized (Supplementary Table 2).

Among TCGA LUAD samples, 101 had mutations in these genes, with frequency of 18%. It was found that SELP exhibited the highest mutation frequency followed by CD86, CSF2RB, LATS2, LCP2, IRF4, and IRF8, etc, while CD74 and TRAFD1did not show any mutations in these samples (Figure 1C). Further analyses revealed a significant mutation co-occurrence relationship between SELP and PTPN6, SELP and IRF8, OAS3 and CSF2RB1, LCP2 and CSF2RB1, along with LCP2 and RBCK1 (Figure 1D).

Overall, the above analyses revealed a highly heterogeneous landscape of genetic and transcriptional alteration in interferon gamma response genes between tumor and normal samples, implying that these genes may play a crucial role in the tumor occurrence and progression.

Gene expression correlation and protein-protein interaction among 24 interferon gamma response genes

For a better understanding of interactions among these 24 interferon gamma response genes, the correlation in gene expression was calculated (Figure 2A). Clearly, ITGB7 and IRF4 were positively correlated with each other, CD69, IRF8, CD86, LCP2, CSF2RB, IL10RA, PTPN6, CD74, CIITA were positively correlated with each other. SELP exhibited a significantly positive correlation with LATS2. TRAFD1, NMI, and OAS3 positively correlated with each other. In addition, NOD1 were negatively correlated with PFKP, PNP, and NMI, while PFKP were negatively correlated with CD74, NOD1, and SELP. Of note, the correlation between IL10RA and IRF8 (0.8), CD86 (0.7), LCP2
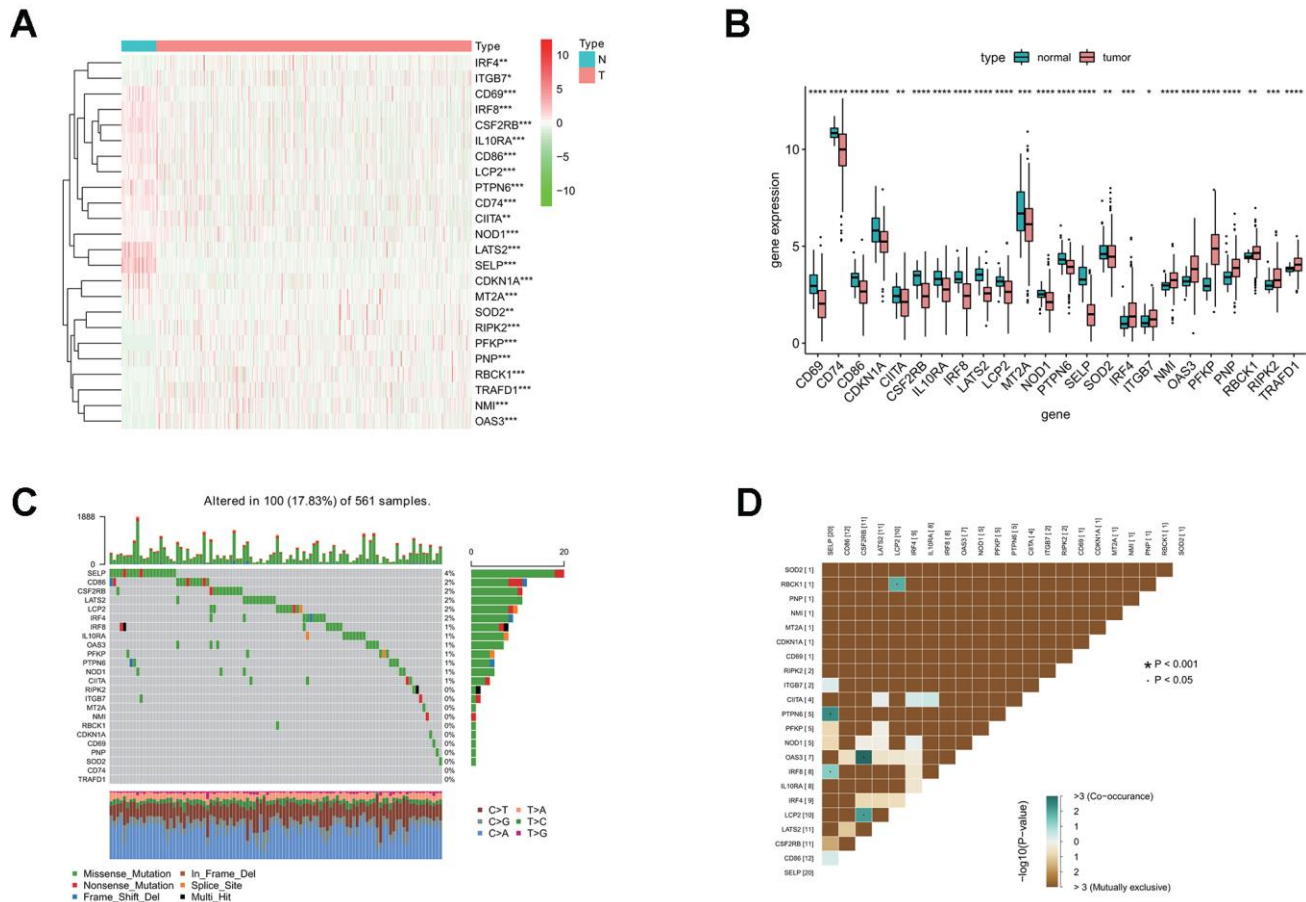

Figure 1. Differential expression of interferon gamma response genes between tumor and normal tissues in the TCGA LUAD cohort. (**A**, **B**) Differential expression of interferon gamma response genes between tumor and normal tissues in the TCGA LUAD cohort is presented in the heatmap (**A**) and the box plot (**B**). ***P<0.001 (normal vs. tumor tissues). N, normal; T, tumor. (**C**) The mutation landscape of 24 interferon gamma response genes in 561 patients from the TCGA LUAD cohort. The middle panel depicts the gene mutation patterns across each sample with different mutation type colored differently. The mutation frequency in each gene is listed on the right of the middle panel, total mutation burden for each sample is shown in the upper barplot, the proportion of each mutation type of genes is shown on the right barplot. The fraction of nucleotide conversions in each sample is indicated by the stacked barplot below. (**D**) The correlation coefficient analysis of mutation co-occurrence in 24 interferon gamma response genes from the TCGA LUAD cohort.
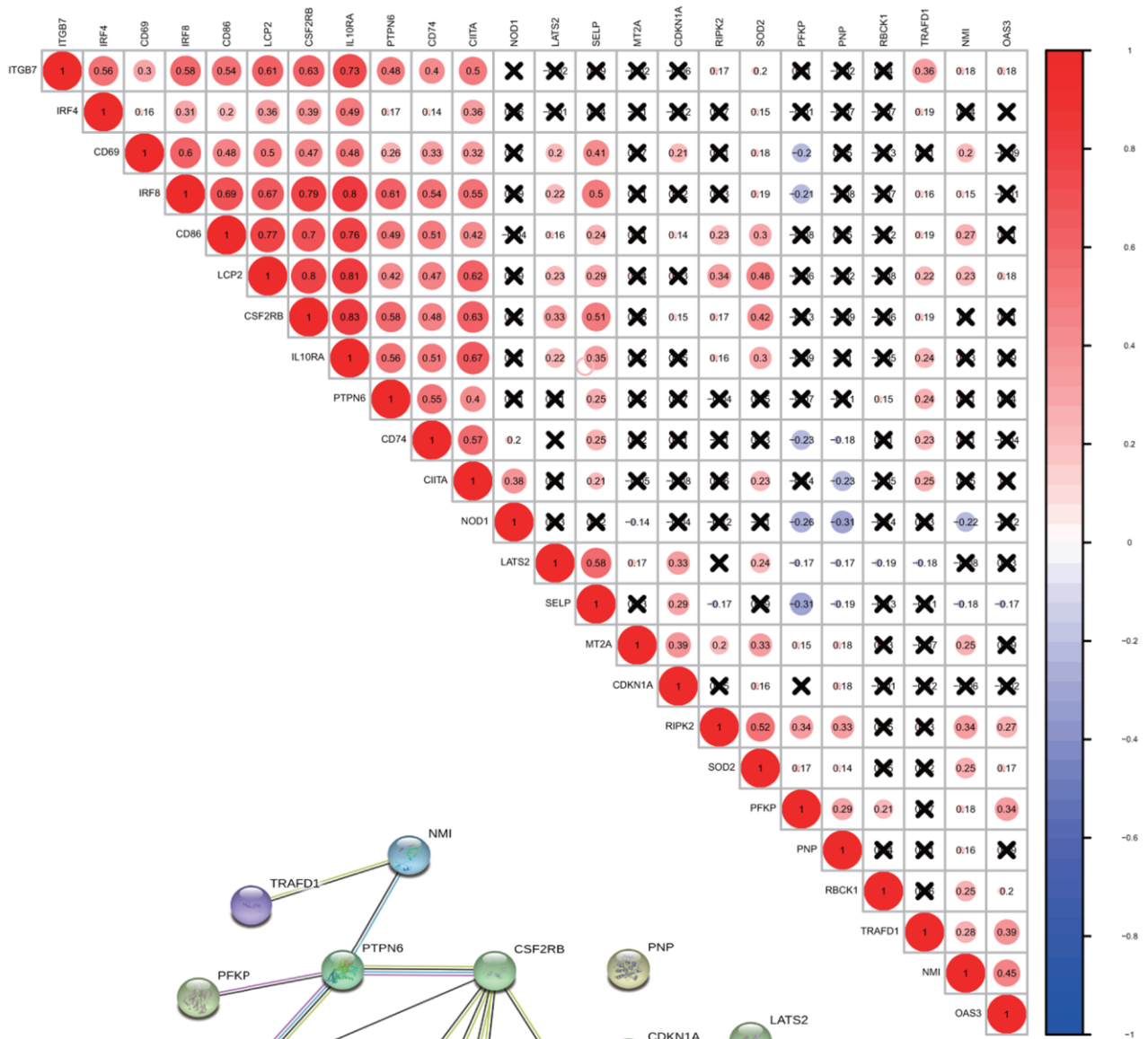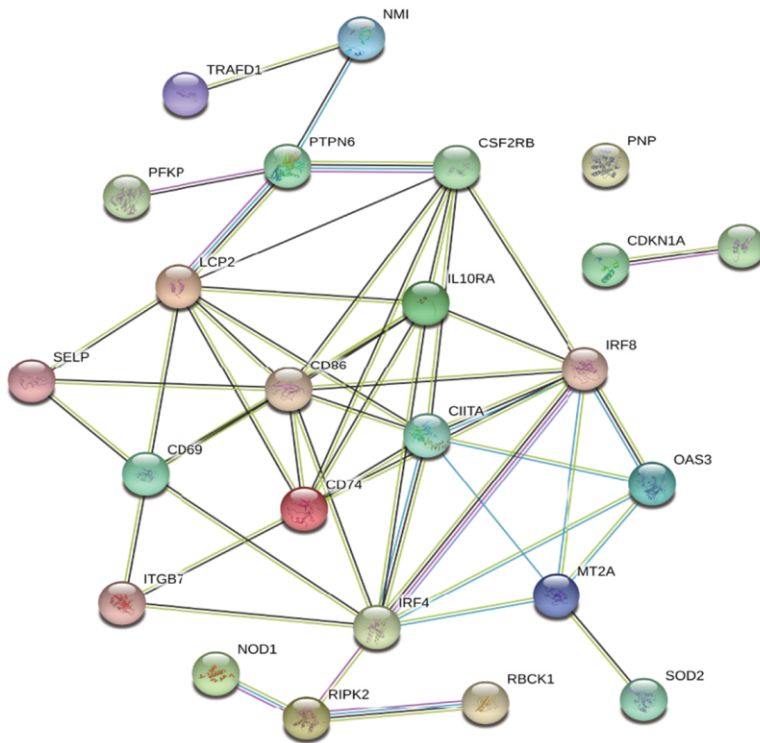
Figure 2. Correlation and interaction among interferon gamma response genes. (**A**) Spearman correlation analysis of the expression among 24 interferon gamma response genes from the TCGA LUAD cohort. (**B**) Protein-protein interactions among 24 interferon gamma response genes.

(0.81) and CSF2RB (0.83), ranked top among all correlations.

Protein-protein interaction among them was also analyzed (Figure 2B). The interactions between these proteins were supported by experimental assays (pink lines), the curated databases (blue lines), co-expression (black lines), or text mining (olive green lines). In addition, LCP2, CSF2RB, CD69, CD86, CD74, IL10RA, IRF4, IRF8, CIITA, MT2A are connected with at least four other proteins, suggesting that these proteins may regulate each other. However, CDKN1A and LATS2 connects only with each other, while PNP appears to have no connection. This may suggest that they are regulated by proteins other than those under study. Overall, the protein-protein interaction analysis among 24 interferon gamma response genes indicates that interferon gamma response pathway is tightly regulated process.

Consensus clustering analysis based on expression of interferon gamma response genes

Consensus clustering analysis indicates that LUAD patients in the TCGA cohort can be classified into two clusters (Figure 3A, 3B), with the range of empirical cumulative distribution (CDF) presented for k = 2 to k = 9 (Supplementary Figure 2A–2J). Therefore, TCGA LUAD cohort was grouped into two clusters based on two fully distinct expression patterns of 24 interferon gamma response genes.
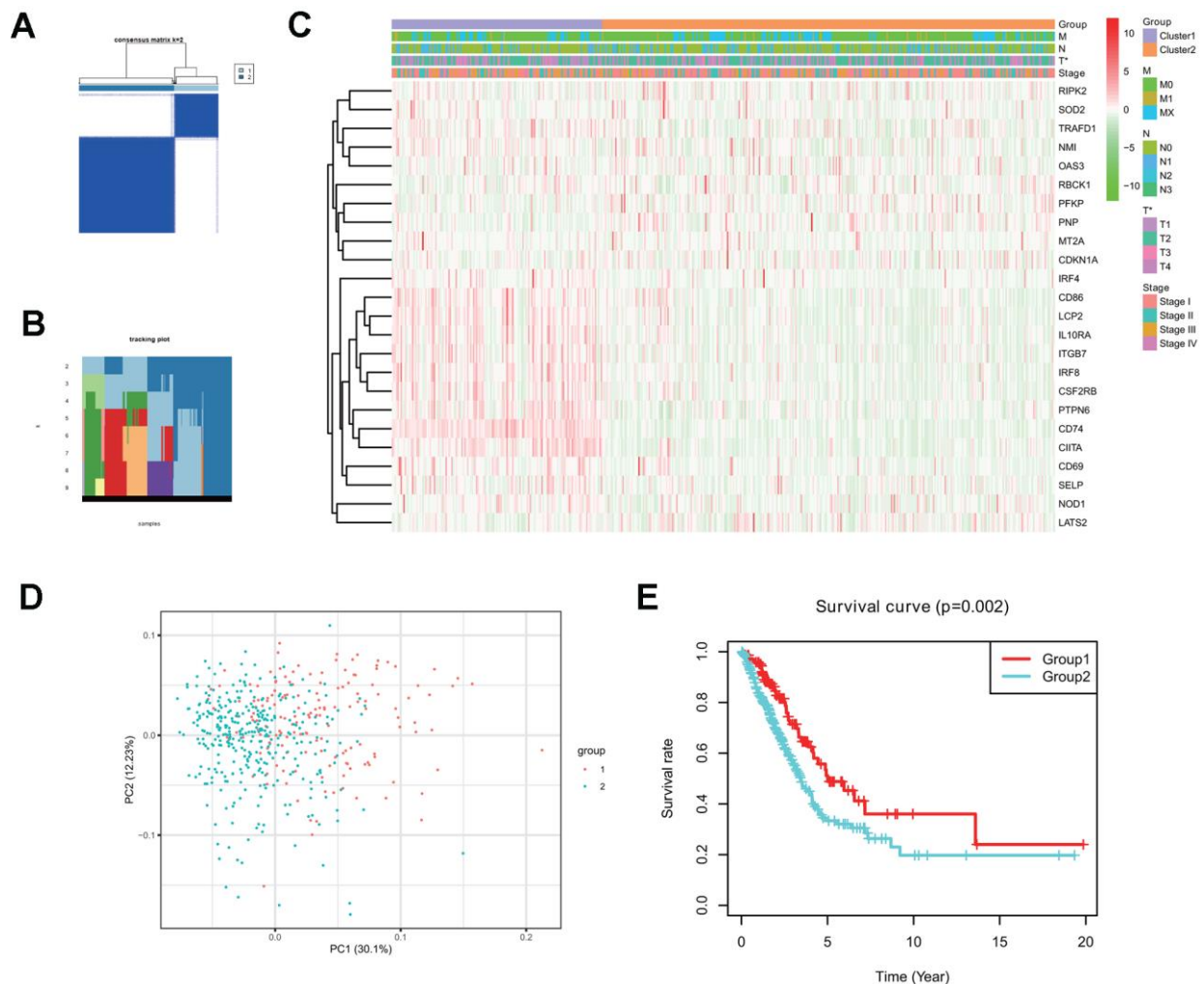

Figure 3. Two clusters in the TCGA LUAD cohort with distinct clinical outcomes and clinicopathological features identified by consensus clustering. (**A**) Consensus clustering matrix at k=2. (**B**) Tracking plot at k=2 to k=10 by consensus clustering. (**C**) Heatmap of interferon gamma response gene expression for two clusters (1 and 2) based on consensus clustering of TCGA LUAD tumor samples, together with the clinical and pathological features (T, N, M, or stage). (**D**) Principal component analysis (PC1 vs. PC2) of 24 interferon gamma response gene expression pattern in the TCGA LUAD cohort. Two clusters are marked with colors (green and red). (**E**) Kaplan-Meier overall survival rate curves for two clusters in the TCGA LUAD cohorts. P=0.002 (group 1 vs. group 2).

The clinical features between two clusters (clusters 1 and 2) were presented in the heatmap (Figure 3C). Cluster 2 was characterized by the decreased expression of IRF4, CD86, LCP2, IL10RA, ITGB7, IRF8, CSF2RB, PTPN6, CD74, CIITA, CD69, SELP, and NOD1, and the increased expression of other genes, including RIPK2, SOD2, TRAFD1, NM1, OAS3, RBCK1, PFKP, PNP, MT2A, and CDKN1A. Furthermore, cluster 2 was significantly associated with late T stage (P < 0.05).

Next, principle component analysis (PCA) was carried for comparing gene expression patterns between two clusters. Our PCA analysis revealed that these two clusters were markedly different (Figure 3D). finally, cluster 2 had a markedly reduced overall survival rate, as compared with that in cluster1 (P = 0.002) (Figure 3E).

The above analysis indicates that distinct expression patterns of interferon gamma response genes may serve as prognostic markers in LUAD patient survival.

Identification of a risk signature comprising of 7 interferon gamma response genes in LUAD

Using the univariate cox regression model, we analyzed the prognostic value of 24 interferon gamma response genes in the TCGA cohort (P<0.05; Figure 4A). Among them, 14 genes, including CD74, CIITA, IRF8, PTPN6, CSF2RB, NOD1, ITGB7, LCP2, CD69, IL10RA, SELP, IRF4, CD86, and TRAFD1, were the protective genes with a hazard ratio (HR) of less than 1, while 10 genes, including MT2A, PFKP, RIPK2, NMI, OAS3, RBCK1, LATS2, CDKN1A, PNP, and SOD2, were the risk genes with a HR of more than 1.

For a more precise prediction of LUAD prognosis using interferon gamma response genes, the cox regression algorithm penalized by LASSO was utilized (Figure 4B, 4C). After cross validation, seven genes, including MT2A, PFKP, NMI, LATS2, CD74, PTPN6, CSF2RB, were chosen for calculation of a risk signature. Risk score of each patient was calculated with the gene expression value of seven genes and corresponding LASSO regression coefficient. The detailed formulation is: Risk=0.00164* MT2A expression + 0.00279*PFKP expression + 0.02065*NMI expression + 0.02099*LATS2 expression -0.00020*CD74 expression -0.00566*PTPN6 expression - 0.01433*CSF2RB expression. In addition, Kaplan Meier survival curve analysis using the log-rank method was carried out for these seven genes. The results showed that high expression of CD74 (p = 0.007), CSF2RB (p = 0.0096), or PTPN6 (p = 0.0083), showed the favorite prognosis, while high expression of MT2A (p = 0.0028), NMI (p <

0.0001), LATS2 (p = 0.0015), or PFKP (p = 0.0011) was associated with poor prognosis in LUAD (Figure 5).

We have also used the two-stage log-rank test to perform survival analysis with intersected curves for MT2A or CSF2RB gene. For MT2A gene, LRPV (p-value of the log-rank test) is 0.0028, MTPV (p-value of the suggested stage-II test) is 0.8000, and TSPV (p-value of the two-stage test) is 0.0028. In addition, for CSF2RB gene, LRPV is 0.0096, MTPV is 0.0000, and TSPV is 0.0096. Therefore, the conclusion is consistent with the previous one.

LUAD patients in the TCGA cohort were classified into the low or high risk group, based on the median risk score of 0.1116, the overall survival probability in the high risk group is much lower than that in the low risk group (P=2.508e-8; Figure 6A). The risk score distribution and associated survival status also indicates higher risk score had more chances of death (Figure 6B, 6C). The expression levels of seven screened interferon gamma response genes in these two risk groups are presented in the heatmap (Figure 6D). LATS2, PEKP, MT2A, and NMI are upregulated in the high risk group, while CD74, PTPN6, CSF2RB are down-regulated in the high risk group. In addition, the correlation between the risk score and each clinical feature was analyzed, the late T stage (P<0.001), or stage (P<0.01) showed significant association with the high risk group (Figure 6D).

Meanwhile, HR of the risk score was 4.549 based on univariate Cox proportional hazards regression analysis (95% confidence interval (CI): 3.133–6.605; P<0.001; Figure 6E). Furthermore, multivariate Cox proportional hazards regression analysis revealed that HR of the risk score was 3.613 (95% CI: 2.358–5.536; P<0.001; Figure 6F). Finally, the ROC curve analysis showed that prognosis prediction had an area under the ROC curve (AUC) value of 0.770 when based on the risk score alone, or the AUC value of 0.798 when based on the combination of the risk score and the stage information (1 years; Figure 6G).

These findings indicated that the risk scores calculated based on the risk signature could accurately predict the prognosis and clinical features of the TCGA LUAD cohort.

**Validation of the risk signature using an independent cohort from the GEO database**

To confirm the risk signature independently, we applied seven-gene prognostic signature to an independent lung cancer cohorts in the GEO database (GSE72094). The clinicopathological information of GSE72094 patients was summarized here (Supplementary Table 1B).

The high risk group had a markedly reduced survival probability than the low risk group (GSE72094: p=7.341e-3; Figure 7A). The increase in the risk scores of patients is correlated with more death in patients (Figure 7B, 7C), implying the higher risk score carries more chance of death for patients.

The expression levels of seven genes were also compared between the high and low risk group, as shown in the heatmaps (Figure 7D). PFKP, MT2A, LATS2, and NMI expressions were increased, while CD74, PTPN6, CSF2RB expressions are decreased in the high risk group in the GSE72094 cohort.

A

|	pvalue	Hazard ratio
MT2A	<0.001	1.002(1.001–1.002)
PFKP	<0.001	1.007(1.003–1.011)
CD74	0.001	1.000(1.000–1.000)
RIPK2	0.004	1.032(1.010–1.054)
CIITA	0.002	0.915(0.865–0.969)
NMI	0.004	1.042(1.013–1.071)
OAS3	0.010	1.012(1.003–1.022)
IRF8	0.012	0.948(0.909–0.988)
PTPN6	0.006	0.961(0.935–0.989)
CSF2RB	0.008	0.945(0.906–0.985)
RBCK1	0.014	1.013(1.003–1.024)
NOD1	0.007	0.910(0.849–0.975)
ITGB7	0.014	0.843(0.736–0.966)
LATS2	0.018	1.061(1.010–1.115)
CDKN1A	0.017	1.007(1.001–1.013)
LCP2	0.030	0.953(0.913–0.995)
CD69	0.047	0.962(0.926–0.999)
PNP	0.014	1.017(1.003–1.031)
IL10RA	0.027	0.961(0.928–0.996)
SELP	0.026	0.914(0.845–0.989)
SOD2	0.022	1.008(1.001–1.014)
IRF4	0.020	0.933(0.879–0.989)
CD86	0.046	0.962(0.926–0.999)
TRAFD1	0.050	0.974(0.949–1.000)
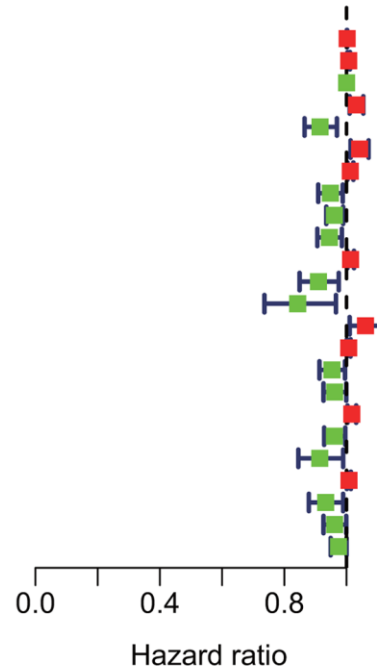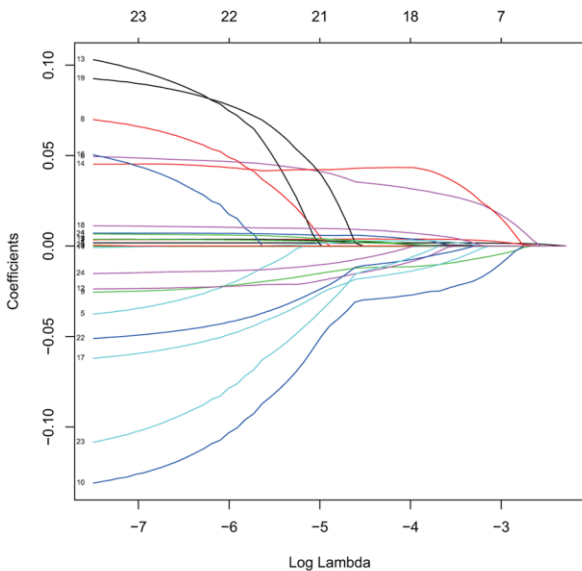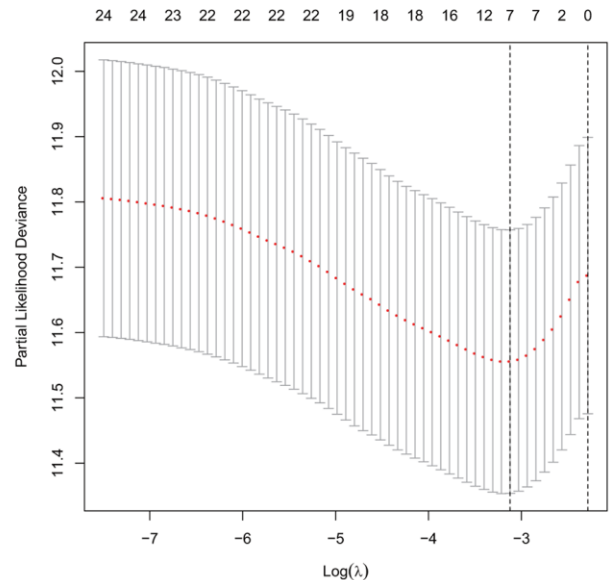
Figure 4. Univariate cox and LASSO Cox regression analysis for overall survival related interferon gamma response genes. (**A**) Forest plots showing the prognostic value detection of various clinical features and the risk score, in which the HRs, corresponding 95% confidence intervals, and p values are displayed. (**B**) LASSO regression analysis was used to calculate the coefficient of interferon gamma response genes. (**C**) Seven genes were selected as active covariates to determine the prognostic value after 10-fold cross-validation for the LASSO model.

Interestingly, TP53 or EGFR mutation is significantly associated with the high risk group.

Univariate and multivariate Cox regression analysis were carried out to evaluate whether the risk score can serve as an independent prognostic factor compared to other clinical features, including gender, age, gene mutations, or stage. In univariate cox regression analysis, for GSE72094, HR of the risk score is 12.359 (95% CI: 3.715 to 41.122) with p <0.001 (Figure 7E). In multivariate cox analysis, for GSE72094, HR of the risk score is 8.297 (95% CI: 3.715 to 41.122) with p <0.001 (Figure 7F); This analysis indicates that the risk signature could serve as independent prognostic factors in this lung cancer data.

The ROC curve analysis was done to assess the prognostic accuracy of the risk signature. For GSE72094, the prognostic prediction has the AUC value of 0.637 based on the risk score, which was higher than other clinical features, including gender (0.580), age (0.513), KRAS mutation (0.552), LKB1 mutation (0.502), but was a little lower than stage (0.637). The AUC value

increased to 0.729 when prognostic predication was assessed by the combination of both the risk score and stage information (3 years; Figure 7G).

Nomogram of TCGA LUAD and GSE72094 cohorts were also presented to quantify the whole risk score for seven genes based on the clinical features (Supplementary Figure 3A, 3B).

Functional analysis of genes associated with the high and low risk group

Genes associated with the high and low risk group of LUAD patients from the TCGA cohort were analyzed by GSEA for enrichment analysis of KEGG pathways, in order to examine the potential biological functions in each group. Genes associated with the high risk group are enriched in pathways including proteasome, cell cycle, DNA replication, spliceosome, DNA mismatch repair, etc, while genes associated with the low risk group are enriched in pathways including thyroid disease, allograft rejection, intestinal immune network for IGA production,
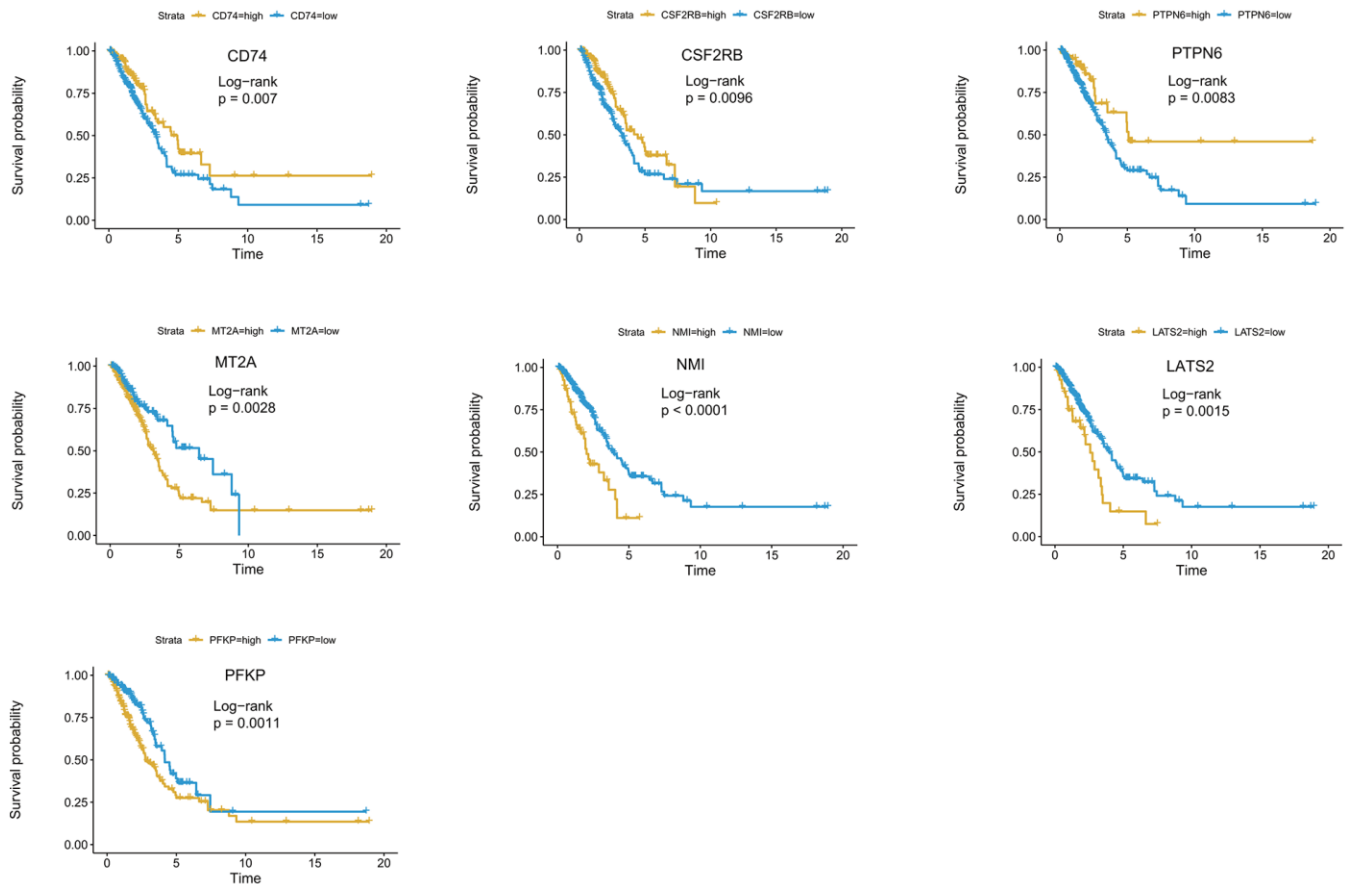

Figure 5. Kaplan-Meier survival curve analysis of seven genes in the risk signature. The survival curves and risk tables of seven genes in the TCGA LUAD cohort were presented with the log-rank test results. Among them, high expression of CD74, CSF2RB, or PTPN6 is associated with favorable prognosis, while high expression of MT2A, NMI, LATS2, or PFKP, is associated with poor prognosis in LUAD.

or graft versus host disease, which are related to immune function (Supplementary Figure 4A).

We carried out similar analysis for the GSE72094 cohort, genes in the high risk group were enriched in pathways including mismatch repair, base excision repair, cell cycle, pyrimidine metabolism, and p53 signaling pathway, etc, and genes in the low risk group were enriched in pathways including GnRH signaling pathway, Fc epsilon RI signaling pathway, valine, leucine and isoleucine degradation, other glycan degradation, etc (Supplementary Figure 4B).
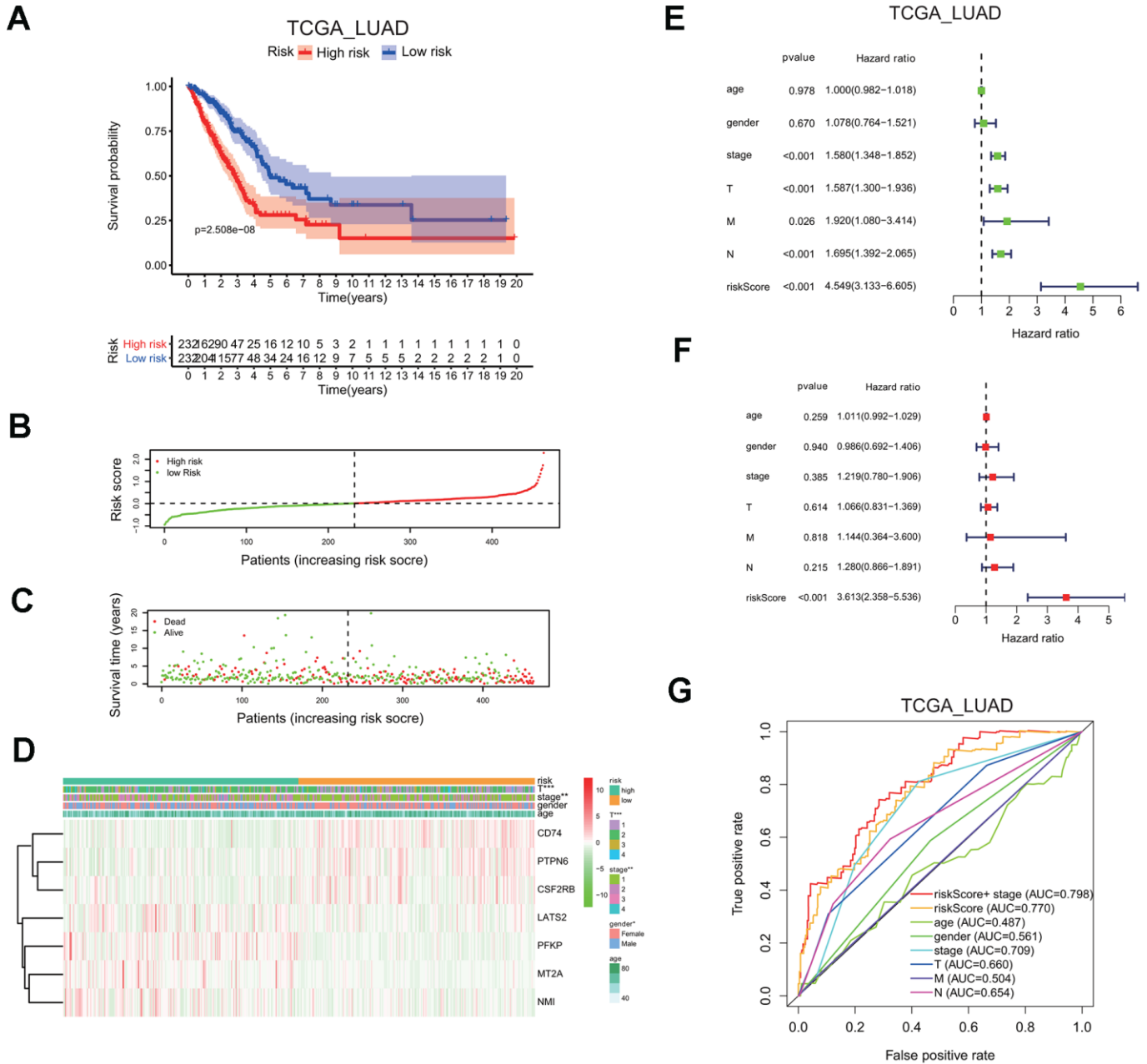

Figure 6. Seven-gene risk signature can predict the prognosis in the TCGA LUAD cohort. (**A**) Kaplan-Meier curves of overall survival probability of the high vs. low risk group. (**B**) Risk score distribution in the TCGA LUAD cohort. (**C**) Survival time and survival status distribution in the TCGA LUAD cohort. (**D**) Heatmap showing the expression level for seven interferon gamma response genes among the high or low risk group in the TCGA LUAD cohort. (**E**, **F**) Univariate (**E**) and multivariate (**F**) Cox regression analysis of the association between clinicopathological features, the risk score, and patient overall survival confirmed the signature as an independent factor of patient overall survival in the TCGA LUAD cohort. (**G**) ROC curves for 3-year survival prediction and clinical characteristics, including age, gender, stage (T, N, or M) in the TCGA LUAD cohort. AUC, area under curve.

Overall, genes in the high risk group were more likely enriched in pathways including cell cycle, DNA replication, mismatch repair pathways, etc, while genes in the low risk group were more likely enriched in immune function related pathways. Thus, this analysis uncovers the underlying difference in cellular pathways in the high and low risk groups.

The risk signature defines distinct immune response, immune cell infiltration, and immune checkpoint gene expression in lung adenocarcinoma

In order to further investigate the underlying connection between risk groups and immune function or cell signaling pathways in lung cancer, GSVA analysis was
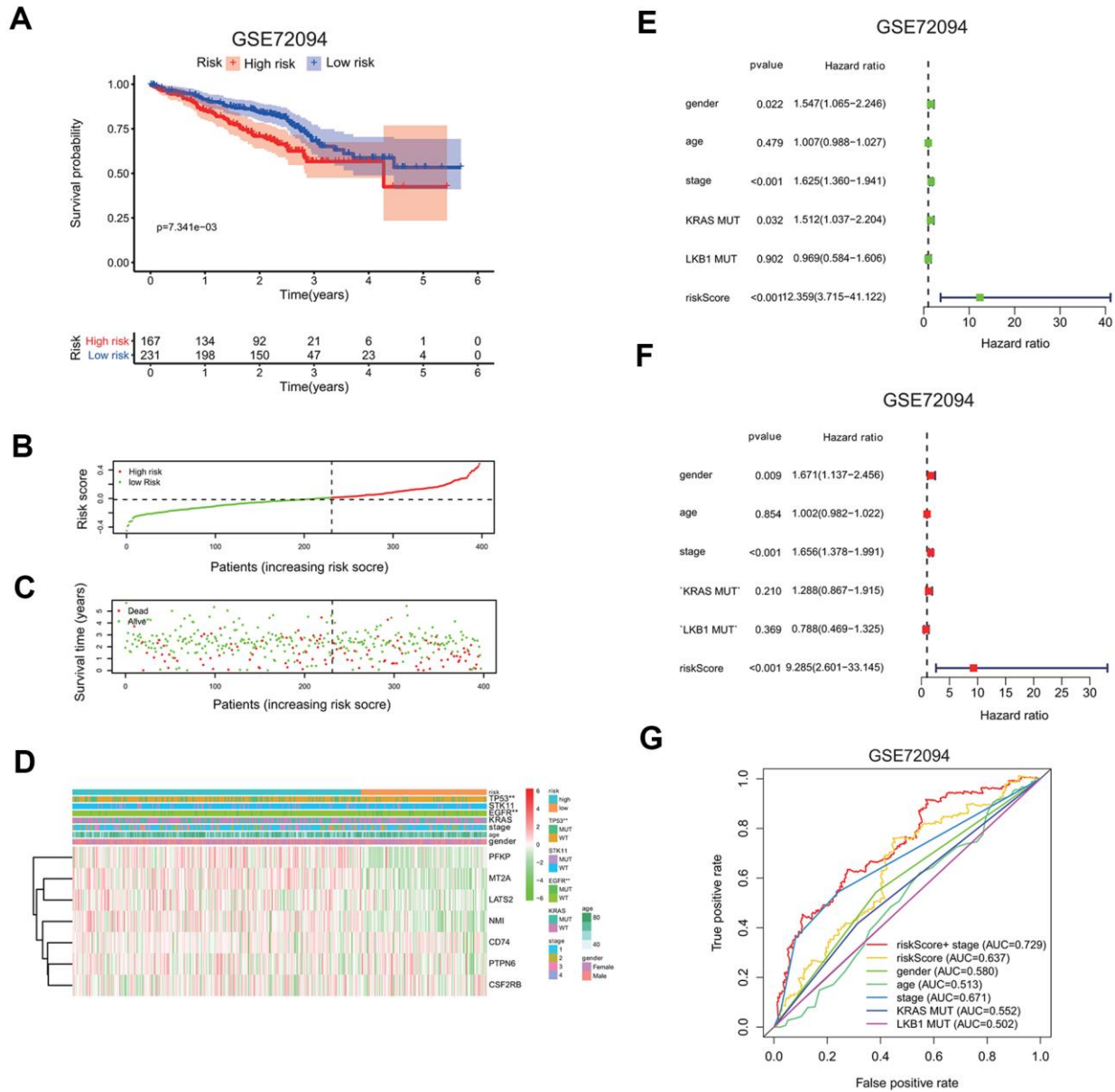

Figure 7. Validation of seven-gene risk signature in an independent LUAD cohort. (**A**) Kaplan-Meier curves of overall survival of the high or low risk group in the GSE72094 cohort. (**B**) Risk score distribution in the GSE72094 cohort. (**C**) Survival time and survival status distribution in the GSE72094 cohort. (**D**) Heatmap showing the expression level of seven interferon gamma response genes in the low or high risk group in the GSE72094 cohort. Association of various clinical features with two groups (high vs. low risk) were determined. *P<0.05, **P<0.01 and ***P<0.001. (**E**, **F**) Univariate (**E**) and multivariate (**F**) Cox regression analysis of the association between clinical features (gender, age, stage, KRAS or LKB1 mutation), risk score, and patient overall survival in the GSE72094 cohort. Hazard ratio and associated p value are shown. (**G**) ROC curves for 3-year survival prediction and clinical characteristics, including age, gender, stage, KRAS or LKB1 mutation in the GSE72094 cohort.

used to calculate the enrichment levels of different gene sets for each sample.

First, we examined the association between risk groups and seven immune metagenes, including interferon, STAT1, MHC-I, HCK, LCK, MHC-II, and IgG, representing different aspects of immune response [25]. GSVA results for the TCGA and GSE72094 cohorts are presented in the heatmap (Figure 8A, 8B), revealing significant difference in the enrichment level of these metagenes between the high and low risk groups.

To quantify their associations, the correlogram was used to display correlation between the risk score, and seven immune metagenes, based on the Pearson's correlation coefficient between the risk score and seven metagenes (Figure 8C). The analysis indicated that the risk score was negatively correlated with HCK, LCK, MHC-II, or IgG metagenes in the TCGA LUAD cohort. To validate these findings, GSVA were also carried out in the GSE72094 cohort, and showed that the risk score was negative correlated with HCK, LCK, MHC-II, or IgG metagenes as well (Figure 8D), indicating that the risk score derived from our prognostic gene signature was negatively associated with immune activities including antigen presentation and T cell activation.

We also extended the GSVA analysis of each sample in these two LUAD cohorts to several KEGG biological pathways, and found that the high risk group were enriched with base excision repair, mismatch repair, or nucleotide excision repair pathway in all two LUAD cohorts examined (Supplementary Figure 5A, 5B). The correlogram confirmed that the risk score is positively correlated with three DNA damage repair pathways, including base excision repair, nucleotide excision repair, and mismatch repair, but negatively correlated with PPAR signaling, cell adhesion molecules, and basal cell carcinoma related pathways (Supplementary Figure 5C, 5D).

Next, we asked whether two risk groups have different immune cells infiltration profile by using CIBERSORT analysis. The relative enrichment levels of 22 immune cells were shown in the radar plots (Figure 9A, 9B). Wilcox test was carried out to determine the difference
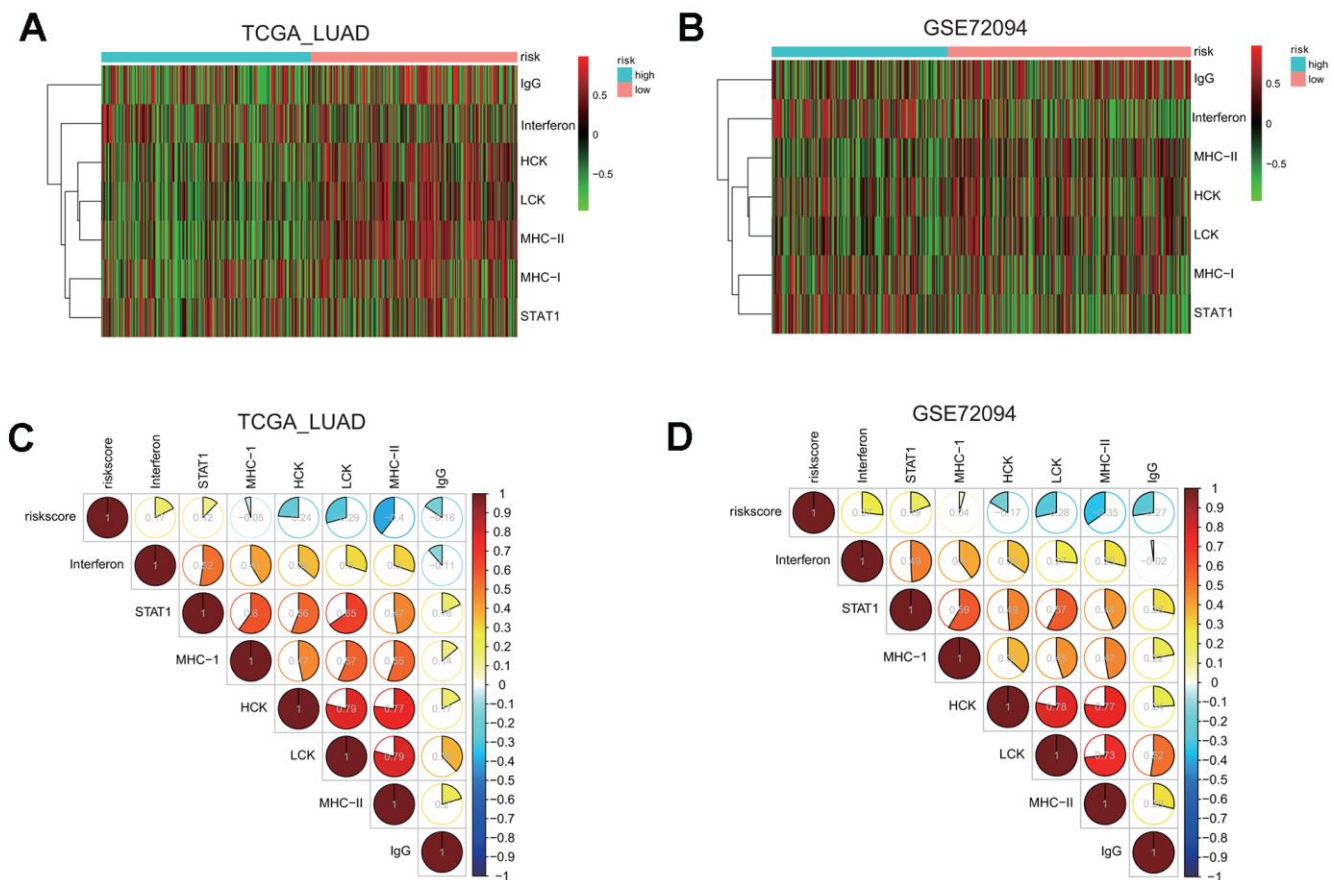

Figure 8. Gene set variation analysis (GSVA) and correlation coefficient analysis between immune-related metagenes and the risk score in two LUAD cohorts. (**A**, **B**) Heatmaps of the enrichment levels of immune-related metagenes, including IgG, Interferon, HCK, LCK, MHC-I, MHC-II, and STAT1, in the high or low risk group in the LUAD cohorts: TCGA (**A**) and GSE72094 (**B**). (**C**, **D**) Correlograms of the risk score and seven immune-related metagenes in the LUAD cohorts: TCGA (**C**) and GSE72094 (**D**).

in the enrichment level of immune cells between the high or low risk groups. In the TCGA LUAD cohort, the high risk group had the markedly higher levels of naïve B cells, activated memory CD4 T cells, macrophage M0 cells, activated dendritic cells, and neutrophils, but the lower level of memory B cells, resting memory CD4 T cells, regulatory T cells, resting

dendritic cells, and resting mast cells (Supplementary Figure 6A).

In the GSE72094 LUAD cohort, the high risk group had the markedly higher level of plasma cells, activated memory CD4 T cells, macrophage M0 and M1 cells, activated dendritic cells, and neutrophils, but the lower
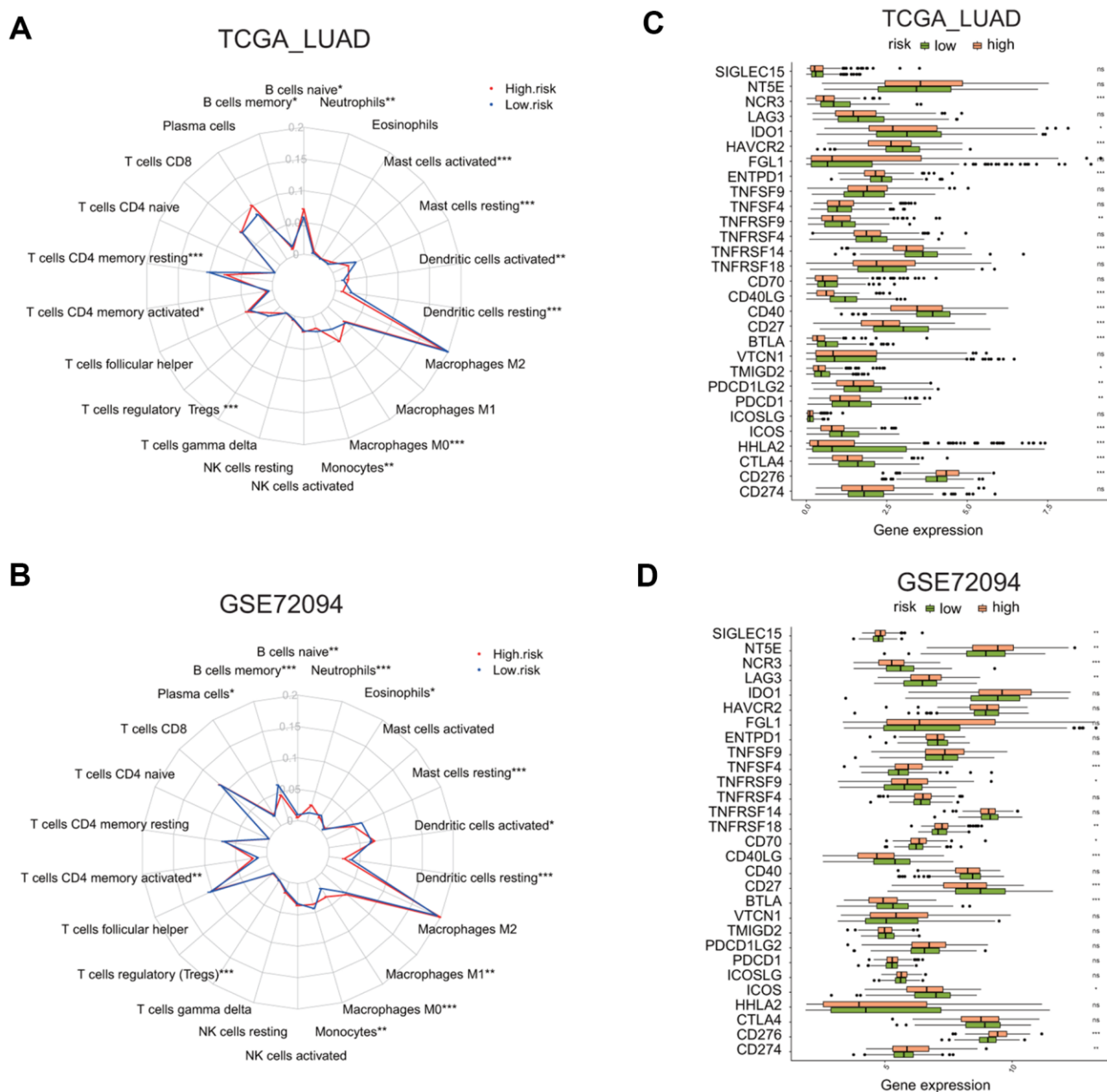

Figure 9. Comparison of immune cell infiltration and immune checkpoint gene expression between the high and low risk groups in two LUAD cohorts. (**A**, **B**) Radar plots show the distribution of the enrichment levels of 22 immune cells in the high (red line) or low (blue line) risk group in the LUAD cohorts: TCGA (**A**) and GSE72094 (**B**). Statistically significant difference in the enrichment level of immune cells between the high and low risk groups is indicated by the asterisks. *P < 0.05, **P < 0.01, ***P < 0.001. (**C**, **D**) Bar plots show the comparison of the expression of 29 immune checkpoint genes between the low and high risk groups in two LUAD cohorts: TCGA (**C**) and GSE72094 (**D**), by *P < 0.05, **P < 0.01, ***P < 0.001.

level of memory B cells, monocytes, resting dendritic cells, resting mast cells, eosinophils (Supplementary Figure 6B).

We also analyzed the differential expression of 29 immune checkpoint molecules or related TNF superfamily members, including CD274, CD276, CTLA4, HHLA2, ICOS, ICOSLG, PDCD1, PDCD1LG2, TMIGD2, VTCN1, BTLA, CD27, CD40, CD40LG, CD70, TNFRSF18, TNFRSF14, TNFRSF4, TNFRSF9, TNFSF4, TNFSF9, ENTPD1, FGL1, HAVCR2, IDO1, LAG3, NCR3, NT5E, SIGLEC15 [26–29] (Figure 9C, 9D). The analysis revealed that, in the TCGA LUAD cohort, the high risk group has a higher expression of CD276, but the reduced expression of CTLA4, HHLA2, ICOS, PDCD1, PDCD1LG2, TMIGD2, BTLA, CD27, CD40, CD40LG, TNFRSF14, TNFRSF9, ENTPD1, HAVCR2, IDO1, or NCR3 (Supplementary Figure 7A, p < 0.05). In the GSE72094 LUAD cohort, the high risk group has a higher expression of CD274, CD276, CD70, TNFRSF18, TNFRSF9, TNFSF4, LAG3, NT5E, or SIGLEC15, but reduced expression of ICOS, BTLA, CD27, CD40LG, or NCR3 (Supplementary Figure 7B, p < 0.05). This analysis indicates that the high or low risk group has distinctive immune checkpoint gene expression profiles.

Consistent with the above analysis, we also found that, in the TCGA LUAD cohort, the high risk group has higher tumor purity, higher tumor mutation burden (TMB), but lower stromal score, lower immune score (Supplementary Figure 8A–8C, P < 0.05).

DISCUSSION

The present study examined the expression, mutation pattern, and protein-protein interaction of 24 interferon gamma response genes in lung adenocarcinoma. To screen for interferon gamma response genes with prognostic value, first, consensus clustering of lung cancer patients was carried out based on 24 interferon gamma response gene expression. two LUAD subgroups, namely cluster 1 and 2, were identified. Specifically, the cluster 2 subgroup has significantly lower survival rate and exhibited a close correlation with late T stage. Second, univariate cox regression analysis was carried out to determine the genes with significant HR values for survival analysis; third, LASSO-penalized cox regression was used to enhance the accuracy of our prognosis model to identifying the most significant variables. A risk signature for prognosis was generated based on seven interferon gamma response genes, which can divide LUAD patients into the low or high risk group. Finally, the risk signature was successfully validated in an independent GEO dataset, and could serve as an independent

predictor for lung cancer prognosis, indicating the general applicability of this risk signature.

The benefits of these approaches are discussed here. Consensus clustering is a powerful unsupervised machine learning algorithm, which can produce a combined clustering unattainable by any single clustering method and are less sensitive to noise, outliers or sample variations. It can be used to mine clinically relevant and previously unknown disease classifications in gene expression data. Next, cox regression analysis is method for investigating the effect of several variables upon the time a specified event happening. It generates a better estimate of these functions than the Kaplan-Meier method when the assumptions of the Cox model are met and the fit of the model is strong. In addition, LASSO-penalized cox regression produces simpler and more interpretable models that incorporate only a reduced set of the predictors.

Interestingly, both MT2A and LATS2 were significantly down-regulated in the lung cancer samples compared to the normal control in the TCGA LUAD cohort (Figure 1B), but the univariate cox analysis of overall survival revealed that both MT2A and LATS2 were the risk genes with a HR of more than 1 (Figure 4), which was validated by our Kaplan Meier survival curve analysis (Figure 5). We validated the accuracy of our analysis by getting the same results using the web-based analysis tool (http://gepia.cancer-pku.cn/index.html). The discrepancy may be due to several reasons, such as: different splicing forms of MT2A or LATS2 proteins may have different roles in tumor development; different lengths of patient follow-up used in the survival analysis may yield different HRs, etc.

Among these seven genes, MT2A acts as anti-oxidant, protects cells against hydroxyl free radicals, is important in homeostatic control of metal in the cell. HMBOX1 interacts with MT2A to regulate intracellular free zinc level, to inhibit apoptosis and promote autophagy in VECs [30]. PFKP encodes a member of the phosphofructokinase. A protein family, plays a key role in glycolysis regulation. PFKP is up-regulated in lung cancer and regulates glucose metabolism [31]. It is a prognostic marker in breast cancer [32]. LATS2 encodes a serine/threonine protein kinase. It can phosphorylate and stabilize SNAI1 in the nucleus, which promotes epithelial-mesenchymal transition and tumor cell invasion. LATS2 is as a poor prognostic marker in non-small cell lung cancer [33]. NMI interacts with CMYC, NMYC, or STATs, and enhances STAT-mediated transcription upon cytokines IL2 and IFN-gamma stimulation. NMI promotes cell proliferation through TGFbeta/Smad pathway by up-

regulating STAT1 in colorectal cancer [34]. CD74 binds class II major histocompatibility complex (MHC) and serves as a chaperone regulating antigen presentation during immune response [35]. It also binds the cytokine macrophage migration inhibitory factor (MIF) to promote cell survival and proliferation [36]. PTPN6 encodes a member of the protein tyrosine phosphatase (PTP) family, which regulates multiple cellular processes, such as cell growth, differentiation, and tumorigenesis [37]. CSF2RB is the common beta chain of the high affinity receptor for IL-3, IL-5 and CSF. Defects in CSF2RB is associated with a lung condition called protein alveolar proteinosis (PAP) [38].

Using the as-constructed risk signature, LUAD patients from two different cohorts can be divided into the high and low risk group, and the high risk group showed significantly poorer survival rate than the low risk group. GSEA analysis revealed that genes in the high risk group are associated with biological processes, such as, cell cycle, DNA replication, base excision repair, and mismatch repair, etc, while genes in the low risk group are enrichment in immune related pathways, such as, allograft rejection, intestinal immune network for IGA production, or graft versus host disease, etc. Furthermore, GSVA analysis indicates the high risk group is negatively correlated with several immune profiling signature (including interferon, LCK, HCK, etc), but highly correlated with DNA repair pathways, including DNA mismatch repair, nuclear excision repair, etc. This result indicates that DNA repair pathway may be involved in negatively regulating immune pathways in lung cancer. Finally, the high risk group has higher tumor purity, and higher tumor burden, but lower stroma score, immune score, and estimate score, which is consistent with our previous analysis.

Immune cell infiltration in tumor microenvironment is an important regulator of tumor progression, our analysis found that there was a significant difference in immune cell infiltration between the high and low risk group of lung cancer patients in three different cohorts. The high risk group has the higher enrichment of immune cells including macrophage M0 cells, activated dendritic cells, activated memory CD4 T cells; while the low risk group has higher enrichment of immune cells including memory B cells, resting memory CD4 T cells, resting dendritic cells, resting mast cells, monocytes, etc. This analysis indicates that the high risk score may represent an immunosuppressive micro-environment.

In addition, several immune checkpoint genes were also differentially expressed in the high and low risk group of lung cancer patients in the cohorts, the high risk

group has the higher expression of CD276 in both TCGA and GSE72094 LUAD cohorts, while the low risk group has the higher expression of CD27, CD40LG, or NCR3 in both cohorts, indicating the possible use of blockers targeting different immune checkpoint genes for different risk groups.

In conclusion, our study comprehensively illustrated the expression and mutation patterns, possible role and prognostic significance of important interferon gamma response genes in lung adenocarcinoma patients, and derived a risk signature composed of seven genes, which could accurately predict the prognosis in lung adenocarcinoma patients. The present study will contribute to enhancing our understanding of interferon gamma response genes, its role in regulating cellular pathways, affecting immune cell-infiltration and immune checkpoint gene expression in tumor micro-environment. This study may help to guide more effective immunotherapy strategies again lung cancer.

MATERIALS AND METHODS

Data acquisition

Gene expression data, somatic mutation data, and clinical data of LUAD patients were downloaded from the TCGA dataset (https://tcga-data.nci.nih.gov/) using the GDC Data Transfer Tool. A total of 551 LUAD patient samples (497 tumor and 54 normal samples) were used in this study to identify the differentially expressed interferon gamma response genes. The GSE72094 dataset was acquired from the GEO database (https://www.ncbi.nlm.nih.gov/geo/). The clinicopathological information of two LUAD cohorts were summarized (Supplementary Table 1). For TCGA LUAD cohort, only 476 samples without missing clinical features were included in the Supplementary Table 1.

Bioinformatic analysis

Difference comparison between two groups (for example, gene expression in tumor vs. normal samples) was carried out using Wilcoxon tests, with p value < 0.05 as statistically significant. Difference comparisons of three or more groups were carried out using One-way ANOVA and Kruskal-Wallis tests. Correlation coefficients between genes were calculated by Spearman correlation coefficient analysis using the 'corrplot' R package. In addition, the interactions between interferon gamma response genes/proteins were obtained from the STRING website (http://www.string-db.org/).

LUAD were clustered in various groups based on the expression pattern of interferon gamma response genes using the 'ConsensusClusterPlus' R package. Principal

component analysis (PCA) analysis of LUAD samples were investigated using the 'FactoMineR' and 'factoextra' R packages. The mutation landscape in the TCGA LUAD cohort was presented by the 'maftools' R package. The 'estimate' R package calculates stromal and immune scores that represent the presence of stromal and immune cells in tumor tissue, respectively. Kaplan Meier survival curve analysis was carried out using the 'survival' R package, the optimal cutpoint value for each gene was determined by the surv_cutpoint function in the 'survminer' R package. Two-stage log-rank test in the 'TSHRC' R package was used to perform the survival analysis with crossed curves.

All statistical P values were two-tailed, and were statistically significant when $p < 0.05$. All data analysis was conducted in R software (version 3 .6.1).

Construction of a signature based on interferon gamma response genes

A univariate Cox regression model was adopted to determine the hazard ratios (HR) of prognosis prediction for interferon gamma response genes. Univariate or multivariate Cox analysis was employed to determine the prognostic value for the risk signature or clinical features using the 'forest' R package.

Genes were further screened by Least Absolute Shrinkage and Selection Operator (LASSO) regression analysis, followed by 10-fold cross validation using the 'glmnet' R package. Seven genes with their regression coefficient (Coef) were selected, the risk score for each patient was calculated through linearly multiplying the expression level with Coef of each gene, according to the following formula: Risk score $= \text{Coef}_{gene1} \times \text{expression}_{gene1} + \text{Coef}_{gene2} \times \text{expression}_{gene2} + \cdots + \text{Coef}_{gene\,n} \times \text{expression}_{gene\,n}$.

Confirmation of the signature based on interferon gamma response genes

All patients in the LUAD cohorts were separated into the high or low risk group based on the median risk score. Survival curves were drawn based on the risk score using the Kaplan-Meier method. Student's t-test and Pearson's test were used to compare the significant association between risk scores and various clinical features.

Subsequently, the receiver operating characteristic (ROC) curves were performed to assess the sensitivity and specificity of survival prediction by the risk signature using the 'survivalROC' R package. Univariate and multivariate Cox proportional hazards regression analysis was performed to determine whether the risk score is an independent predictor for prognosis.

Gene set enrichment analysis (GSEA) and gene set variation analysis (GSVA)

GSEA for the mRNAs associated with the high or low risk group was carried out using c2.cp.kegg.v7.1. symbols.gmt as gene sets database at 1,000 random sample permutations using JAVA application (http://software.broadinstitute.org/gsea/index.jsp).

To investigate the difference on biological pathways between high and low risk groups, GSVA enrichment analysis was carried out using the "GSVA" R package. GSVA, a non-parametric and unsupervised method, is used for calculates sample-wise gene set enrichment scores [39]. The gene sets for pathways are in the "c2.cp.kegg.v7.1.symbols" document from the MSigDB database.

Immune cell infiltration estimation

The CIBERSORT method was used to calculate the enrichment levels of immune cell infiltration in lung cancer cohorts. The difference in the immune cell infiltration between the low and high risk groups was carried out using Wilcoxon tests, with the p value < 0.05 as statistically significant.

The gene set for each immune cell signature, as well as various pro- or anti-tumor immune cell subtypes, came from this study [40]. The enrichment scores determined by CIBERSORT were used to represent the relative abundance of each immune cell in LUAD samples.

Data availability statement

The data that support the findings of this study are openly available from the TCGA dataset (https://tcga-data.nci.nih.gov/). The GSE72094 dataset is available from the GEO database (https://www.ncbi.nlm.nih.gov/geo/).

AUTHOR CONTRIBUTIONS

BY.Y., JX.B., HY.W, QW.L, FZ.Y., WJ.Z., L.Z., and W.L. collected the data and performed the analyses. HY.W., LX.W., XY.H., YT.Z., K.F., and F.Z. validated and interpreted the results. XY.H. and F.Z. conceived the project, XY.H., and F.Z. designed the work flow, and wrote the manuscript with contributions from all authors.

CONFLICTS OF INTEREST

FUNDING

REFERENCES

1. Kasahara T, Hooks JJ, Dougherty SF, Oppenheim JJ. Interleukin 2-mediated immune interferon (IFN-gamma) production by human T cells and T cell subsets. J Immunol. 1983; 130:1784–89. PMID:6403613

2. Matsushita H, Hosoi A, Ueha S, Abe J, Fujieda N, Tomura M, Maekawa R, Matsushima K, Ohara O, Kakimi K. Cytotoxic T lymphocytes block tumor growth both by lytic activity and IFNγ-dependent cell-cycle arrest. Cancer Immunol Res. 2015; 3:26–36. https://doi.org/10.1158/2326-6066.CIR-14-0098 PMID:25127875

3. Yu J, Wei M, Becknell B, Trotta R, Liu S, Boyd Z, Jaung MS, Blaser BW, Sun J, Benson DM Jr, Mao H, Yokohama A, Bhatt D, et al. Pro- and antiinflammatory cytokine signaling: reciprocal antagonism regulates interferon-gamma production by human natural killer cells. Immunity. 2006; 24:575–90. https://doi.org/10.1016/j.immuni.2006.03.016 PMID:16713975

4. Bao Y, Liu X, Han C, Xu S, Xie B, Zhang Q, Gu Y, Hou J, Qian L, Qian C, Han H, Cao X. Identification of IFN-γ-producing innate B cells. Cell Res. 2014; 24:161–76. https://doi.org/10.1038/cr.2013.155 PMID:24296781

5. Ye J, Ortaldo JR, Conlon K, Winkler-Pickett R, Young HA. Cellular and molecular mechanisms of IFN-gamma production induced by IL-2 and IL-12 in a human NK cell line. J Leukoc Biol. 1995; 58:225–33. https://doi.org/10.1002/jlb.58.2.225 PMID:7643015

6. Carson WE, Ross ME, Baiocchi RA, Marien MJ, Boiani N, Grabstein K, Caligiuri MA. Endogenous production of interleukin 15 by activated human monocytes is critical for optimal production of interferon-gamma by natural killer cells in vitro. J Clin Invest. 1995; 96:2578–82. https://doi.org/10.1172/JCI118321 PMID:8675621

7. Takeda K, Tsutsui H, Yoshimoto T, Adachi O, Yoshida N, Kishimoto T, Okamura H, Nakanishi K, Akira S. Defective NK cell activity and Th1 response in IL-18-deficient mice. Immunity. 1998; 8:383–90. https://doi.org/10.1016/s1074-7613(00)80543-9 PMID:9529155

8. Matikainen S, Paananen A, Miettinen M, Kurimoto M, Timonen T, Julkunen I, Sareneva T. IFN-alpha and IL-18 synergistically enhance IFN-gamma production in human NK cells: differential regulation of Stat4 activation and IFN-gamma gene expression by IFN-alpha and IL-12. Eur J Immunol. 2001; 31:2236–45. https://doi.org/10.1002/1521-4141(200107)31:7<2236::AID-IMMU2236>3.0.CO;2-G PMID:11449378

9. Sercan O, Hämmerling GJ, Arnold B, Schüler T. Innate immune cells contribute to the IFN-gamma-dependent regulation of antigen-specific CD8+ T cell homeostasis. J Immunol. 2006; 176:735–39. https://doi.org/10.4049/jimmunol.176.2.735 PMID:16393956

10. Baldridge MT, King KY, Boles NC, Weksberg DC, Goodell MA. Quiescent haematopoietic stem cells are activated by IFN-gamma in response to chronic infection. Nature. 2010; 465:793–97. https://doi.org/10.1038/nature09135 PMID:20535209

11. Duque G, Huang DC, Dion N, Macoritto M, Rivas D, Li W, Yang XF, Li J, Lian J, Marino FT, Barralet J, Lascau V, Deschênes C, et al. Interferon-γ plays a role in bone formation in vivo and rescues osteoporosis in ovariectomized mice. J Bone Miner Res. 2011; 26:1472–83. https://doi.org/10.1002/jbmr.350 PMID:21308779

12. Xiang J, Moyana T. Cytotoxic CD4+ T cells associated with the expression of major histocompatibility complex class II antigen of mouse myeloma cells secreting interferon-gamma are cytolytic in vitro and tumoricidal in vivo. Cancer Gene Ther. 1998; 5:313–20. PMID:9824051

13. Taniguchi T. Transcription factors IRF-1 and IRF-2: linking the immune responses and tumor suppression. J Cell Physiol. 1997; 173:128–30. https://doi.org/10.1002/(SICI)1097-4652(199711)173:2<128::AID-JCP8>3.0.CO;2-P PMID:9365509

14. Nozawa H, Oda E, Ueda S, Tamura G, Maesawa C, Muto T, Taniguchi T, Tanaka N. Functionally inactivating point mutation in the tumor-suppressor IRF-1 gene identified in human gastric cancer. Int J Cancer. 1998; 77:522–27.
https://doi.org/10.1002/(sici)1097-0215(19980812)77:4<522::aid-ijc8>3.0.co;2-w
PMID:9679752

15. Ferber IA, Brocke S, Taylor-Edwards C, Ridgway W, Dinisco C, Steinman L, Dalton D, Fathman CG. Mice with a disrupted IFN-gamma gene are susceptible to the induction of experimental autoimmune encephalomyelitis (EAE). J Immunol. 1996; 156:5–7.
PMID:8598493

16. Lee SJ, Jang BC, Lee SW, Yang YI, Suh SI, Park YM, Oh S, Shin JG, Yao S, Chen L, Choi IH. Interferon regulatory factor-1 is prerequisite to the constitutive expression and IFN-gamma-induced upregulation of B7-H1 (CD274). FEBS Lett. 2006; 580:755–62.
https://doi.org/10.1016/j.febslet.2005.12.093
PMID:16413538

17. Liang SC, Latchman YE, Buhlmann JE, Tomczak MF, Horwitz BH, Freeman GJ, Sharpe AH. Regulation of PD-1, PD-L1, and PD-L2 expression during normal and autoimmune responses. Eur J Immunol. 2003; 33:2706–16.
https://doi.org/10.1002/eji.200324228
PMID:14515254

18. Abiko K, Matsumura N, Hamanishi J, Horikawa N, Murakami R, Yamaguchi K, Yoshioka Y, Baba T, Konishi I, Mandai M. IFN-γ from lymphocytes induces PD-L1 expression and promotes progression of ovarian cancer. Br J Cancer. 2015; 112:1501–09.
https://doi.org/10.1038/bjc.2015.101 PMID:25867264

19. Spranger S, Spaapen RM, Zha Y, Williams J, Meng Y, Ha TT, Gajewski TF. Up-regulation of PD-L1, IDO, and T(regs) in the melanoma tumor microenvironment is driven by CD8(+) T cells. Sci Transl Med. 2013; 5:200ra116.
https://doi.org/10.1126/scitranslmed.3006504
PMID:23986400

20. Post AE, Smid M, Nagelkerke A, Martens JW, Bussink J, Sweep FC, Span PN. Interferon-stimulated genes are involved in cross-resistance to radiotherapy in tamoxifen-resistant breast cancer. Clin Cancer Res. 2018; 24:3397–408.
https://doi.org/10.1158/1078-0432.CCR-17-2551
PMID:29661777

21. de Veer MJ, Holko M, Frevel M, Walker E, Der S, Paranjape JM, Silverman RH, Williams BR. Functional classification of interferon-stimulated genes identified using microarrays. J Leukoc Biol. 2001; 69:912–20.
PMID:11404376

22. Castro F, Cardoso AP, Gonçalves RM, Serre K, Oliveira MJ. Interferon-gamma at the crossroads of tumor immune surveillance or evasion. Front Immunol. 2018; 9:847.
https://doi.org/10.3389/fimmu.2018.00847
PMID:29780381

23. Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. Nature. 2014; 511:543–50.
https://doi.org/10.1038/nature13385 PMID:25079552

24. Zheng S, Luo X, Dong C, Zheng D, Xie J, Zhuge L, Sun Y, Chen H. A B7-CD28 family based signature demonstrates significantly different prognoses and tumor immune landscapes in lung adenocarcinoma. Int J Cancer. 2018; 143:2592–601.
https://doi.org/10.1002/ijc.31764 PMID:30152019

25. Rody A, Holtrich U, Pusztai L, Liedtke C, Gaetje R, Ruckhaeberle E, Solbach C, Hanker L, Ahr A, Metzler D, Engels K, Karn T, Kaufmann M. T-cell metagene predicts a favorable prognosis in estrogen receptor-negative and HER2-positive breast cancers. Breast Cancer Res. 2009; 11:R15.
https://doi.org/10.1186/bcr2234 PMID:19272155

26. Wang J, Sanmamed MF, Datar I, Su TT, Ji L, Sun J, Chen L, Chen Y, Zhu G, Yin W, Zheng L, Zhou T, Badri T, et al. Fibrinogen-like protein 1 is a major immune inhibitory ligand of LAG-3. Cell. 2019; 176:334–47.e12.
https://doi.org/10.1016/j.cell.2018.11.010
PMID:30580966

27. Zhang C, Zhang Z, Zhang G, Zhang Z, Luo Y, Wang F, Wang S, Che Y, Zeng Q, Sun N, He J. Clinical significance and inflammatory landscapes of a novel recurrence-associated immune signature in early-stage lung adenocarcinoma. Cancer Lett. 2020; 479:31–41.
https://doi.org/10.1016/j.canlet.2020.03.016
PMID:32201203

28. Pardoll DM. The blockade of immune checkpoints in cancer immunotherapy. Nat Rev Cancer. 2012; 12:252–64.
https://doi.org/10.1038/nrc3239 PMID:22437870

29. Aggarwal BB. Signalling pathways of the TNF superfamily: a double-edged sword. Nat Rev Immunol. 2003; 3:745–56.
https://doi.org/10.1038/nri1184 PMID:12949498

30. Ma H, Su L, Yue H, Yin X, Zhao J, Zhang S, Kung H, Xu Z, Miao J. HMBOX1 interacts with MT2A to regulate autophagy and apoptosis in vascular endothelial cells. Sci Rep. 2015; 5:15121.
https://doi.org/10.1038/srep15121 PMID:26456220

31. Shen J, Jin Z, Lv H, Jin K, Jonas K, Zhu C, Chen B. PFKP is highly expressed in lung cancer and regulates glucose metabolism. Cell Oncol (Dordr). 2020; 43:617–29.

https://doi.org/10.1007/s13402-020-00508-6
PMID:32219704

32. Yeerken D, Hong R, Wang Y, Gong Y, Liu R, Yang D, Li J, Fan J, Chen J, Zhang W, Zhan Q. PFKP is transcriptionally repressed by BRCA1/ZBRK1 and predicts prognosis in breast cancer. PLoS One. 2020; 15:e0233750.
https://doi.org/10.1371/journal.pone.0233750
PMID:32470015

33. Wu A, Li J, Wu K, Mo Y, Luo Y, Ye H, Mai Z, Guo K, Wang Y, Li S, Chen H, Luo W, Yang Z. LATS2 as a poor prognostic marker regulates non-small cell lung cancer invasion by modulating MMPs expression. Biomed Pharmacother. 2016; 82:290–97.
https://doi.org/10.1016/j.biopha.2016.04.008
PMID:27470365

34. Ji D, Feng Y, Peng W, Li J, Gu Q, Zhang Z, Qian W, Wang Q, Zhang Y, Sun Y. NMI promotes cell proliferation through TGFβ/Smad pathway by upregulating STAT1 in colorectal cancer. J Cell Physiol. 2020; 235:429–41.
https://doi.org/10.1002/jcp.28983 PMID:31230364

35. Jasanoff A, Wagner G, Wiley DC. Structure of a trimeric domain of the MHC class II-associated chaperonin and targeting protein ii. EMBO J. 1998; 17:6812–18.
https://doi.org/10.1093/emboj/17.23.6812
PMID:9843486

36. Leng L, Metz CN, Fang Y, Xu J, Donnelly S, Baugh J, Delohery T, Chen Y, Mitchell RA, Bucala R. MIF signal transduction initiated by binding to CD74. J Exp Med. 2003; 197:1467–76.

https://doi.org/10.1084/jem.20030286
PMID:12782713

37. St-Denis N, Gupta GD, Lin ZY, Gonzalez-Badillo B, Veri AO, Knight JDR, Rajendran D, Couzens AL, Currie KW, Tkach JM, Cheung SWT, Pelletier L, Gingras AC. Phenotypic and Interaction Profiling of the Human Phosphatases Identifies Diverse Mitotic Regulators. Cell Rep. 2016; 17:2488–501.
https://doi.org/10.1016/j.celrep.2016.10.078
PMID:27880917

38. Suzuki T, Maranda B, Sakagami T, Catellier P, Couture CY, Carey BC, Chalk C, Trapnell BC. Hereditary pulmonary alveolar proteinosis caused by recessive CSF2RB mutations. Eur Respir J. 2011; 37:201–04.
https://doi.org/10.1183/09031936.00090610
PMID:21205713

39. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinformatics. 2013; 14:7.
https://doi.org/10.1186/1471-2105-14-7
PMID:23323831

40. Charoentong P, Finotello F, Angelova M, Mayer C, Efremova M, Rieder D, Hackl H, Trajanoski Z. Pan-cancer Immunogenomic Analyses Reveal Genotype-Immunophenotype Relationships and Predictors of Response to Checkpoint Blockade. Cell Rep. 2017; 18:248–62.
https://doi.org/10.1016/j.celrep.2016.12.019
PMID:28052254
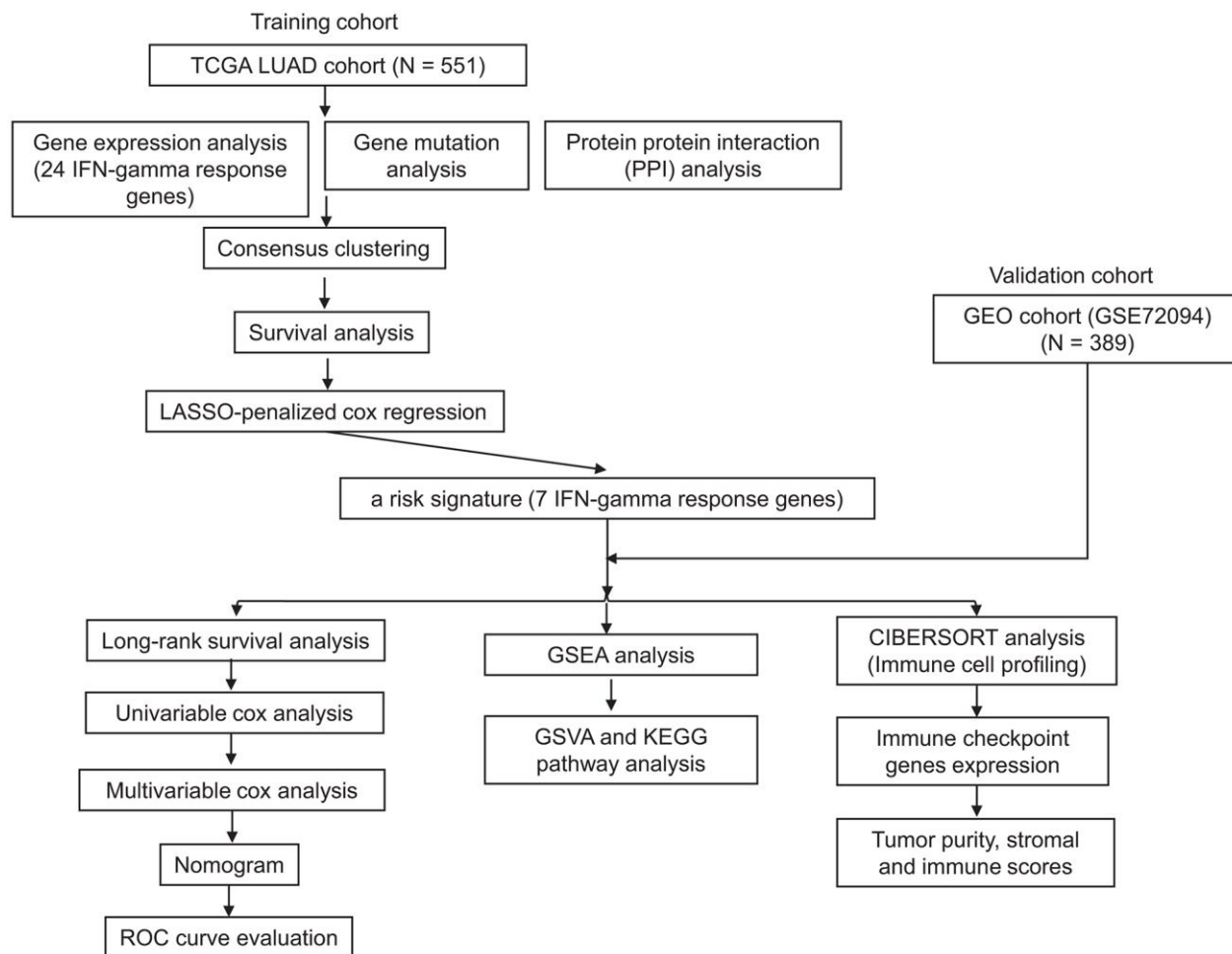
Supplementary Figure 1. Flow chart of the study.
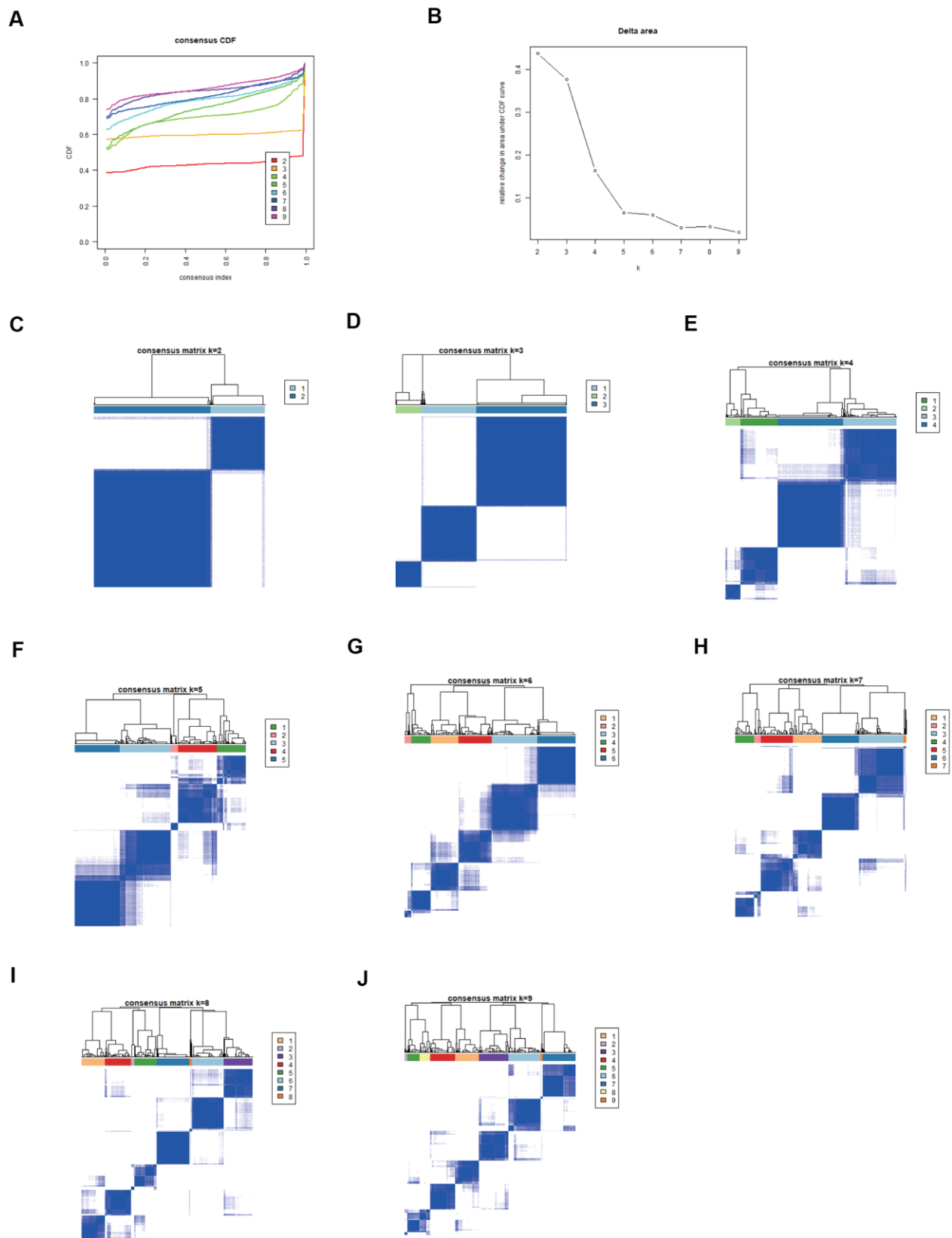
Supplementary Figure 2. Consensus clustering of 24 interferon gamma response genes identified two clusters of LUAD in the TCGA LUAD cohort. (**A**) Cumulative distribution function (CDF) plot displays consensus distributions at k=2 to k=9. (**B**) Relative changes in the area under the CDF curve at k=2 to k=9. (**C–J**) Consensus matrix display at k=2 (**C**) to k=9 (**J**).
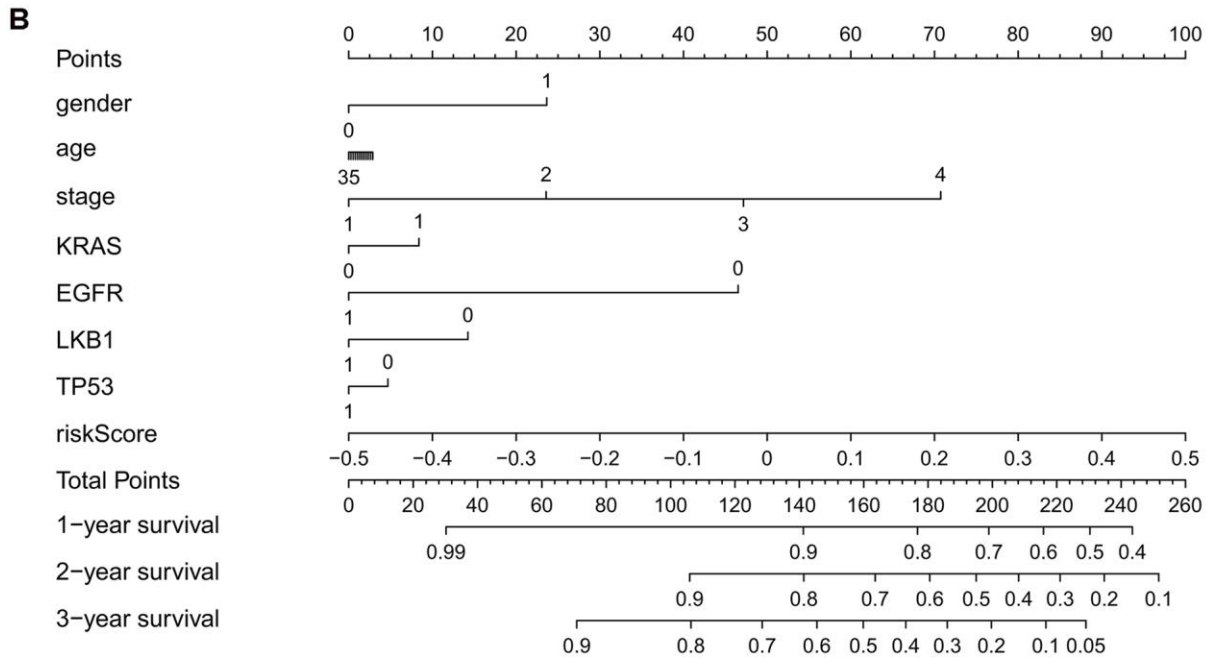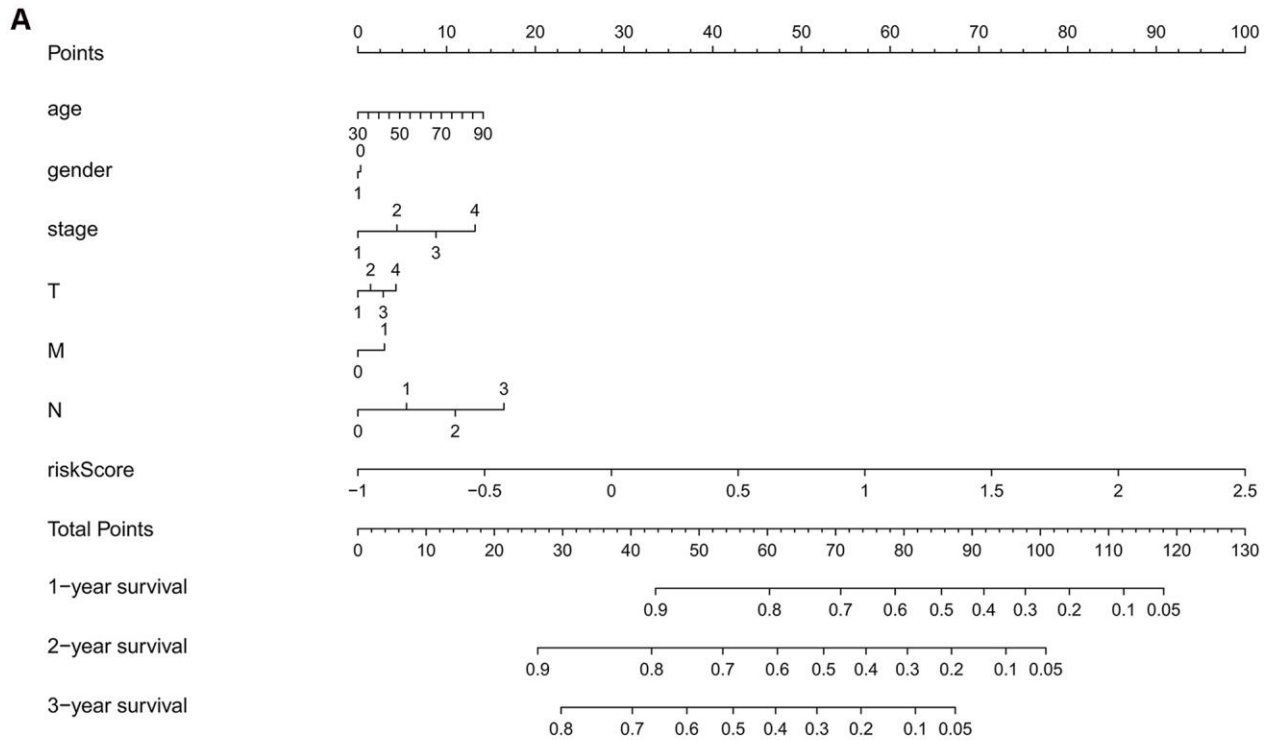
Supplementary Figure 3. Nomogram of TCGA LUAD and GSE72094 cohorts to quantify the whole risk score for seven genes based on the clinical features. (**A**) LUAD cohort. (**B**) GSE72094 cohort.

A TCGA_LUAD

- KEGG_PROTEASOME
- KEGG_CELL_CYCLE
- KEGG_DNA_REPLICATION
- KEGG_MISMATCH_REPAIR

- KEGG_GRAFT_VERSUS_HOST_DISEASE
- KEGG_INTESTINAL_IMMUNE_NETWORK_FOR_IGA_PRODUCTION
- KEGG_ALLOGRAFT_REJECTION
- KEGG_AUTOIMMUNE_THYROID_DISEASE

High risk<----------->Low risk

B GSE72094

- KEGG_MISMATCH_REPAIR
- KEGG_BASE_EXCISION_REPAIR
- KEGG_CELL_CYCLE
- KEGG_PYRIMIDINE_METABOLISM

- KEGG_OTHER_GLYCAN_DEGRADATION
- KEGG_FC_EPSILON_RI_SIGNALING_PATHWAY
- KEGG_GNRH_SIGNALING_PATHWAY
- KEGG_VALINE_LEUCINE_AND_ISOLEUCINE_DEGRADATION

High risk<----------->Low risk

Supplementary Figure 4. Gene set enrichment analysis (GSEA) of the mRNAs associated with the high or low risk group in two LUAD cohorts. (**A**, **B**) Top enriched KEGG pathways in the high risk group are represented by the curves above the x-axis in the graph. Top enriched KEGG pathways in the low risk group are represented by the curves below the x-axis in the graph (p-value < 0.05) for two LUAD cohorts: TCGA (**A**) and GSE72094 (**B**). The names of enriched KEGG pathways are listed on the right side.
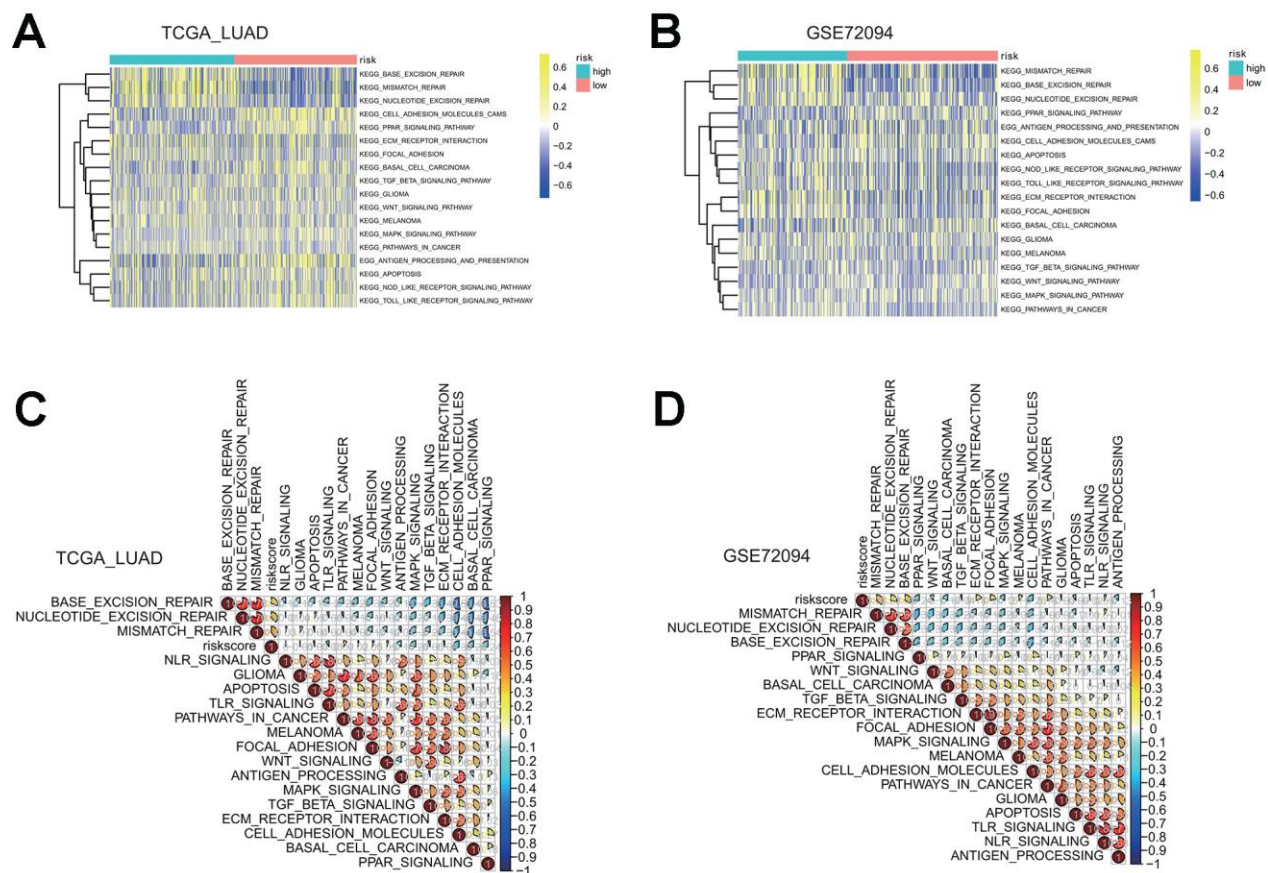
Supplementary Figure 5. GSVA and correlogram analysis of KEGG biological pathways based on risk groups or risk scores in two LUAD cohorts. (**A**, **B**) Heatmaps visualizing the enrichment levels of KEGG biological pathways in the high or low risk group in the LUAD cohorts: TCGA (**A**) and GSE72094 (**B**). Yellow represents enriched pathways, and blue represents repressed pathways. (**C**, **D**) Correlograms demonstrating the correlation coefficient between the risk score and KEGG pathways in the LUAD cohorts: TCGA (**C**) and GSE72094 (**D**). Red color depicts the positive correlation, and blue color depicts the negative correlation.

A
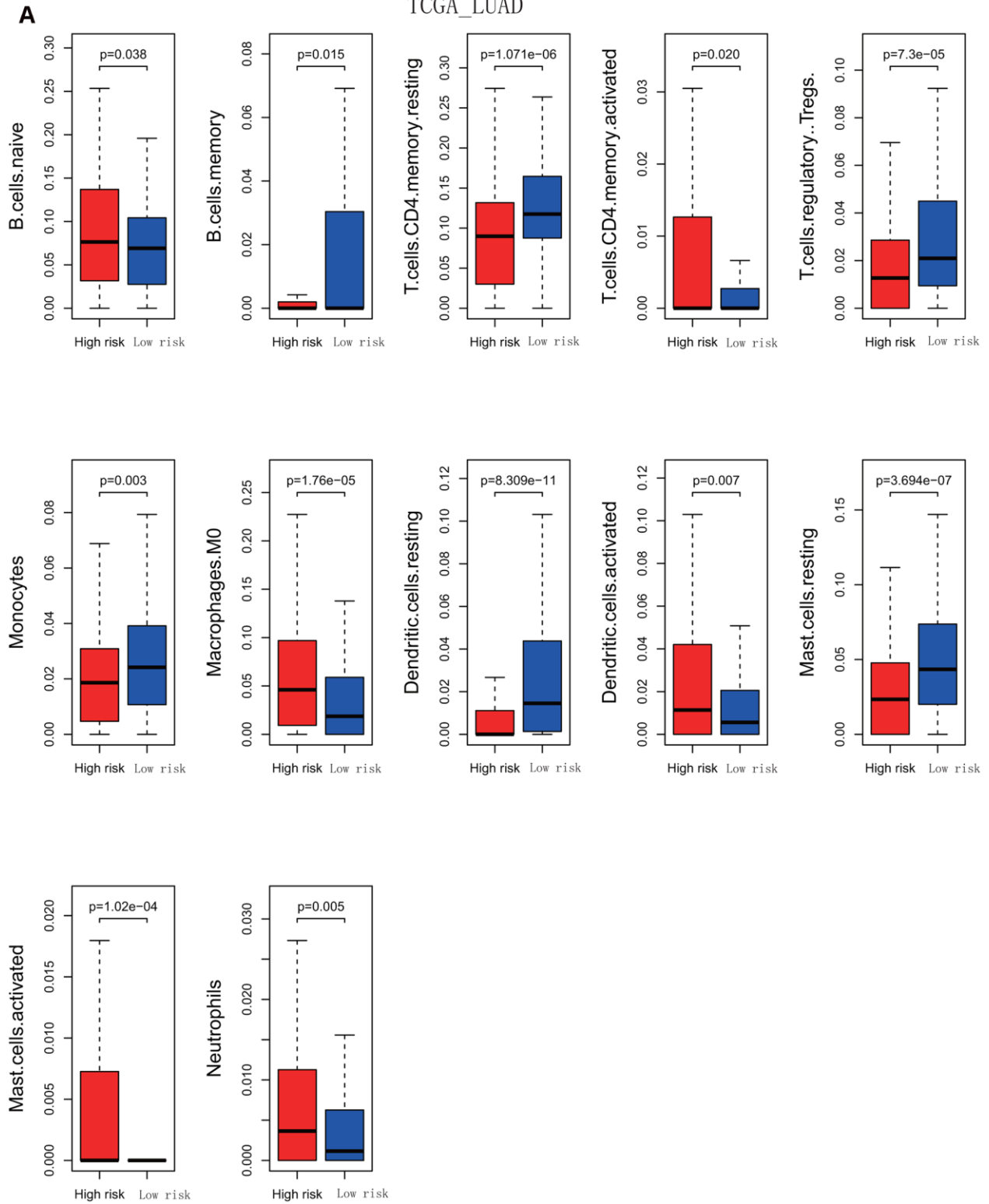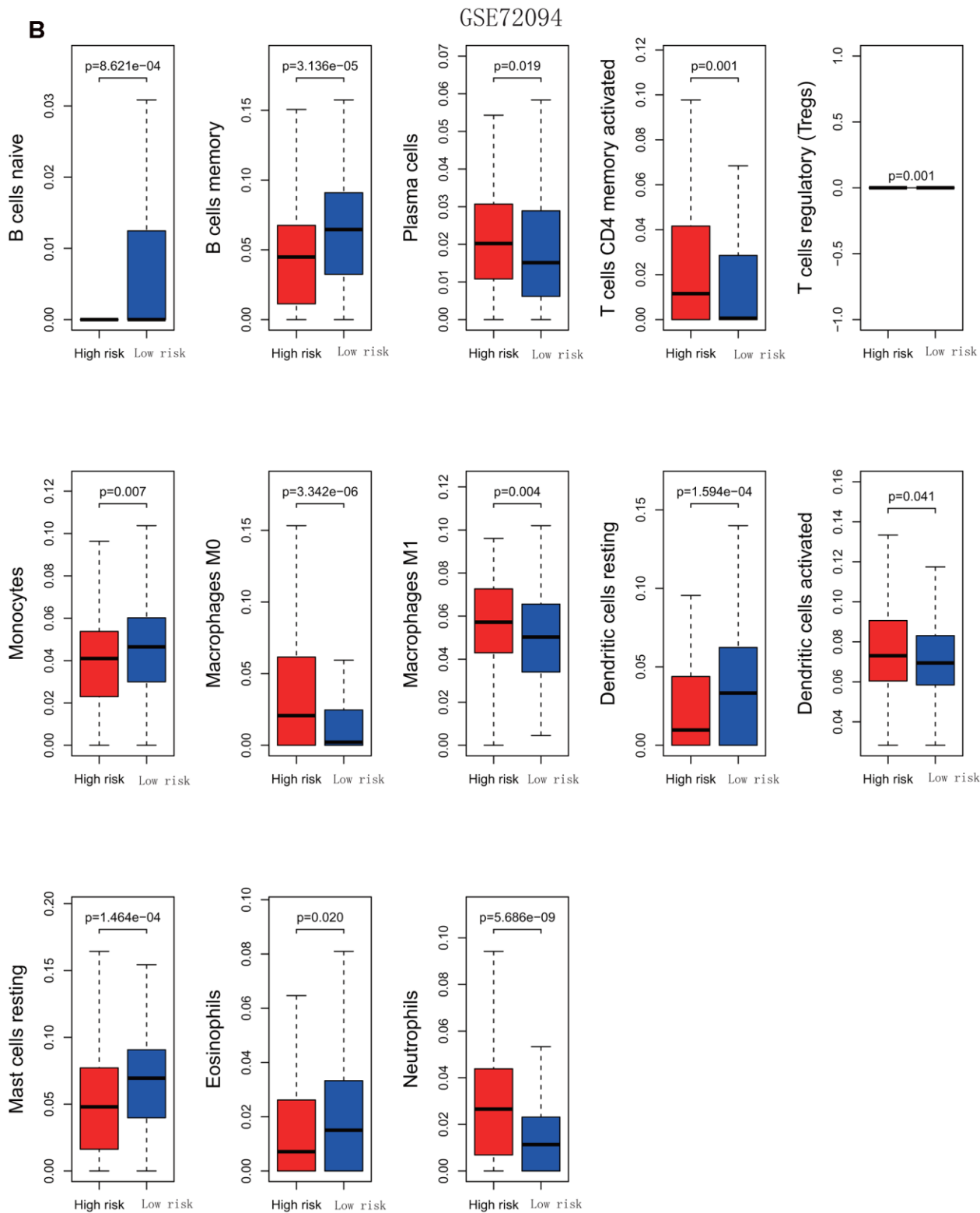
Supplementary Figure 6. Comparison of relative immune cell enrichment levels based on the CIBERSORT analysis results. (**A**, **B**) Bar plots show the statistically significant difference in the enrichment levels of immune cells between the high (red color) and low (blue color) risk groups in two LUAD cohorts: TCGA (**A**) and GSE72094 (**B**).

TCGA_LUAD
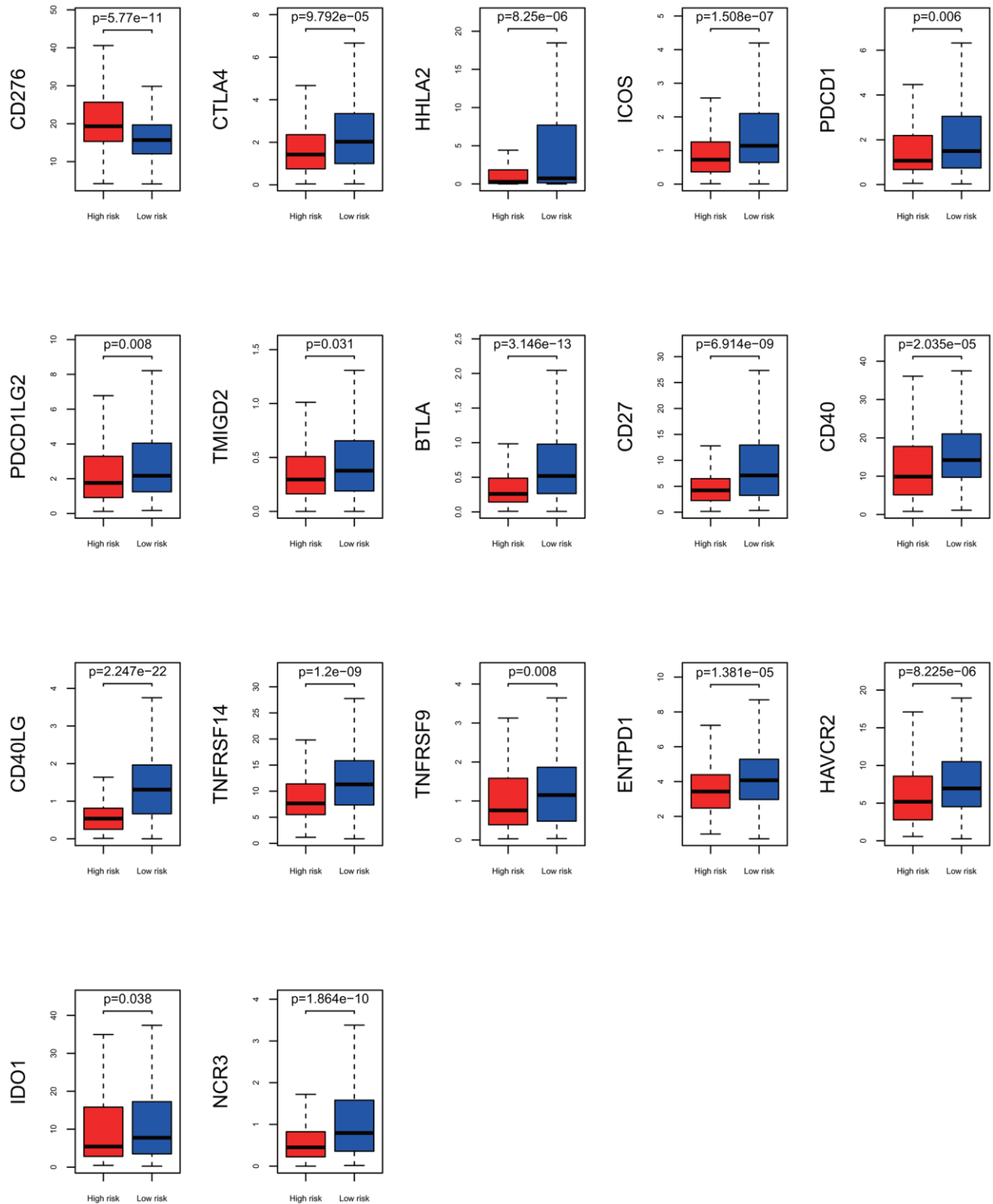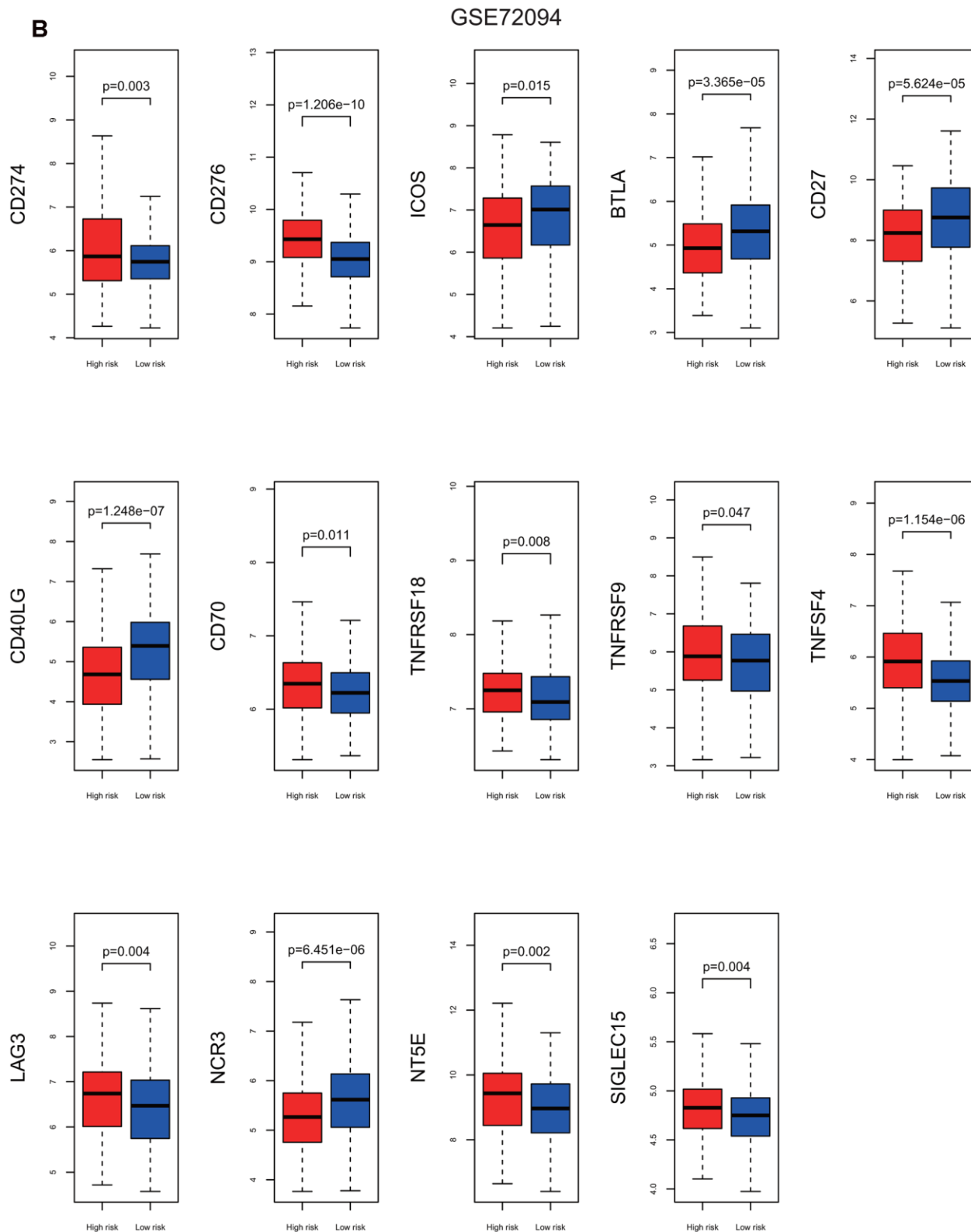
Supplementary Figure 7. Comparison of gene expression of 29 immune checkpoint genes in the LUAD cohorts. (**A**, **B**) Bar plots show the statistically significant difference in gene expression levels of immune checkpoint genes between the high (red color) and low (blue color) risk groups in two LUAD cohorts: TCGA (**A**) and GSE72094 (**B**).
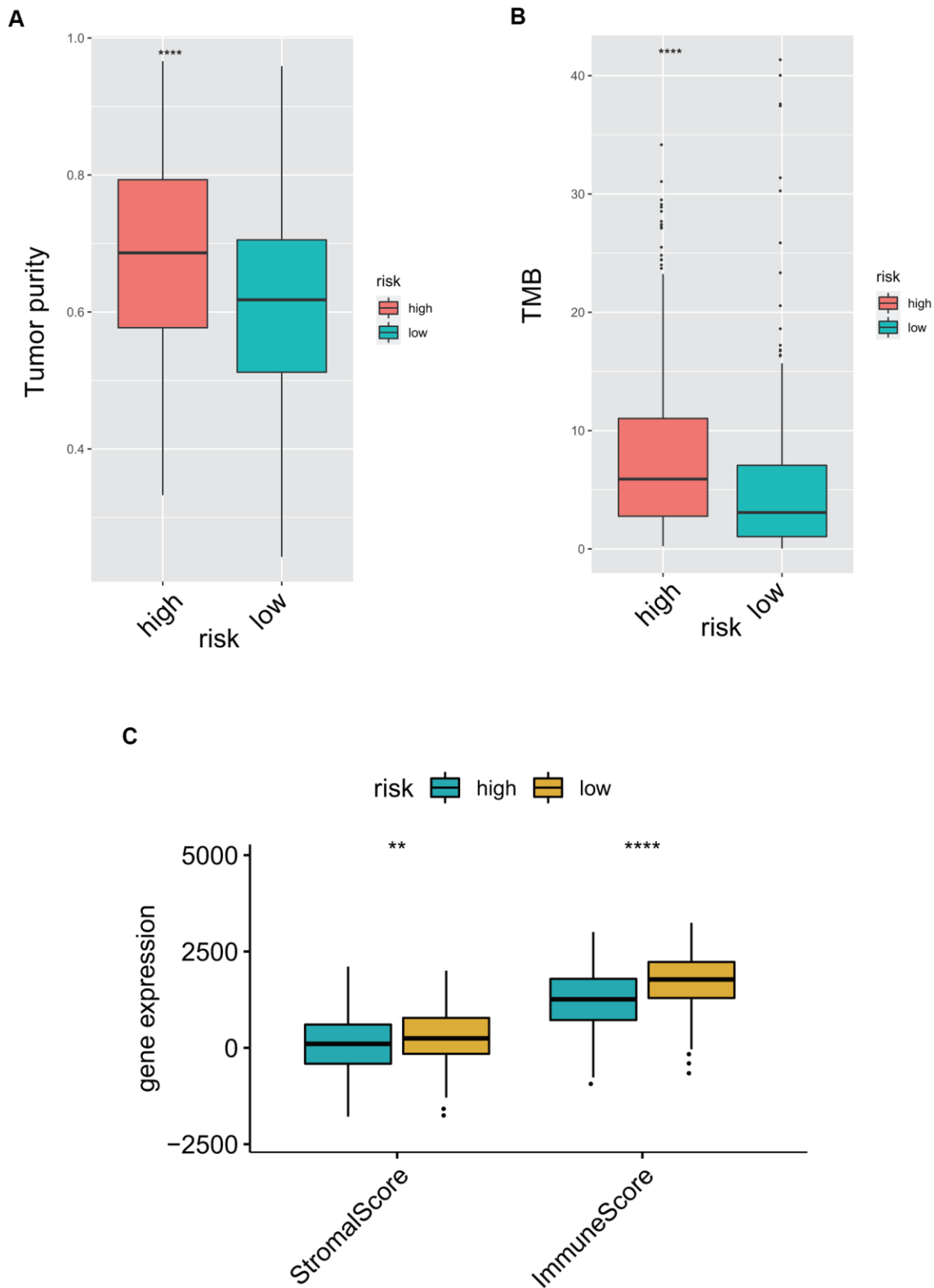
Supplementary Figure 8. Comparison of tumor purity, tumor burden, stromal score, or immune score between the high and low risk group in the TCGA LUAD cohort. (**A**, **B**) Bar plots show the comparison of tumor purity (**A**) and tumor burden (TMB) (**B**) between the high and low risk groups in the TCGA LUAD cohort. (**C**) Comparison of stromal score and immune score between the high and low risk group in the TCGA LUAD cohort.

Supplementary Tables

Supplementary Table 1. Clinical features of two LUAD cohorts.

	Death (N=154)	alive (N=322)	Overall (N=476)
Gender			
Male	75 (48.7%)	144 (44.7%)	219 (46.0%)
Female	79 (51.3%)	178 (55.3%)	257 (54.0%)
Age (years)			
Mean (SD)	65.7 (10.7)	65.0 (9.69)	65.2 (10.0)
Median [Min, Max]	67.0 [40.0, 88.0]	66.0 [33.0, 87.0]	66.0 [33.0, 88.0]
Stage			
I	53 (34.4%)	206 (64.0%)	259 (54.4%)
II	46 (29.9%)	70 (21.7%)	116 (24.4%)
III	42 (27.3%)	37 (11.5%)	79 (16.6%)
IV	13 (8.4%)	9 (2.8%)	22 (4.6%)
T			
T1	39 (25.3%)	123 (38.2%)	162 (34.0%)
T2	87 (56.5%)	168 (52.2%)	255 (53.6%)
T3	18 (11.7%)	24 (7.5%)	42 (8.8%)
T4	10 (6.5%)	7 (2.2%)	17 (3.6%)

(a)Clinical features of the TCGA LUAD cohort.

	Death (N=109)	alive (N=280)	Total (N=389)
Gender			
Male	58 (53.2%)	114 (40.7%)	172 (44.2%)
Female	51 (46.8%)	166 (59.3%)	217 (55.8%)
Age (years)			
Mean (SD)	70.4 (8.54)	69.1 (9.75)	69.5 (9.43)
Median [Min, Max]	72.0 [46.0, 85.0]	70.0 [38.0, 89.0]	70.0 [38.0, 89.0]
Stage			
I	51 (46.8%)	201 (71.8%)	252 (64.8%)
II	24 (22.0%)	41 (14.6%)	65 (16.7%)
III	26 (23.9%)	31 (11.1%)	57 (14.7%)
IV	8 (7.3%)	7 (2.5%)	15 (3.9%)
KRAS			
Wildtype	64 (58.7%)	190 (67.9%)	254 (65.3%)
Mutation	45 (41.3%)	90 (32.1%)	135 (34.7%)
EGFR			
Wildtype	105 (96.3%)	244 (87.1%)	349 (89.7%)
Mutation	4 (3.7%)	36 (12.9%)	40 (10.3%)
STK11			
Wildtype	91 (83.5%)	234 (83.6%)	325 (83.5%)
Mutation	18 (16.5%)	46 (16.4%)	64 (16.5%)
TP53			
Wildtype	80 (73.4%)	215 (76.8%)	295 (75.8%)
Mutation	29 (26.6%)	65 (23.2%)	94 (24.2%)

(b)Clinical features of the GSE72094 LUAD cohort.

Supplementary Table 2. Analysis of 24 significantly differentially expressed interferon gamma response genes in the TCGA LUAD cohort.

Gene	conMean	treatMean	logFC	pValue	change
CD69	8.936670479	4.47653308	-0.997355581	3.19E-12	Down
CD74	1897.142334	1258.20791	-0.592457585	3.76E-10	Down
CD86	9.607033362	6.223707114	-0.626316828	4.52E-10	Down
CDKN1A	71.73071206	41.03490112	-0.805739475	6.18E-07	Down
CIITA	5.058851136	4.29529146	-0.236053757	0.00187674	Down
CSF2RB	9.875387254	5.517276061	-0.839881155	7.30E-15	Down
IL10RA	9.701275787	6.926449937	-0.486058378	2.96E-07	Down
IRF8	10.10959639	5.46877443	-0.88643594	2.31E-15	Down
LATS2	10.7276258	5.178535474	-1.050714763	1.56E-23	Down
LCP2	8.387419434	6.207701793	-0.434167749	4.55E-08	Down
MT2A	171.2990976	107.9445696	-0.666226884	0.000272554	Down
NOD1	4.764992505	4.034663216	-0.240025692	1.65E-05	Down
PTPN6	21.23585905	15.13571205	-0.488545923	6.81E-08	Down
SELP	10.56550028	2.27969876	-2.212445977	1.03E-29	Down
SOD2	31.28827632	26.96227591	-0.214679901	0.009578905	Down
IRF4	1.273380295	2.803021931	1.138319688	0.00081257	Up
ITGB7	1.223934115	1.669670931	0.448037896	0.03900192	Up
NMI	7.067112293	9.363642417	0.405949008	9.72E-05	Up
OAS3	8.38112837	16.96725326	1.017536638	1.93E-08	Up
PFKP	7.575956699	36.9750878	2.287053583	1.61E-24	Up
PNP	11.79940954	15.63021418	0.405622882	1.59E-05	Up
RBCK1	21.51999684	26.09297107	0.27798336	0.001820364	Up
RIPK2	7.642719185	10.60564426	0.472674335	0.000781467	Up
TRAFD1	13.22724719	16.44497774	0.314134211	3.75E-05	Up

Differential gene expression was analyzed using the 'limma' R package. Information, including gene name, mean gene expression in normal (conMean) or cancer (treatMean) samples, log fold change (logFC) of cancer vs. normal, p value (pValue) of logFC, and gene expression change pattern in cancer, were listed in the table.