

Research Paper

Integration of proteomics and transcriptomics to construct a prognostic signature of renal clear cell carcinoma

Guangyang Cheng^{1#}, Zhaokai Zhou^{1,2,3#}, Shiqi Li^{1#}, Zhuo Ye¹, Yan Wang¹, Jianguo Wen^{1,2,3}, Chuanchuan Ren¹✉

1. Department of Urology, The First Affiliated Hospital of Zhengzhou University, Henan, Zhengzhou 450052, China.
2. Henan Joint International Pediatric Urodynamic Laboratory, The First Affiliated Hospital of Zhengzhou University, Henan, Zhengzhou 450052, China.
3. Bladder Structure and Function Reconstruction Henan Engineering Laboratory, The First Affiliated Hospital of Zhengzhou University, Henan, Zhengzhou 450052, China.

#These authors contributed equally to this work.

✉ Corresponding author: **Chuanchuan Ren**. Department of Urology, The First Affiliated Hospital of Zhengzhou University, Zhengzhou 450052, China (E-mail: cc_ren@zzu.edu.cn).© The author(s). This is an open access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>). See <https://ivyspring.com/terms> for full terms and conditions.

Received: 2024.06.23; Accepted: 2024.08.06; Published: 2024.08.19

Abstract

Background: Protein information is often replaced by RNA data in studies to understand cancer-related biological processes or molecular functions, and proteins of prognostic significance in Kidney clear cell carcinoma (KIRC) remain to be mined.

Methods: The cancer genome atlas program (TCGA) data was utilized to screen for proteins that are prognostically significant in KIRC. Machine learning algorithms were employed to develop protein prognostic models. Additionally, immune infiltration abundance, somatic mutation differences, and immunotherapeutic responses were analyzed in various protein risk subgroups. Ultimately, the validation of protein-coding genes was confirmed by utilizing an online database and implementing quantitative real-time PCR (qRT-PCR).

Results: The patients were divided into two risk categories based on prognostic proteins, and notable disparities in both overall survival (OS) and progression free interval (PFI) were observed between the two groups. The OS was more unfavorable in the high-risk group, and there was a noteworthy disparity in the level of immune infiltration observed between the two groups. In addition, the nomogram showed high accuracy in predicting survival in KIRC patients.

Conclusion: In this research, we elucidated the core proteins associated with prognosis in terms of survival prediction, immunotherapeutic response, somatic mutation, and immune microenvironment. Additionally, we have developed a reliable prognostic model with excellent predictive capabilities.

Keywords: Proteomics; Transcriptomics; Renal clear cell carcinoma; Prognostic signature; tumor microenvironment

Introduction

Renal cancer is the most frequently occurring tumor of the urinary system[1], and also has the highest mortality rate in urological tumors[2]. Around 85% of tumors in the kidneys are renal cell carcinomas, with about 70% being clear cell carcinomas of the kidney. Renal cell carcinoma (RCC) makes up around 3.8% of all newly diagnosed cancers, typically affecting individuals at a median

age of 64 years[3]. However, those with distant metastases experience a significantly lower 5-year survival rate of 0% to 10%[4]. Despite the utilization of molecularly targeted treatments like inhibitors of tyrosine kinase and rapamycin for advanced renal cancer[5, 6], finding a solution to enhance the OS and progression-free survival (PFS) of patients remains a Gordian knot. Constructing novel prognostic models

could weigh better treatment modalities based on patient conditions to maximize clinical benefit and provide novel insights for precision medicine treatment decisions.

Proteins, as important mediators in biological processes and also responsible for the gene coding functions of cells, are demonstrated as a nexus with tumorigenesis and progression[7]. In most research, protein expression data are replaced by RNA sequencing data, but often with poor expected results[8]. To bridge the gap between genomic data and functional proteins, it is necessary to utilize proteomics and convert this information[9]. Recently, proteomic studies have enhanced insights into the biological basis and prognostic assessment of KIRC and revealed potential therapeutic targets[10, 11]. Therefore, further proteomic studies in KIRC, especially biomarker discovery, could provide novel prognostic molecular markers and effective treatment to win the war against KIRC.

For this research, we conducted transcriptomic and proteomic examinations from the TCGA-KIRC data. Comprehensive data analyses revealed different proteomic isoforms of KIRC, and a seven-protein-based prognostic model was constructed by linking the proteomic features of KIRC to clinical outcomes. In addition, we analyzed the relationship of protein prognostic models with immune infiltration and immune microenvironment. These results may provide a basis for predicting the outcome and therapeutic effects of KIRC.

Materials and methods

Data acquisition

The proteomic data, RNA-seq data, and clinical data of KIRC were downloaded from the genomic data commons (GDC) website. Missing values in the protein expression data were interpolated using the "impute" R package, and expression values for the same protein were averaged. For RNA-seq data, genes with overall low expression were removed and expression values for the same genes were averaged. For clinical data, patients with missing essential data were removed. The clinical characteristics of these samples are detailed in Supplementary Table S1.

Screening and validation of model proteins

The "caret" R package was used to randomly divide 474 patients into training and test groups according to 5:5. Proteins that showed a significant correlation with prognosis were identified through COX regression analysis ($p < 0.01$) in the training group. Subsequently, proteins were selected from the pool of prognosis-related proteins using least absolute shrinkage and selection operator (LASSO) regression.

In the end, a multivariate COX regression model was built and the proteins for the final model were obtained through stepwise regression. Risk scores were calculated after screening the modeled proteins using the following formula: $riskscore = \sum_{i=1}^n k_i * X_i$, where k and X denote the relative expression level of the protein and the regression coefficient. The sample was divided into two risk groups using the median of the prognostic protein risk score as the cutoff value.

Construction of nomogram

To facilitate the clinical application of prognostic models, we plotted nomogram using the "rms" and "regplot" R packages, along with calibration curves and receiver operating characteristic (ROC) curves.

Consistent cluster analysis of prognostic proteins

The samples were typed according to the expression of prognostic proteins using the "ConsensusClusterPlus" R package, and the number of samples typed was determined according to the optimal K -value.

Immune infiltration analysis, gene set enrichment analysis (GSEA), and tumor immune dysfunction and exclusion (TIDE)

GSEA analysis was performed using the "clusterProfiler" R package, and GSEA analysis using the "c2.cp.kegg.symbols" gene set downloaded from the molecular signatures database (MSigDB) database. CIBERSORT was utilized to conduct immune infiltration analysis, and a boxplot illustrating the variations in immune cell infiltration among various subgroups was generated. Immunotherapy responses were analyzed using the TIDE website.

Online database to validate expression of prognostic proteins and protein-coding genes

The human protein atlas (HPA) database was utilized to confirm the expression of prototype proteins in both normal and KIRC tissues. Additionally, the tumor immune single-cell hub (TISCH) database was employed to observe the expression of genes encoding proteins in various cell clusters. Furthermore, the cancer single-cell state atlas (CancerSEA) database was utilized to examine the relationship between the function of cancer cells and genes encoding proteins at the individual cell level. The BEST website was utilized to analyze the differential expression of genes.

Cell culture

Human renal cortical proximal tubular epithelial cells (HK-2) and human renal cell adenocarcinoma

cells (769-P, 786-O, and ACHN) were used for this study, both provided by Servicebio, Ltd (Wuhan, China). Among them, HK-2 and ACHN were cultured in MEM medium containing 10% fetal bovine serum (FBS) and 1% penicillin-streptomycin (P/S). 769-P and 786-O were cultured in RPMI 1640 medium containing 10% FBS and 1% P/S in a humidified incubator at 37°C with 5% CO₂.

RNA extraction and qRT-PCR

Total RNA was extracted from cells using RNAeasy™ Animal RNA Extraction Kit (RR0026, Beyotime, Shanghai, China). cDNA synthesis was performed using PrimeScript™ RT reagent Kit with gDNA Eraser (RR047A, Takara, Beijing, China), and real-time quantitative PCR was performed using TB Green® Premix Ex Taq™ II (RR820A, Takara, Beijing, China). qPCR primers were synthesized by Sangon Biotech, Ltd (Shanghai, China) (Table 1). β -actin was used as a standardized internal reference gene. The relative gene expression levels were calculated using the 2^{- $\Delta\Delta$ Ct} method.

Table 1. Primers for qRT-PCR.

Gene	Forward Primer Sequence (5' to 3')	Reverse Primer Sequence (5' to 3')
ERRFI1	GGGCAGGGTATCCATTCT	TCCCTCAACAAGAGCGCA
RPS6KB1	ATATTGCCATGAAGGTGCT	CGATGAAGGGATGCITTTACT
CDKN1A	CTGGCCCTCAAATCGT	CCGCTGCTAATCAAAGTGC
ACACA	GAAGGGGTTTTACICTGTC	GAGGATCGTATGGGGTCTT
BRAF	CAAATTCACCCAGTCCGT	GGTCTCGTTGCCAAAT
ACTB	TCTCCCAAGTCCACACAGG	GGCACGAAGGCTCATCA
PEA15	TGAGGAGGATGAGCTGGA	GGGAGTGGTCTGATGAAGG
DIABLO	CCTACCTGCGTGAGGATTG	GGATCTGCCCTCTTC

Statistical analysis

The prognostic significance was evaluated through Kaplan-Meier (K-M) analysis and COX analysis, comparing groups using the Wilcoxon test for numerical variables and the chi-square test for categorical variables. R4.3.2 was utilized for statistical analyses, with significance defined as $p < 0.05$ unless stated otherwise.

Results

KIRC prognosis-related protein screening

Firstly, we analyzed the proteomics data of TCGA-KIRC, and after screening, we finally obtained 474 valid samples and 36 prognostic proteins that were significantly correlated out of survival ($p < 0.01$) and visualized the results on volcano and forest plots (Figure 1A-B). Subsequently, the LASSO and COX regression model was constructed to further screen prognostic proteins, and the final model identified the best performance when 7 proteins (AC1, P21,

P70S6K_pT389, SMAC, BRAF_pS445, MIG6, PEA15) were identified (Figure 1C-D). Patients were categorized into two groups based on the calculated risk score for each patient using regression coefficients, with the median risk score serving as the cutoff value. Principal component analysis (PCA) found that risk score could cluster samples well (Figure 1E-F).

Prognostic protein model validation

To verify the precision of the model, we also computed the risk scores for the validation group. Next, we conducted a survival analysis on the two groups and illustrated the disparities in survival using K-M curves. To assess the precision of the model, the corresponding ROC graphs were generated. In all cohorts, it was discovered that the high-risk score group had a considerably worse prognosis (Figure 2A-C, G-I). According to the ROC curves, the prognostic protein model exhibited a positive predictive performance for OS and PFI (OS-5 Year: Train Cohort: 0.810, Test Cohort: 0.775, TCGA whole Cohort: 0.788, PFI-5 Year: Train Cohort: 0.784, Test Cohort: 0.721, TCGA Whole Cohort: 0.743, Figure 2D-F, Figure 3A-C). Moreover, we plotted the sample distribution of risk scores for the three cohorts according to subgroups, and the expression of the modeled proteins is shown in the heatmap (Figure 3D-F). Finally, we examined the variations in survival rates among different risk subgroups based on clinical factors. The findings indicated that high-risk patients had unfavorable survival outcomes in subgroups categorized by age (≤ 65 and > 65), gender, tissue Grade levels, clinical stages, and TNM stages ($p < 0.001$, Figure S1A-E). However, the survival differences were not statistically significant in the N1 stage ($p = 0.082$, Figure S1F) and M1 stage ($p = 0.057$, Figure S1G).

Somatic mutation profile of tumor samples

To uncover genetic changes in the subgroups with high and low risks, and understand the connection between genetic mutations and patient survival, we conducted a waterfall plot analysis using copy number variation (CNV) data for TCGA-KIRC (Figure 4A-B). The findings indicated that both subgroups with high and low risk demonstrated elevated rates of mutation. After classifying patients based on the median tumor mutational load (TMB), we observed that the high TMB group exhibited a more unfavorable prognosis compared to the low TMB group (Figure 4C). The final combined survival analysis of TMB grouping and risk score grouping showed that the high TMB and high-risk score group had the most unfavorable prognosis (Figure 4D).

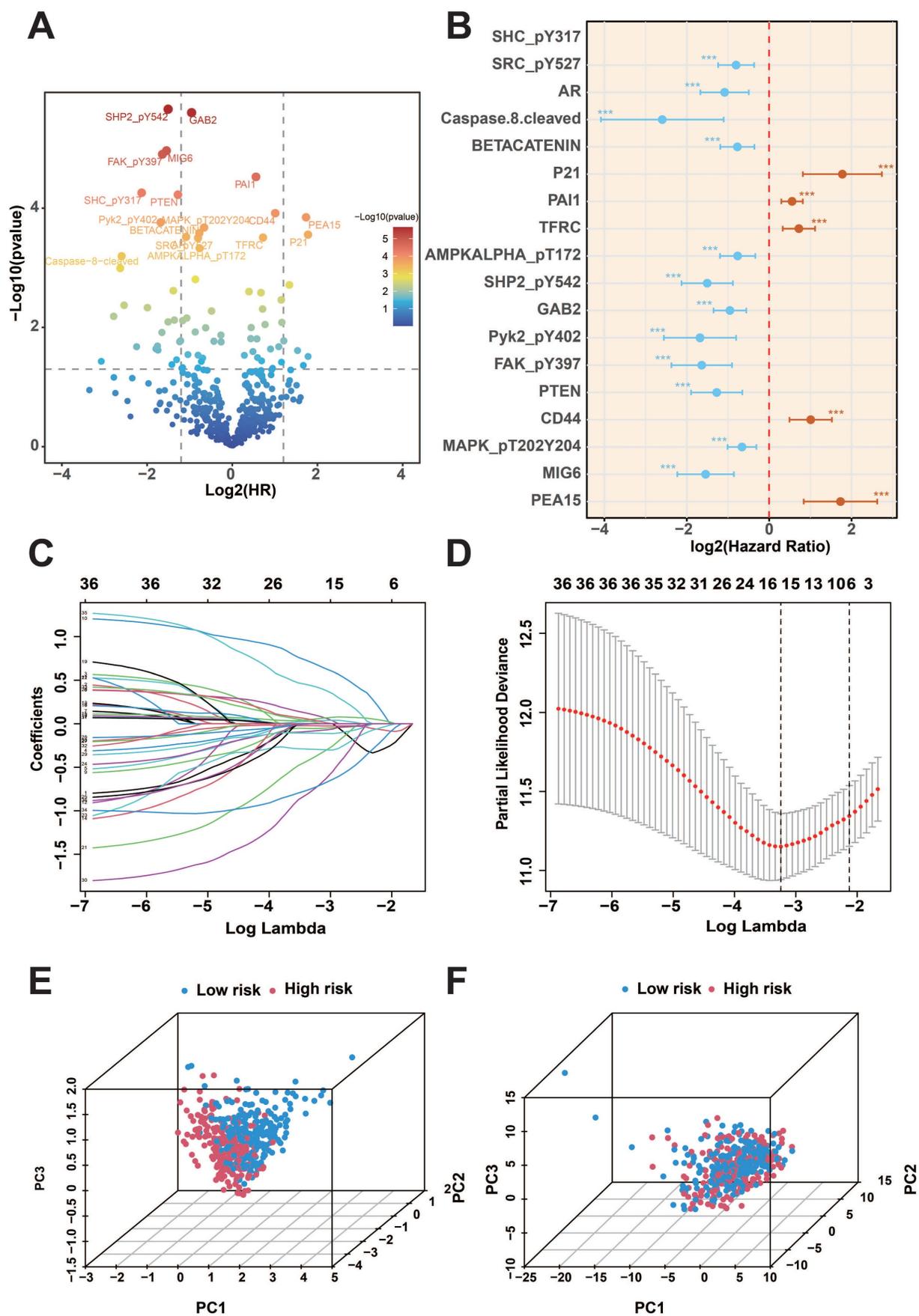


Figure 1. KIRC prognosis-related protein screening. A: COX regression analysis of protein prognostic volcano plots. B: Forest plot of prognostic-related proteins. C-D: Results of LASSO regression analysis and cross-validation. E-F: PCA analysis of risk scores. (***p*<0.001).

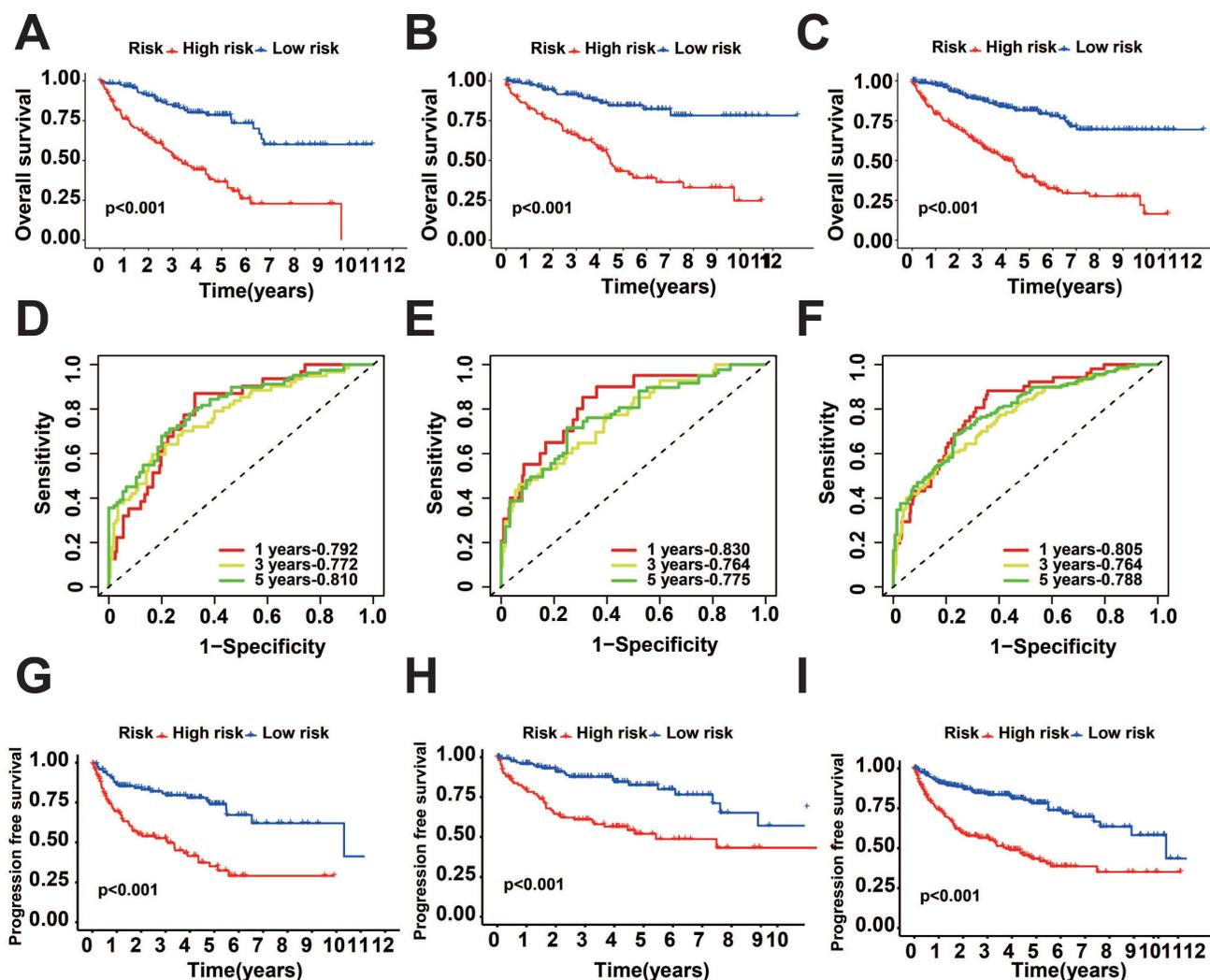


Figure 2. KIRC prognosis-related protein survival analysis. A-C: K-M survival curves for OS between the two groups in the training, test, and overall groups. D-F: ROC curves predicting OS in the training, test, and overall groups. G-I: K-M survival curves for PFI between the two groups in the training, test, and overall groups.

Clinical correlation of prognostic proteins

To further investigate the correlation between the 7 model proteins and the prognosis of KIRC, survival analyses were further performed. The results showed that high expression of SMAC, P21, PEA15, and ACC1 showed a more inferior prognosis (Figure 5B). In contrast, high expression of MIG6, P70S6K_pT389, and BRAF_pS445 showed a favorable prognosis (Figure 5A). This further demonstrated the feasibility of these seven proteins as prognostic markers. Clinical correlation analysis between the two risk subgroups showed that Grade, Stage, T, and M were significantly different between the two subgroups (Figure 5C). Co-expression analysis of the 7 predictive proteins revealed a total of 39 proteins that were significantly co-expressed ($r > 0.5$ and $p < 0.001$) (Figure S2A). The correlation analysis of the 7 prognostic proteins is visualized in the ring diagram (Figure S2B).

Construction of nomogram

To better predict the prognosis of patients with KIRC, we performed uni- and multivariate Cox regression analysis, and the results showed that risk score, Age, Grade, and Stage were significantly correlated with the prognosis (Figure 6A). Subsequently, we utilized these clinical characteristics and risk scores to construct a nomogram. Calculating the total score from the scores of each independent factor in the nomogram could accurately predict the survival of patients (Figure 6B). The nomogram exhibits a promising predictive performance, as indicated by the calibration curves and ROC curves (Figure 6C, D).

Evaluation of prognostically relevant protein subgroups

We performed a consistent clustering analysis based on the expression of model proteins to further

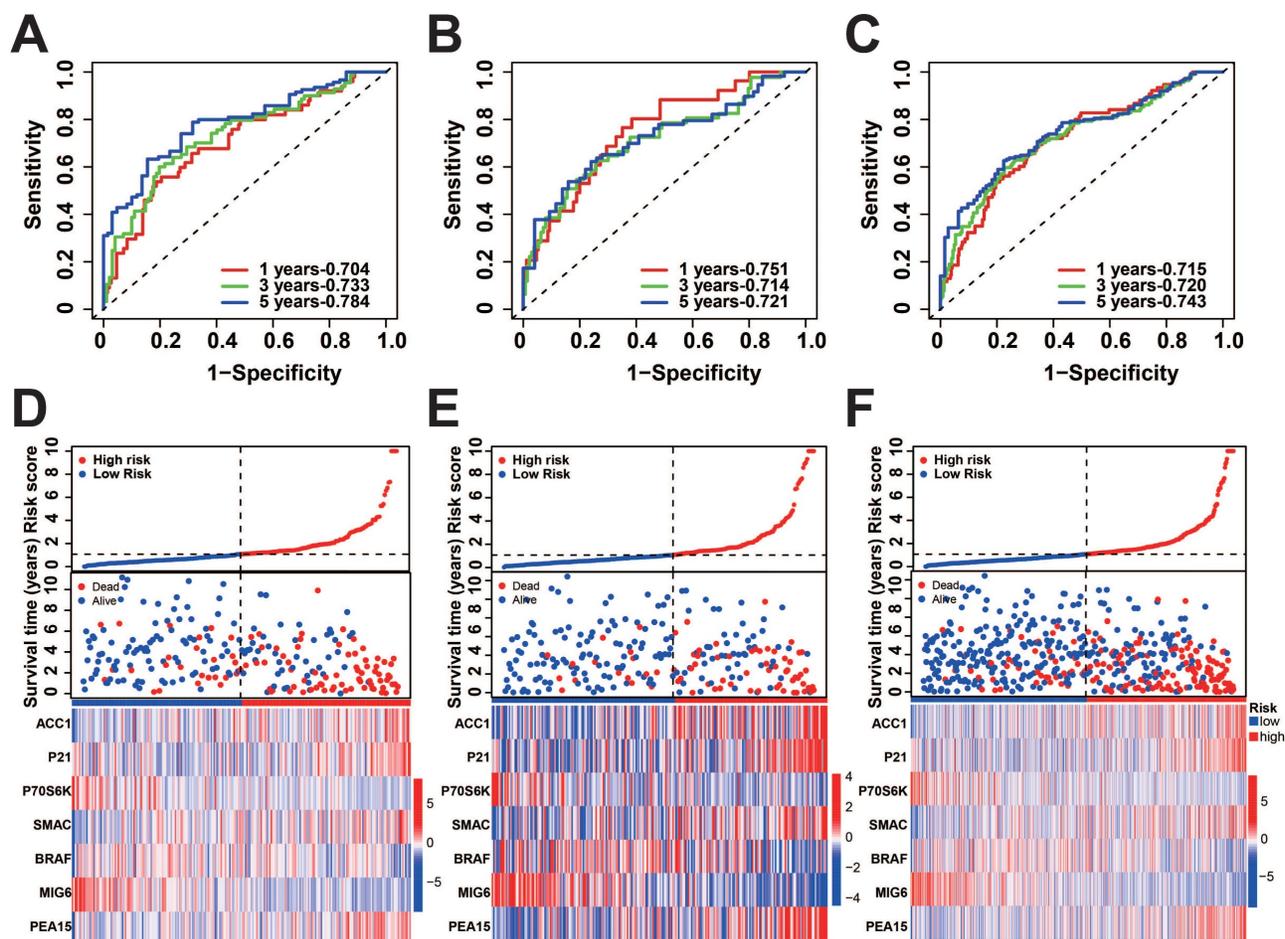


Figure 3. Distribution of risk score. A-C: Predicted ROC curves for PFI in the training, test, and overall groups. D-F: Distribution of risk scores across cohorts.

identify the subgroups of model proteins. We found that the intra-group correlation was the most obvious and the inter-group correlation was lower when $k=2$ (Figure 7A-C). Afterward, an analysis of survival was conducted on the two subgroups of prognostic proteins. The results showed that the C1 group had a considerably more favorable prognosis (Figure 7D). Heatmaps illustrate the expression of model proteins and protein-coding genes across various subgroups. (Figure 7E-F).

Immune cell infiltration and immune checkpoint analysis

The examination of the distribution of immune subtypes in prognostic protein groupings, which were categorized into six subtypes in previous studies, revealed that the high-risk group had a higher prevalence of C1 (healing of wounds), C2 (dominated by IFN-g), C4 (depleted of lymphocytes), and C6 (dominated by TGF-b). Additionally, the low-risk group had a higher percentage of C3 (inflammatory) (Fig. 8A). Similar differences were found between prognostic protein C1 and C2 subgroups (Figure 8B).

Immune infiltration analysis of transcriptomic data using CIBERSORT and MCPcounter showed significant differences in both risk subgroups and protein subgroups ($p < 0.05$, Figure 7C-F). According to the tumor micro-environment (TME) analysis, the high-risk group exhibited considerably elevated immune scores ($p < 0.05$, Figure 8C). Additionally, the stromal scores of the C1 subgroup were significantly higher than those of the C2 subgroup ($p < 0.001$, Figure 8D). TIDE analysis showed a high percentage of immunotherapy nonresponse in high-risk patients and the C2 subgroup (Figure S1H).

Meanwhile, notable variations in immune checkpoint manifestation were observed among the two risk groups as well as the protein subgroups ($p < 0.05$, Figure 8G, H). In the high-risk group, GSEA revealed enrichment of complement and coagulation cascades, cytokine receptor interaction, ECM receptor interaction, hematopoietic cell lineage, and p53 signaling pathway (Figure 9A). In addition, fatty acid metabolism, neurotrophic protein signaling pathways, propionic acid metabolism, type II diabetes, and valine, leucine, and isoleucine

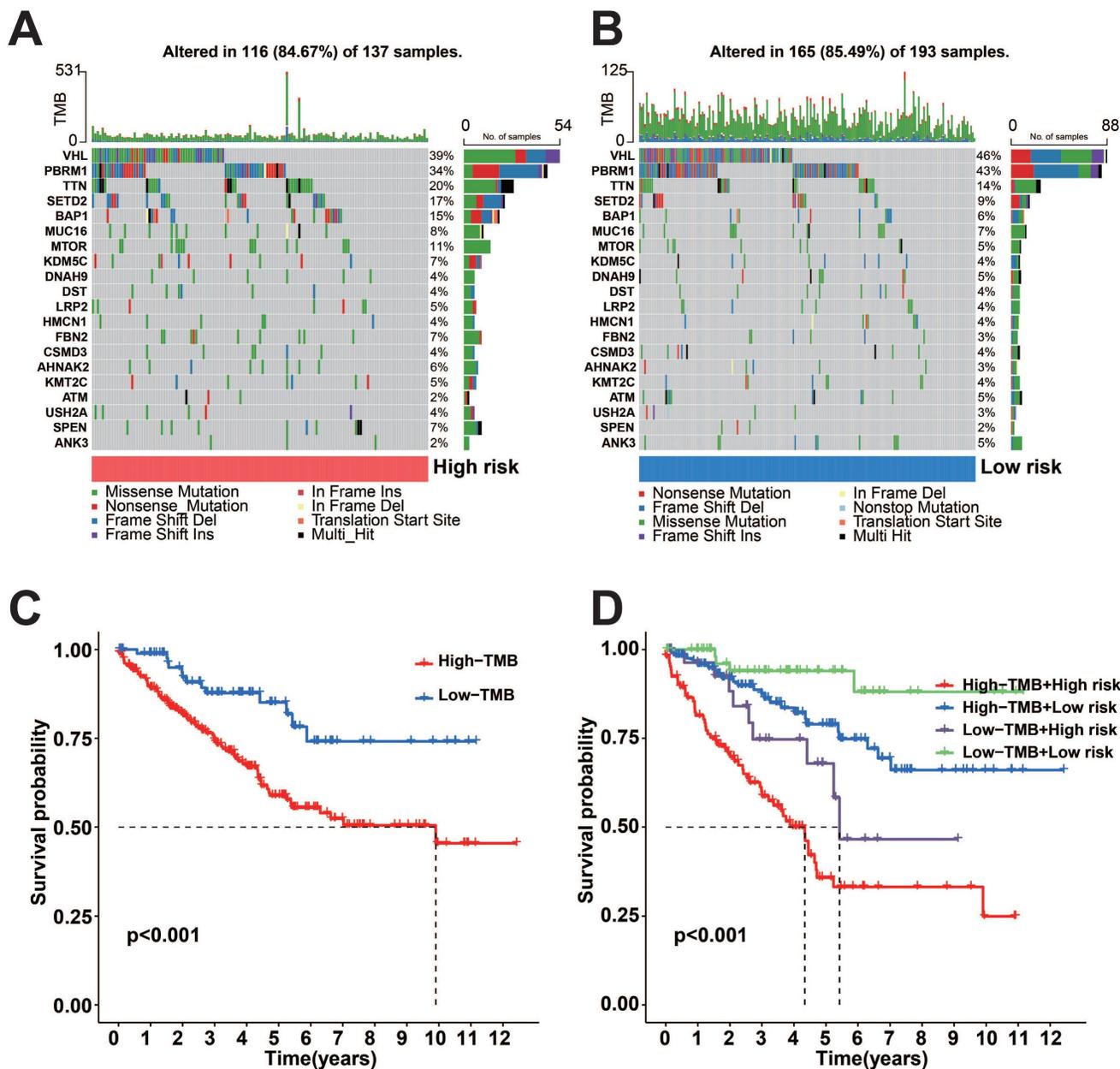


Figure 4. Somatic mutation analysis and survival analysis of model proteins in two risk subgroups. A-B: Gene mutations in the two risk subgroups. C: K-M survival curves for OS between high and low TMB groups. D: K-M survival curves for OS between the two TMB groups and the two risk groups.

degradation pathways were enriched in the low-risk group (Figure 9B). Figure 9C shows that the prognostic protein C1 subgroup had enriched pathways including focal adhesion, interaction of neuroactive ligands and receptors, neurotrophin signaling pathway, pathways related to cancer, and contraction of vascular smooth muscle. In Figure 9D, the complement and coagulation cascades, oxidative phosphorylation, steroid hormone biosynthesis, and vibrio cholerae infection pathways were found to be more abundant in the C2 subgroup.

The relationship between the expression of prognostic protein-coding genes and prognosis

We performed a survival analysis of these seven genes, which showed that all seven genes were significantly associated with prognosis, with high expression of DIABLO (encodes the SMAC protein) and ACACA (encodes the ACC1 protein) showing worse prognosis, while low expression of PEA15 (encodes the PEA15 protein), CDKN1A (encodes P21 protein), ERFFI1 (which encodes MIG6 protein), RPS6KB1 (encodes the P70S6K_pT389 protein) and BRAF (encodes the BRAF_pS445 protein), showed

worse prognosis ($p < 0.05$, Figure 10A-B). To examine the predictive effect of the seven protein-coding genes on prognosis, we constructed the Cox proportional hazards model using these seven genes. The survival analysis indicated that the model built using these seven genes exhibited outstanding predictive ability. K-M survival analysis showed worse prognostic performance in the high-risk score group ($p < 0.05$, Figure 10C-D). Analysis of 47 immune checkpoints revealed that 32 of these immune checkpoint genes exhibited distinct expression patterns ($p < 0.05$, Figure 10E). The analysis of ESTIMATE showed that the immune score and stromal score were higher in the high-risk group ($p < 0.05$, Figure 10F).

Analysis of prognostic protein-coding genes using online databases

We examined the HPA database to analyze the expression of protein-coding genes. The results indicated that ACACA exhibited low expression in KIRC tissues, while BRAF, CDKN1A, DIABLO,

PEA15, and RPS6KB1 showed high expression in KIRC tissues (Figure 11). To examine the variations in gene expression across various cell profiles, we utilized the TISCH database's KIRC_GSE111360 and KIRC_GSE159115 datasets to detect the expression levels of protein-coding genes (Figure 12A-B). The correlation between protein-coding genes and the functional status of tumor cells was analyzed using the CancerSEA database. The results revealed that ACACA was negatively correlated with the cell cycle ($r = -0.35$, $p < 0.05$, Figure 12C), CDKN1A was positively associated with metastasis ($r = 0.33$, $p < 0.01$, Figure 12D), RPS6KB1 was negatively related to invasion ($r = -0.38$, $p < 0.001$, Figure 12E), BRAF was positively linked to Stemness ($r = 0.39$, $p < 0.001$, Figure 12F), ERRFI1 was positively connected to epithelial-mesenchymal transition (EMT) ($r = 0.51$, $p < 0.01$, Figure 12G), and PEA15 was positively correlated with Quiescence ($r = 0.43$, $p < 0.01$, Figure 12H).

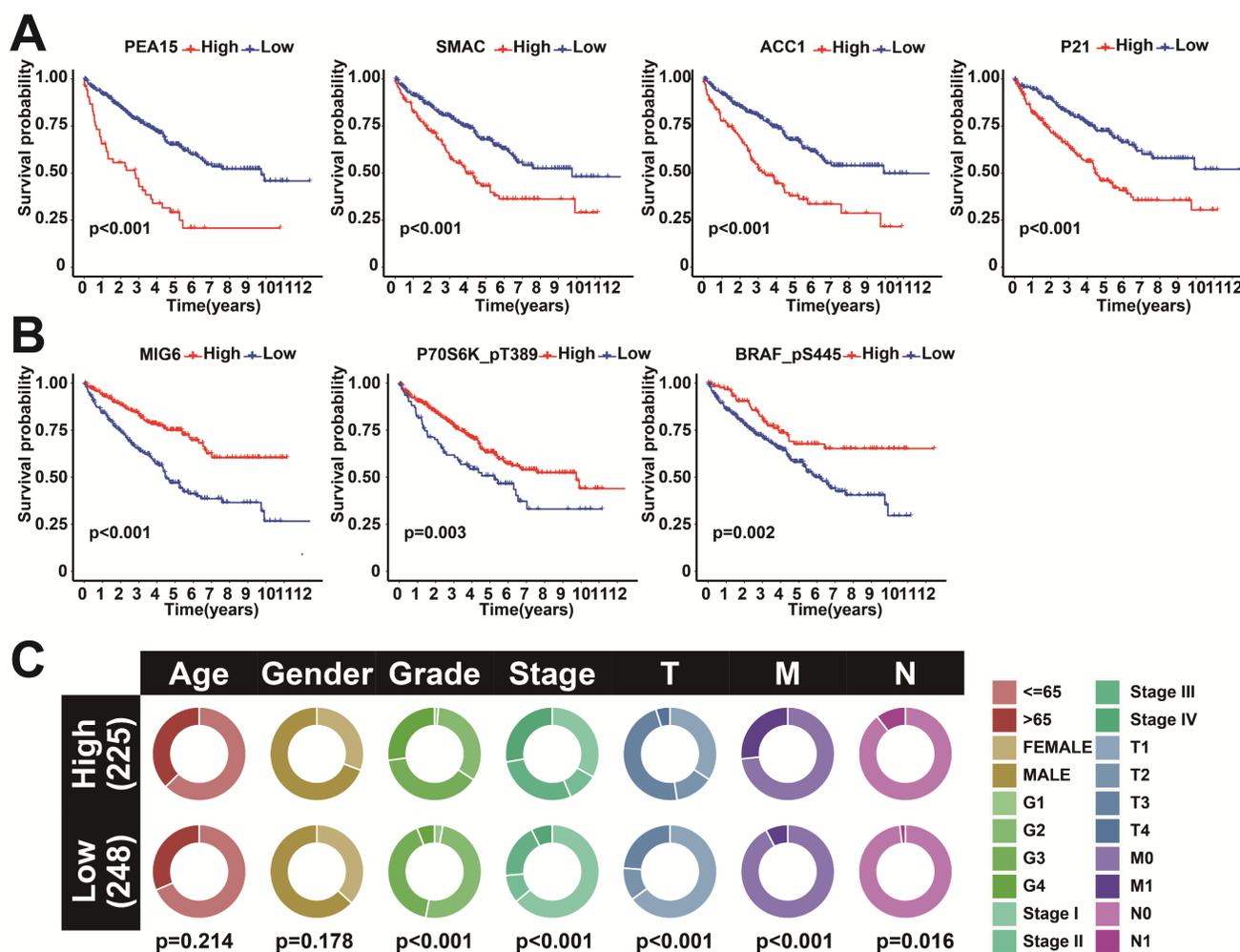


Figure 5. Survival analysis and interaction analysis of seven prognosis-related proteins. A: K-M survival curves between SMAC, P21, PEA15, and ACC1 expression and OS. B: K-M survival curves between MIG6, P70S6K_pT389, and BRAF_pS445 expression and OS. C: Correlation between the two risk score groups and clinical factors.

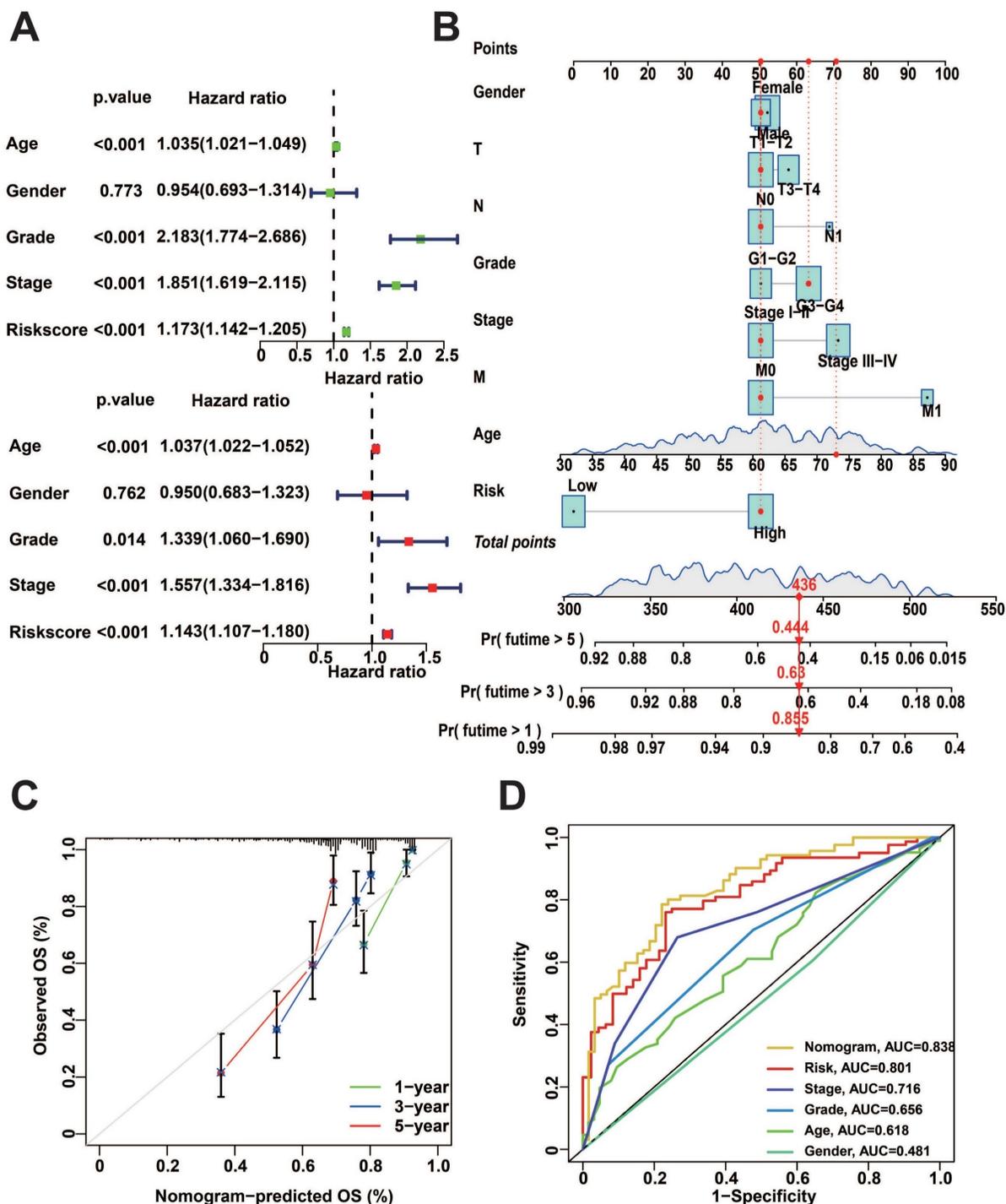


Figure 6. Nomogram construction and performance tests. A: Uni-variate and multi-variate COX regression analysis of clinical factors and risk score. B: Nomogram for predicting patient survival. C: Calibration curves for nomogram. D: Nomogram of ROC curves for OS prediction.

Verification of protein-coding genes by qRT-PCR

We analyzed the differences in expression of seven protein-coding genes between tumor and normal samples at the BEST website (Figure 13A-B). To further investigate the manifestation of these seven protein-encoding genes in cellular models, we analyzed their presence in four different cell

lines, namely HK-2, ACHN, 786-O, and 769-P, through the utilization of qRT-PCR. The results show that BRAF, CDKN1A, DIABLO, PEA15, ERFFI1, and RPS6KB1 are highly expressed in tumor cell lines, which is the same as in the TCGA database. Nevertheless, the cellular expression of ACACA varied from that observed in TCGA, potentially due to the disparity in sample size (Figure 13C).

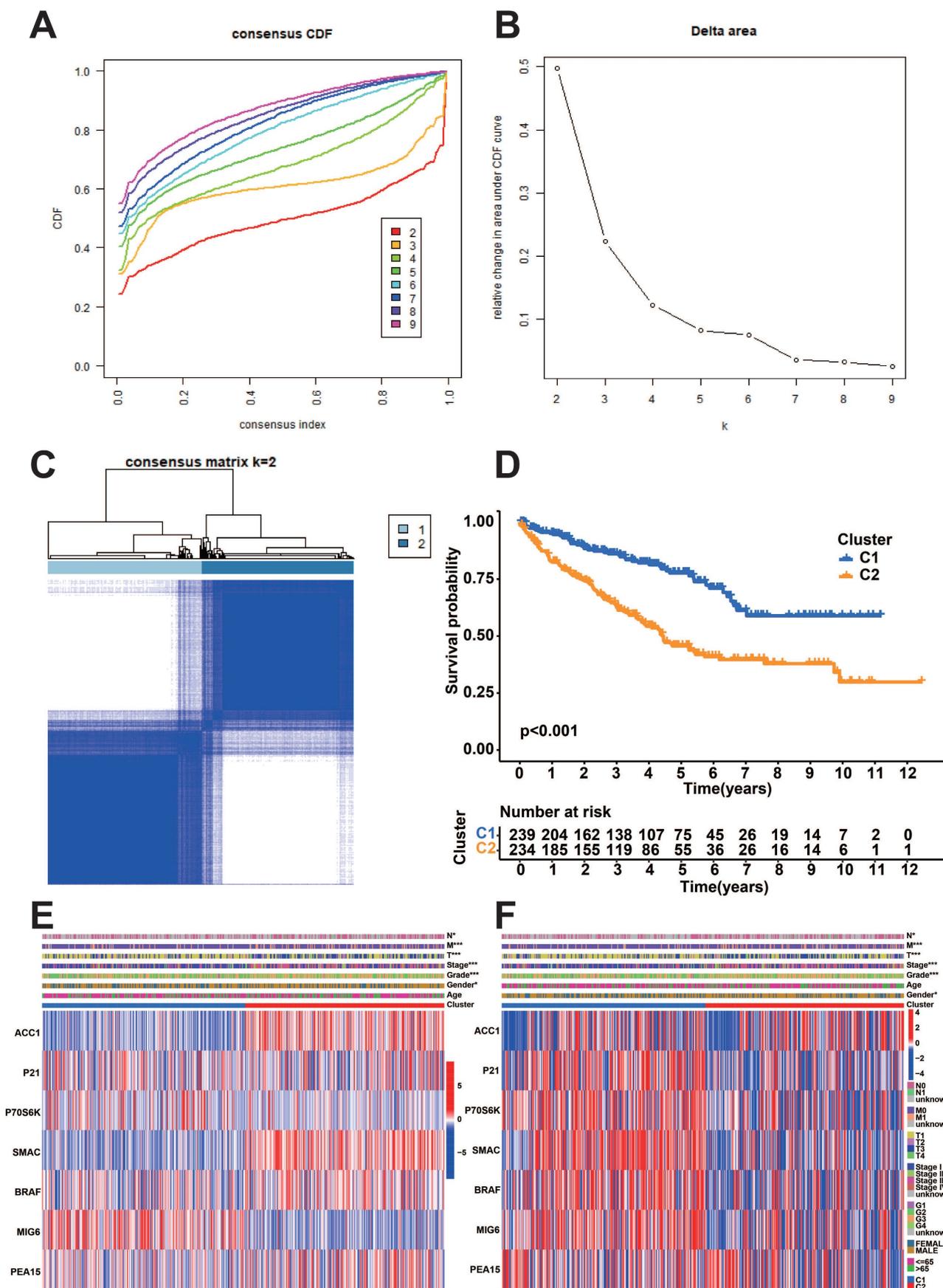


Figure 7. Consistency clustering analysis of prognostic proteins. A: Consistency clustering analysis cumulative distribution function (CDF) curve. B: K-means algorithm. C: Unsupervised clustering of seven model proteins and optimal consensus matrix with k = 2. D: Survival analysis of two prognostic protein subgroups. E-F: Heatmaps depicting the expression patterns of seven model proteins and protein-coding genes in two prognostic protein subgroups.

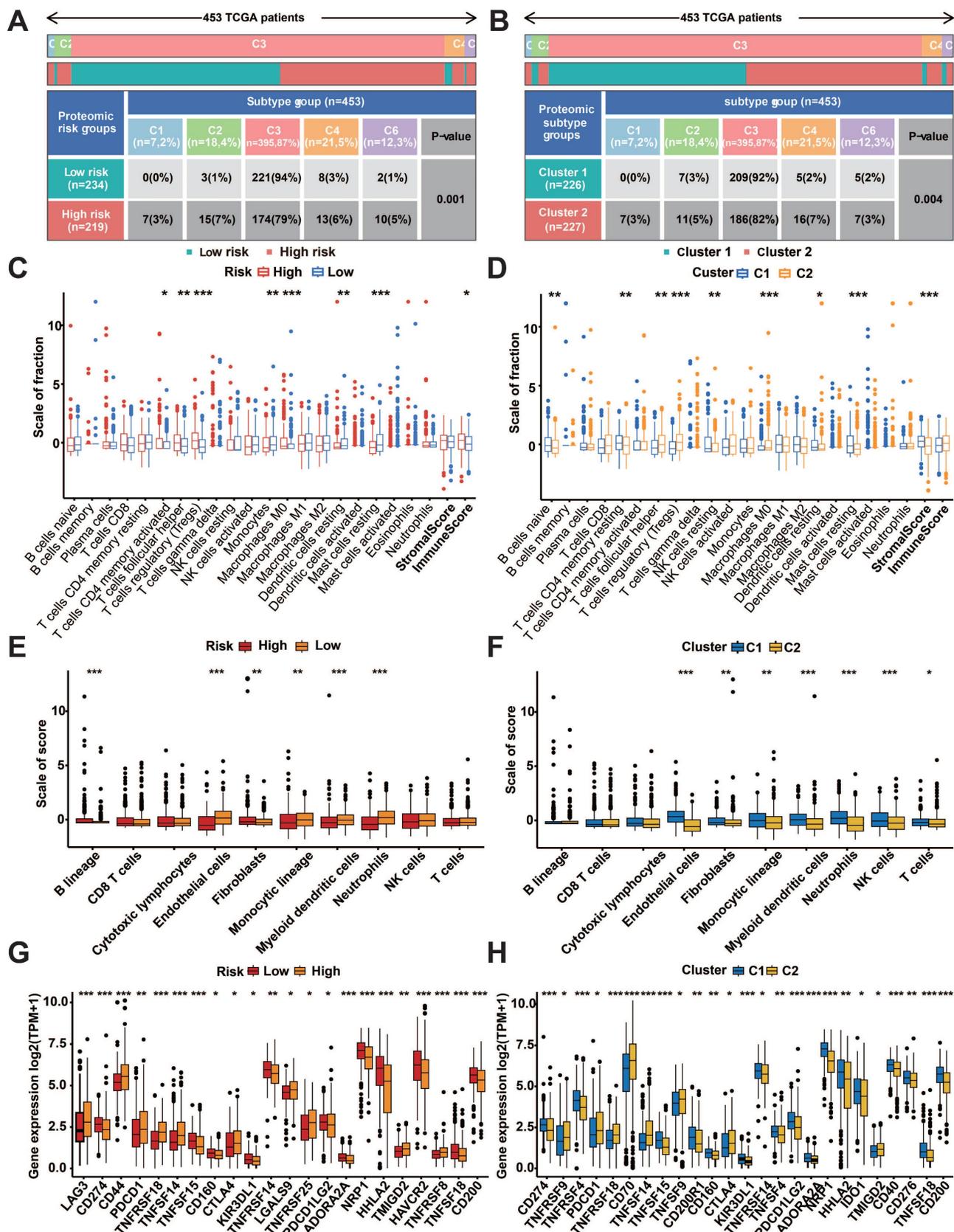


Figure 8. Immune infiltrates and subtypes of prognostic protein subgroups. A: Distribution of immune subtypes in risk subgroups. B: Immunosubtype distribution of prognostic protein subgroups C1 and C2. C: Divergences in the ratio of CIBERSORT immune cell infiltrations and TME scores among risk subgroups. D: Differences in proportions of CIBERSORT immune cell infiltration and TME scores for prognostic protein C1 and C2 subgroups. E: Divergences in the ratio of MCPcounter immune cell penetration among risk subgroups. F: The difference in the proportion of MCPcounter immune cell infiltration between prognostic protein C1 and C2 subgroups. G: Variations in the expression of immune checkpoint molecules among risk subgroups. H: Differences in immune checkpoint expression between prognostic protein C1 and C2 subgroups. (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$).

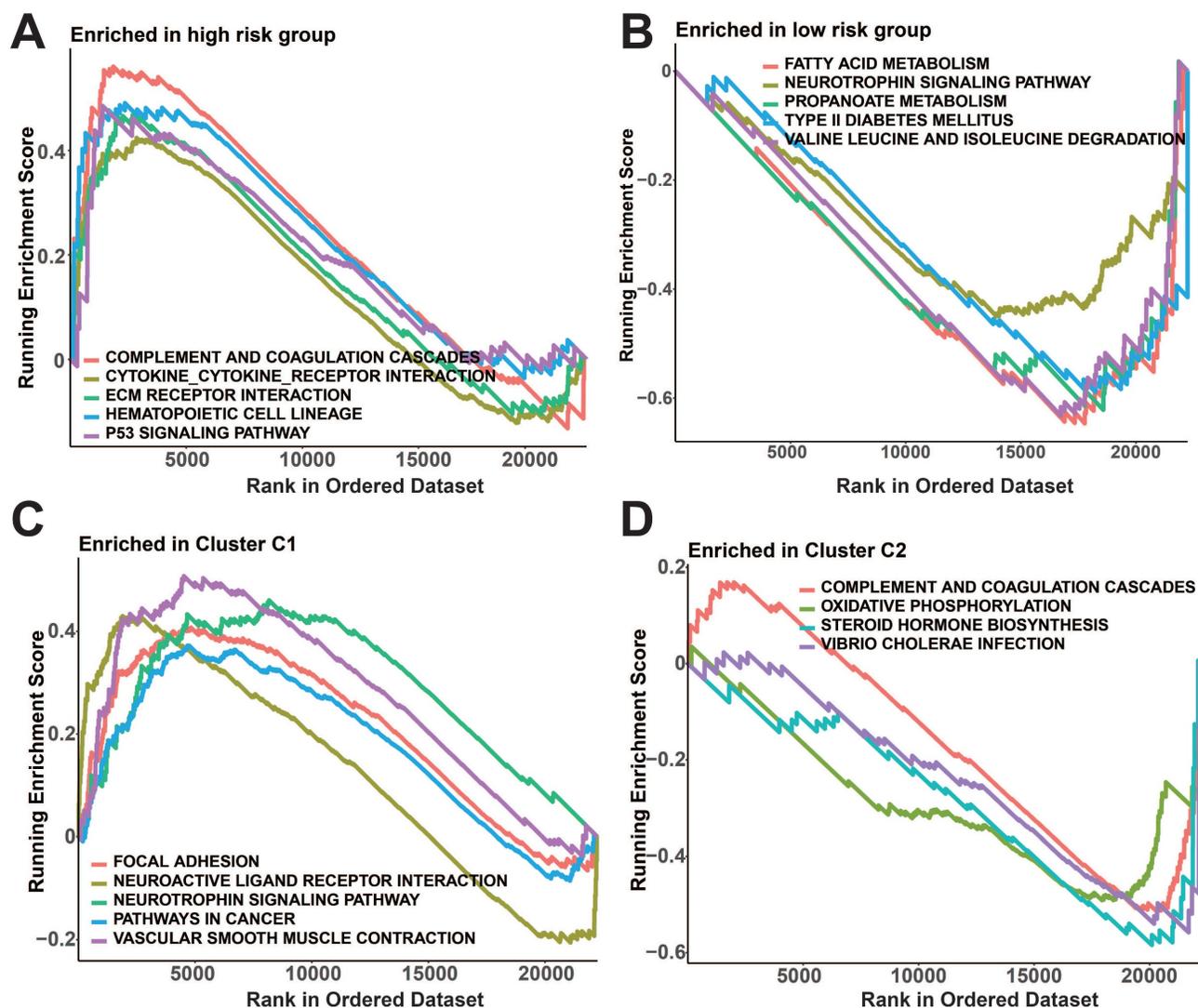


Figure 9. GSEA analysis of prognostic protein groupings. A-B: GSEA analysis of risk subgroups. C-D: GSEA analysis of protein subgroups.

Discussion

KIRC has an insidious onset, is heterogeneous, and is insensitive to conventional radiotherapy[12, 13]. Partial or radical nephrectomy is the primary approach for treating non-metastatic KIRC. After undergoing surgical treatment, around 25%~50% of patients still experience either local recurrence or distant metastasis[14-16]. At present, focused immunotherapies are employed for advanced renal cancer that has spread to distant areas[17], although the effectiveness of these treatments for KIRC patients remains restricted[18]. In this particular situation, healthcare professionals require more accurate biological indicators to establish risk categorization for anticipating patient prognosis and response to immunotherapy, as well as to offer direction for clinical diagnosis and treatment.

By conducting survival analysis and LASSO on the training cohort, we developed a prognostic

model that consists of seven proteins. The model's predictive performance was validated in the test cohort, showcasing its exceptional accuracy. Furthermore, the nomogram was developed to forecast survival by amalgamating the risk scores of the prognostic proteins with clinical factors including age, gender, tumor stage, and grading.

Earlier studies on the proteomic characterization of KIRC have demonstrated a strong association between proteomics and the tumor immune microenvironment. These studies have also revealed that distinct immune subtypes exhibit diverse clinical characteristics, prognoses, and responses to treatment[10]. Hence, we conducted analyses on immune cell infiltration and immune checkpoint expression in various risk groups and subtype subgroups, revealing noteworthy disparities among the different subgroups. Analysis of prognostic protein subtypes also indicated that the C1 cluster exhibited a

significantly more favorable prognosis compared to the C2 cluster. There are also studies showing that the Pearson correlation between protein and mRNA data was higher than 0.75 in immune and stromal-derived features[19]. Hence, we addition-

ally conducted a survival assessment on predictive protein-coding genes and developed a reliable prognostic model using them, demonstrating strong predictive accuracy.

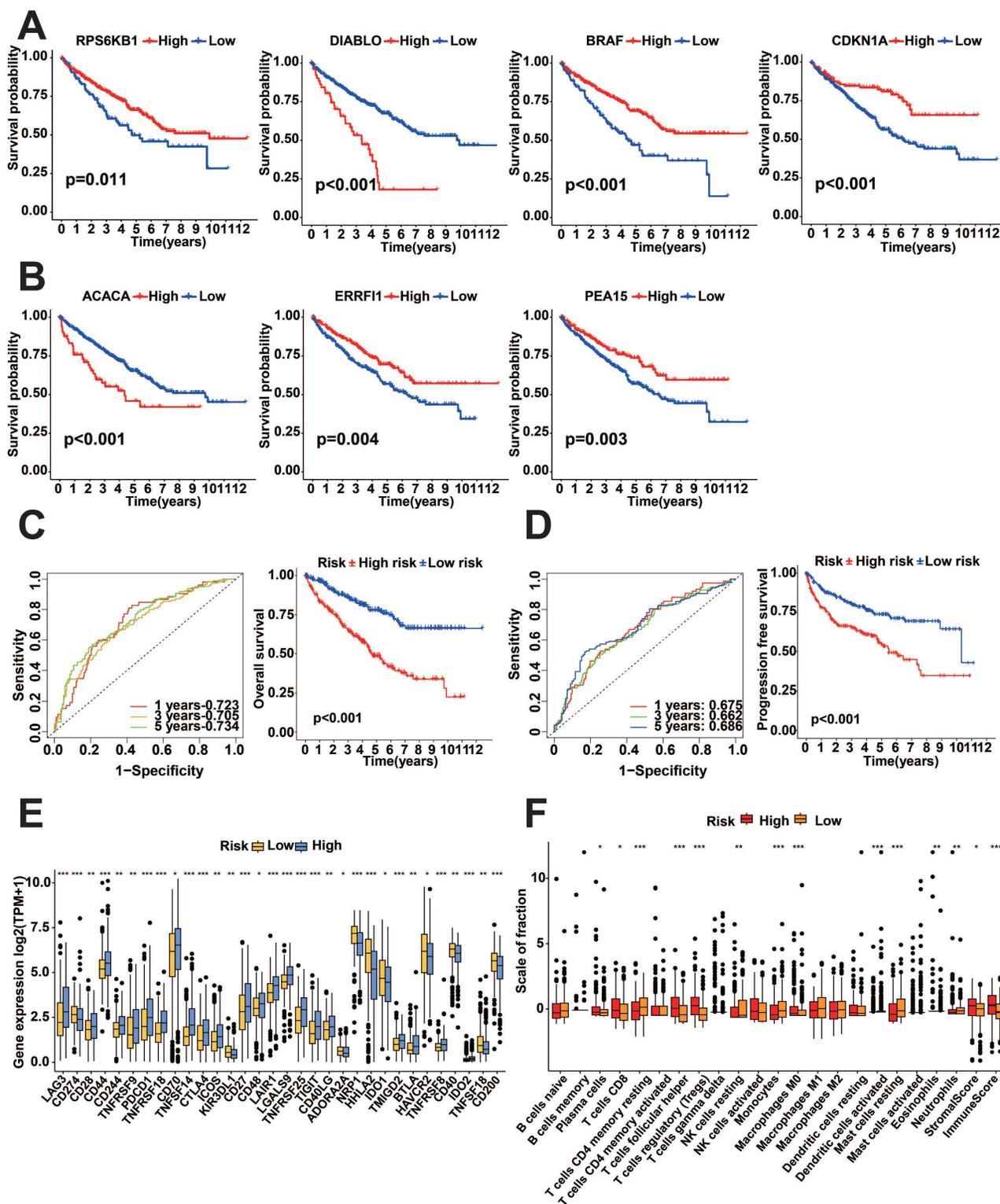


Figure 10. Survival analysis of TCGA-KIRC prognostic protein-encoding genes. A-B: K-M survival curves of prognostic protein-encoding genes for survival analysis. C: K-M survival curves and ROC curves for OS survival analysis for protein-coding genes. D: K-M survival curves and ROC curves for PFI survival analysis for protein-coding genes. E: Differential evaluation of immune checkpoint expression for risk subgroups of protein-coding genes. F: Analysis of immune infiltration in risk subgroups of protein-coding genes. (*p<0.05, **p<0.01, ***p<0.001, ****p<0.0001).

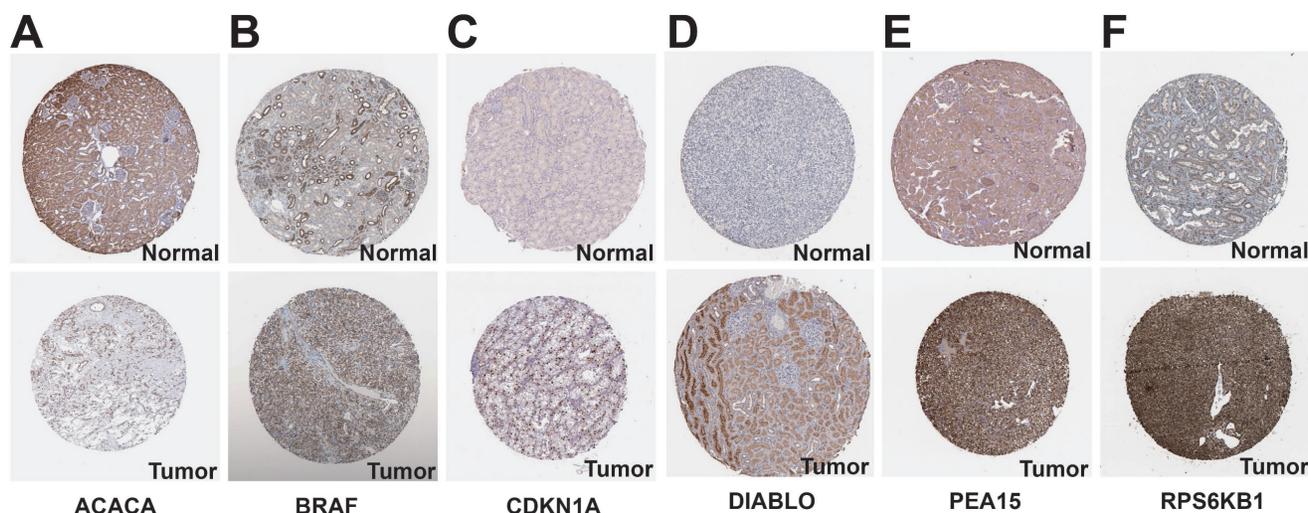


Figure 11. Expression of prognostic protein-encoding genes in tumor tissues and normal tissues in the HPA database. A: ACACA. B: BRAF. C: CDKN1A. D: DIABLO. E: PEA15. F: RPS6KB1.

There are often intrinsic mechanisms underlying the prognostic differences between risk groups[20], and our GSEA results showed significant enrichment of complement and coagulation cascades, cytokine receptor interactions, ECM receptor interactions, hematopoietic cell lineage, and p53 signaling pathway in the high-risk group. Activation of the complement system and coagulation cascade in the tumor microenvironment promotes inflammatory response and tumor progression[21, 22]. Cytokine receptor interactions regulate immune responses and inflammatory processes[23], while ECM receptor interactions play a key role in tumor growth and metastasis[24], and finally, p53 is an important tumor suppressor gene[25], which suggests that dysfunction of p53 may be present in the high-risk group, and that the above results may be associated with a poorer prognosis in the high-risk group. In contrast, the low-risk group was significantly enriched in metabolism-related signaling pathways, suggesting that the metabolic status of patients in the low-risk group may be associated with their better prognosis[26]. We suggest further studies to determine whether these pathways may serve as new therapeutic targets or biomarkers.

Among the prognostic protein-coding genes, ACACA played a crucial role in lipid metabolism as a significant enzyme and as a key enzyme in regulating the initial production of fatty acids [27]. Bioactive lipids could promote or inhibit the development and metastasis of renal cancer by regulating the stability and transcriptional activity of HIF-2 α , thus affecting the proliferation, migration, angiogenesis, immune escape, and other processes of renal cancer cells[28, 29]. It has been shown that ACACA-induced lipid

synthesis and lipid accumulation led to hepatocellular carcinoma progression[30], whereas inhibition of ACACA similarly inhibited hepatocellular carcinoma progression[31]. The research focused on ACACA and kidney cancer has not made significant advancements. However, compounds that inhibit lipid absorption or transportation have demonstrated anti-cancer effects in laboratory and animal studies. Furthermore, a number of these medications have progressed to clinical trials [32]. The p21 protein, encoded by the gene CDKN1A, interacted with a variety of proteins involved in cell cycle inhibition, as well as cell differentiation, and migration, thereby acting as an oncogenic or pro-oncogenic agent[33, 34]. Up-regulation of P21 signaling or p21 mRNA stability was found to inhibit the growth of KIRC [35, 36]. Immunohistochemical analysis of tissue microarrays from 366 KIRC patients showed that p21 might contribute to prognosis and play different biological functions in limited and metastatic renal cell carcinoma[37]. The RPS6KB1 gene encodes ribosomal protein S6 kinase (p70S6K), which is used to control cell survival, proliferation, and metabolism through the PI3K/mTOR signaling pathway[38]. Several studies have shown that p70S6K is associated with the pathophysiology of various tumors such as prostate and colorectal cancers[39-41]. Smac was a pro-apoptotic mitochondrial protein, and phospholipid metabolism and phosphatidylethanolamine synthesis regulated by Smac and phosphatidylethanolamine interactions were critical for cancer cell proliferation[42, 43]. Studies have demonstrated that patients with RCC who exhibited low levels of Smac expression had a fourfold increase in the risk of death compared to patients with high expression[44].

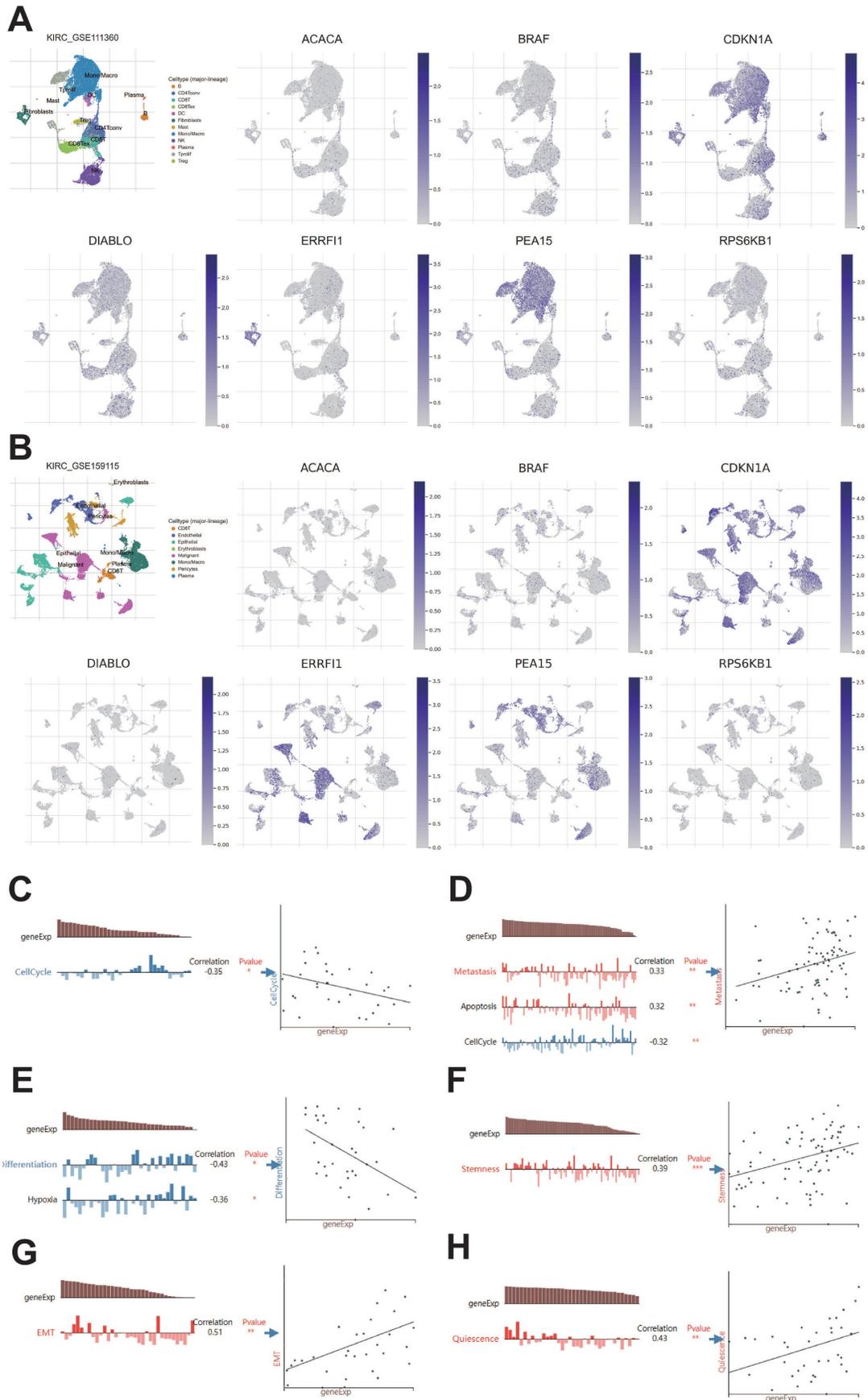


Figure 12. TISCH database and CancerSEA database analysis of protein-coding genes. **A:** Gene expression from the KIRC-GSE11360 dataset in the TISCH database. **B:** Gene expression from the KIRC-GSE159115 dataset in the TISCH database. **C-H:** Correlation of prognostic protein-encoding genes with cancer cell functional status at the single-cell level.

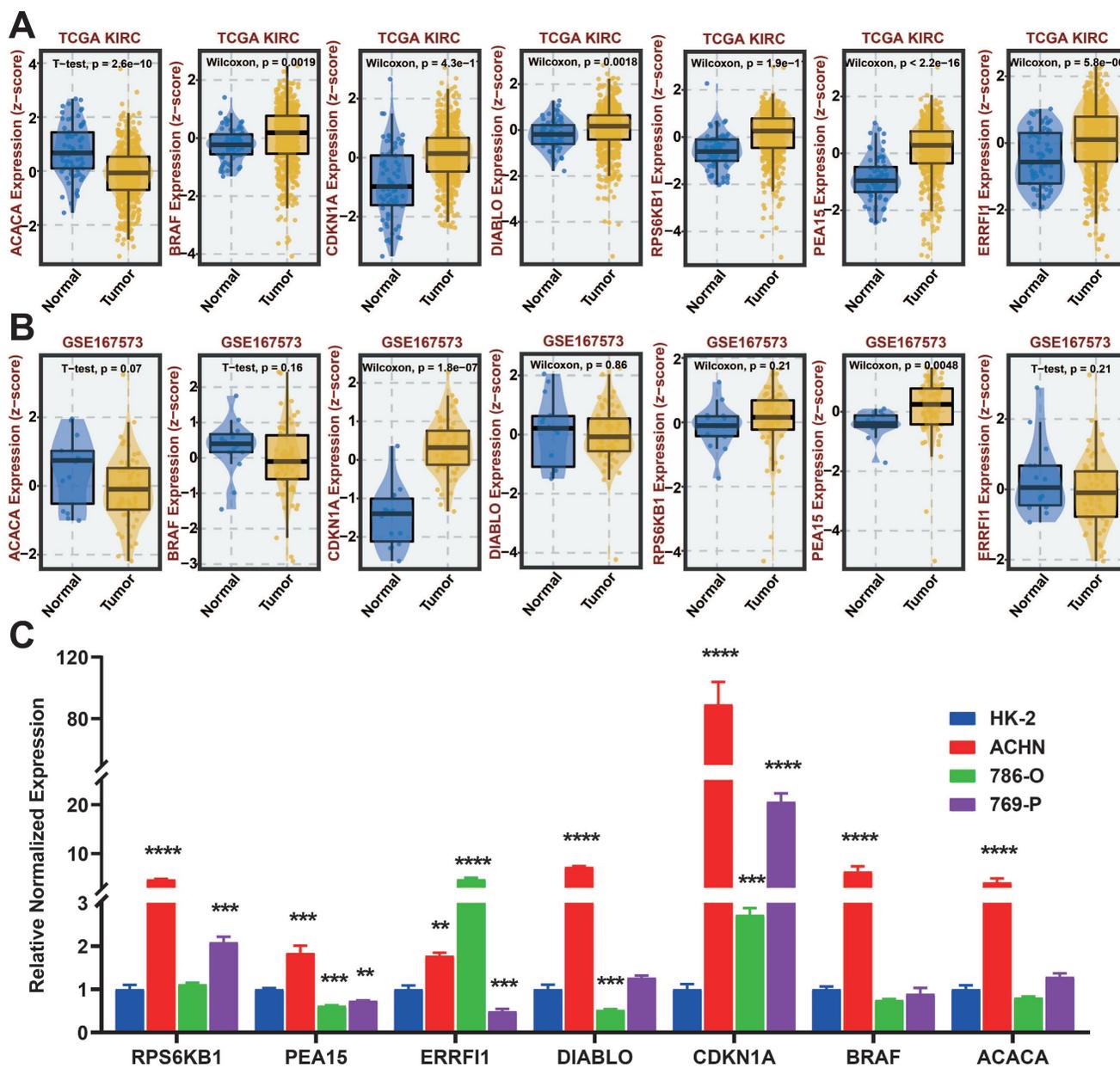


Figure 13. qRT-PCR of seven protein-coding genes. A: Gene expression from the TCGA-KIRC dataset in the BEST website. B: Gene expression from the GSE167573 dataset in the BEST website. C: Expression of seven protein-coding genes in cell lines examined by qRT-PCR.

BRAF functioned as a serine/threonine kinase that played a role in controlling the MAPK signaling pathway. BRAF was mutated in a variety of cancers, leading to aberrant activation of the signaling pathway. Currently, the FDA has approved small molecule inhibitors that target BRAF for the treatment of advanced melanoma and non-small cell lung cancer[45, 46]. ERRF1 acts as a suppressor of EGFR by directly attaching to EGFR, restraining the enzymatic function of EGFR, and facilitating the degradation of EGFR in lysosomes[47]. Research has indicated that ERRF1 is significant in the development of lung cancer, endometrial cancer, and breast cancer[48]. PEA15 was widely expressed in

breast cancer, and studies have shown that PEA15 dephosphorylation might be associated with breast cancer progression and drug resistance[49, 50].

Although our model exhibits excellent predictive performance, there are still certain constraints in our research. Our study was based on the analysis and validation of data collected through public databases, the inclusion of corresponding external validation data may give better results, and prospective studies are needed to further evaluate the model. Moreover, it is crucial to conduct thorough functional experiments to clarify the interrelated mechanisms of the seven prognostic proteins.

Conclusion

In general, we created a proteomics-driven framework that can forecast the outlook, immune surroundings, and medication reactions of patients with KIRC. The model relies on 7 prognostic proteins and clinical factors. The model's potential for reliable prognostic prediction in KIRC was demonstrated through validation against the HPA database and qRT-PCR experiments.

Abbreviations

KIRC: Kidney clear cell carcinoma
TCGA: The Cancer Genome Atlas Program
qRT-PCR: Quantitative Real-time PCR
OS: Overall survival
PFI: Progression Free Interval
RCC: Renal cell carcinoma
PFS: Progression-free survival
GDC: Genomic Data Commons
LASSO: Least absolute shrinkage and selection

operator

ROC: Receiver Operating Characteristic
GSEA: Gene set enrichment analysis
TIDE: Tumor Immune Dysfunction and

Exclusion

HPA: Human Protein Atlas
TISCH: Tumor Immune Single-cell Hub
CancerSEA: Cancer Single-cell State Atlas
FBS: Fetal bovine serum
P/S: Penicillin-streptomycin
K-M: Kaplan-Meier
PCA: Principal component analysis
CNV: Copy number variation
TMB: Tumor mutational load
CDF: Cumulative Distribution Function
TME: Tumor micro-environment
EMT: Epithelial-Mesenchymal Transition.

Supplementary Material

Supplementary figures and table.

<https://www.medsci.org/v21p2215s1.zip>

Acknowledgements

Funding

This work was supported by the National Natural Science Foundation of China (Grant numbers [82000724]).

Data availability

Public data used in this work can be acquired from the TCGA Research Network portal, the HPA database, the TISCH database, CancerSEA database and the TIDE database.

Author contributions

Guangyang Cheng: Writing – review & editing, Writing – original draft, Methodology, Data curation, Software. Zhaokai Zhou: Writing – review & editing, Investigation, Visualization. Shiqi Li: Writing – review & editing, Validation. Zhuo Ye: Writing – review & editing. Yan Wang: Writing – review & editing. Jianguo Wen: Writing – review & editing, Resources. Chuanchuan Ren: Writing – review & editing, Resources, Supervision.

Competing Interests

The authors have declared that no competing interest exists.

References

- Znaor A, Lortet-Tieulent J, Laversanne M, Jemal A, Bray F. International variations and trends in renal cell carcinoma incidence and mortality. *Eur Urol.* 2015; 67: 519-30.
- Dy GW, Gore JL, Forouzanfar MH, Naghavi M, Fitzmaurice C. Global Burden of Urologic Cancers, 1990-2013. *Eur Urol.* 2017; 71: 437-46.
- Motzer RJ, Jonasch E, Agarwal N, Alva A, Baine M, Beckermann K, et al. Kidney Cancer, Version 3.2022, NCCN Clinical Practice Guidelines in Oncology. *J Natl Compr Canc Netw.* 2022; 20: 71-90.
- Linehan WM, Schmidt LS, Crooks DR, Wei D, Srinivasan R, Lang M, et al. The Metabolic Basis of Kidney Cancer. *Cancer Discov.* 2019; 9: 1006-21.
- Escudier B, Eisen T, Stadler WM, Szczylik C, Oudard S, Staehler M, et al. Sorafenib for treatment of renal cell carcinoma: Final efficacy and safety results of the phase III treatment approaches in renal cancer global evaluation trial. *J Clin Oncol.* 2009; 27: 3312-8.
- Motzer RJ, Escudier B, Oudard S, Hutson TE, Porta C, Bracarda S, et al. Efficacy of everolimus in advanced renal cell carcinoma: a double-blind, randomised, placebo-controlled phase III trial. *Lancet.* 2008; 372: 449-56.
- Nusinow DP, Szpyt J, Ghandi M, Rose CM, McDonald ER, 3rd, Kalocsay M, et al. Quantitative Proteomics of the Cancer Cell Line Encyclopedia. *Cell.* 2020; 180: 387-402 e16.
- Liu Y, Beyer A, Aebersold R. On the Dependency of Cellular Protein Levels on mRNA Abundance. *Cell.* 2016; 165: 535-50.
- Tan HT, Lee YH, Chung MCM. Cancer proteomics. *Mass Spectrometry Reviews.* 2012; 31: 583-605.
- Qu Y, Feng J, Wu X, Bai L, Xu W, Zhu L, et al. A proteogenomic analysis of clear cell renal cell carcinoma in a Chinese population. *Nat Commun.* 2022; 13: 2052.
- Clark DJ, Dhanasekaran SM, Petralia F, Pan J, Song X, Hu Y, et al. Integrated Proteogenomic Characterization of Clear Cell Renal Cell Carcinoma. *Cell.* 2019; 179: 964-83 e31.
- Blanco AI, Teh BS, Amato RJ. Role of radiation therapy in the management of renal cell cancer. *Cancers (Basel).* 2011; 3: 4010-23.
- Li Y, Lih TM, Dhanasekaran SM, Mannan R, Chen L, Cieslik M, et al. Histopathologic and proteogenomic heterogeneity reveals features of clear cell renal cell carcinoma aggressiveness. *Cancer Cell.* 2023; 41: 139-63 e17.
- Ghatalia P, Gordetsky J, Kuo F, Dulaimi E, Cai KQ, Devarajan K, et al. Prognostic impact of immune gene expression signature and tumor infiltrating immune cells in localized clear cell renal cell carcinoma. *J Immunother Cancer.* 2019; 7: 139.
- Janzen NK, Kim HL, Figlin RA, Belldegrun AS. Surveillance after radical or partial nephrectomy for localized renal cell carcinoma and management of recurrent disease. *Urol Clin North Am.* 2003; 30: 843-52.
- Jian Y, Yang K, Sun X, Zhao J, Huang K, Aldanakh A, et al. Current Advance of Immune Evasion Mechanisms and Emerging Immunotherapies in Renal Cell Carcinoma. *Front Immunol.* 2021; 12: 639636.
- Barata PC, Rini BI. Treatment of renal cell carcinoma: Current status and future directions. *CA Cancer J Clin.* 2017; 67: 507-24.
- Hsieh JJ, Le VH, Oyama T, Ricketts CJ, Ho TH, Cheng EH. Chromosome 3p Loss-Orchestrated VHL, HIF, and Epigenetic Dereglulation in Clear Cell Renal Cell Carcinoma. *J Clin Oncol.* 2018; 36: JCO2018792549.
- Clark DJ, Dhanasekaran SM, Petralia F, Pan J, Song X, Hu Y, et al. Integrated Proteogenomic Characterization of Clear Cell Renal Cell Carcinoma. *Cell.* 2020; 180: 207.
- Wang J, Ling S, Ni J, Wan Y. Novel gammadelta T cell-based prognostic signature to estimate risk and aid therapy in hepatocellular carcinoma. *BMC Cancer.* 2022; 22: 638.
- Wang MJ, Sun Y, Song Y, Ma JN, Wang ZQ, Ding XQ, et al. Mechanism and Molecular Targets of Ejiao Siwu Decoction for Treating Primary Immune Thrombocytopenia Based on High-Performance Liquid Chromatograph,

- Network Pharmacology, Molecular Docking and Cytokines Validation. *Front Med (Lausanne)*. 2022; 9: 891230.
22. Wang Y, Miao X, Jiang Y, Wu Z, Zhu X, Liu H, et al. The synergistic antitumor effect of IL-6 neutralization with NVP-BEZ235 in hepatocellular carcinoma. *Cell Death Dis*. 2022; 13: 146.
 23. Xie B, Li K, Zhang H, Lai G, Li D, Zhong X. Identification and validation of an immune-related gene pairs signature for three urologic cancers. *Aging (Albany NY)*. 2022; 14: 1429-47.
 24. Kakkassery V, Gemoll T, Kraemer MM, Sauer T, Tura A, Ranjbar M, et al. Protein Profiling of WERI-RB1 and Etoposide-Resistant WERI-ETOR Reveals New Insights into Topoisomerase Inhibitor Resistance in Retinoblastoma. *Int J Mol Sci*. 2022; 23: 4058.
 25. Richardson JS, Sethi G, Lee GS, Malek SN. Chalepin: isolated from *Ruta angustifolia* L. Pers induces mitochondrial mediated apoptosis in lung carcinoma cells. *BMC Complement Altern Med*. 2016; 16: 389.
 26. Liu T, Chen L, Gao G, Liang X, Peng J, Zheng M, et al. Development of a Gene Risk Signature for Patients of Pancreatic Cancer. *J Healthc Eng*. 2022; 2022: 4136825.
 27. Hunkeler M, Hagmann A, Stutfeld E, Chami M, Guri Y, Stahlberg H, et al. Structural basis for regulation of human acetyl-CoA carboxylase. *Nature*. 2018; 558: 470-4.
 28. Stepanovska Tanturovska B, Manaila R, Fabbro D, Huwiler A. Lipids as Targets for Renal Cell Carcinoma Therapy. *Int J Mol Sci*. 2023; 24: 3272.
 29. Heravi G, Yazdanpanah O, Podgorski I, Matherly LH, Liu W. Lipid metabolism reprogramming in renal cell carcinoma. *Cancer Metastasis Rev*. 2022; 41: 17-31.
 30. Peng JY, Cai DK, Zeng RL, Zhang CY, Li GC, Chen SF, et al. Upregulation of Superenhancer-Driven lncRNA FASRL by USF1 Promotes De Novo Fatty Acid Biosynthesis to Exacerbate Hepatocellular Carcinoma. *Adv Sci (Weinh)*. 2022; 10: e2204711.
 31. Lally JSV, Ghoshal S, DePeralta DK, Moaven O, Wei L, Masia R, et al. Inhibition of Acetyl-CoA Carboxylase by Phosphorylation or the Inhibitor ND-654 Suppresses Lipogenesis and Hepatocellular Carcinoma. *Cell Metab*. 2019; 29: 174-82 e5.
 32. Jeong DW, Lee S, Chun YS. How cancer cells remodel lipid metabolism: strategies targeting transcription factors. *Lipids Health Dis*. 2021; 20: 163.
 33. Xiong Y, Hannon GJ, Zhang H, Casso D, Kobayashi R, Beach D. p21 is a universal inhibitor of cyclin kinases. *Nature*. 1993; 366: 701-4.
 34. Kreis NN, Louwen F, Yuan J. The Multifaceted p21 (Cip1/Waf1/CDKN1A) in Cell Differentiation, Migration and Cancer Therapy. *Cancers (Basel)*. 2019; 11: 1220.
 35. Zhu L, Wang J, Kong W, Huang J, Dong B, Huang Y, et al. LSD1 inhibition suppresses the growth of clear cell renal cell carcinoma via upregulating P21 signaling. *Acta Pharm Sin B*. 2019; 9: 324-34.
 36. Di J, Wang H, Zhao Z, Zhao G, Qin X, Han Z, et al. CPEB4 Inhibit Cell Proliferation via Upregulating p21 mRNA Stability in Renal Cell Carcinoma. *Front Cell Dev Biol*. 2021; 9: 687253.
 37. Weiss RH, Borowsky AD, Seligson D, Lin PY, Dillard-Telm L, Beldegrun AS, et al. p21 is a prognostic marker for renal cell carcinoma: implications for novel therapeutic approaches. *J Urol*. 2007; 177: 63-8; discussion 8-9.
 38. Bahrami BF, Ataie-Kachoei P, Pourgholami MH, Morris DL. p70 Ribosomal protein S6 kinase (Rps6kb1): an update. *J Clin Pathol*. 2014; 67: 1019-25.
 39. An HJ, Park M, Kim J, Han YH. miR-5191 functions as a tumor suppressor by targeting RPS6KB1 in colorectal cancer. *Int J Oncol*. 2019.
 40. Cai C, Chen QB, Han ZD, Zhang YQ, He HC, Chen JH, et al. miR-195 Inhibits Tumor Progression by Targeting RPS6KB1 in Human Prostate Cancer. *Clin Cancer Res*. 2015; 21: 4922-34.
 41. Wang-Bishop L, Chen Z, Goma A, Lockhart AC, Salaria S, Wang J, et al. Inhibition of AURKA Reduces Proliferation and Survival of Gastrointestinal Cancer Cells With Activated KRAS by Preventing Activation of RPS6KB1. *Gastroenterology*. 2019; 156: 662-75 e7.
 42. Pandey SK, Paul A, Shteinfein-Kuzmine A, Zalk R, Bunz U, Shoshan-Barmatz V. SMAC/Diablo controls proliferation of cancer cells by regulating phosphatidylethanolamine synthesis. *Mol Oncol*. 2021; 15: 3037-61.
 43. Arbiser JL. Diablo: A Double-Edged Sword in Cancer? *Mol Ther*. 2018; 26: 678-9.
 44. Martinez-Ruiz G, Maldonado V, Ceballos-Cancino G, Grajeda JP, Melendez-Zajgla J. Role of Smac/DIABLO in cancer progression. *J Exp Clin Cancer Res*. 2008; 27: 48.
 45. Zaman A, Wu W, Bivona TG. Targeting Oncogenic BRAF: Past, Present, and Future. *Cancers (Basel)*. 2019; 11.
 46. El-Nassan HB. Recent progress in the identification of BRAF inhibitors as anti-cancer agents. *Eur J Med Chem*. 2014; 72: 170-205.
 47. Cui M, Liu D, Xiong W, Wang Y, Mi J. ERRF1 induces apoptosis of hepatocellular carcinoma cells in response to tryptophan deficiency. *Cell Death Discov*. 2021; 7: 274.
 48. Xu D, Li C. Gene 33/Mig6/ERRF1, an Adapter Protein with Complex Functions in Cell Biology and Human Diseases. *Cells*. 2021; 10: 1574.
 49. Li L, Wang N, Xiong Y, Guo G, Zhu M, Gu Y. Transcription Factor FOSL1 Enhances Drug Resistance of Breast Cancer through DUSP7-Mediated Dephosphorylation of PEA15. *Mol Cancer Res*. 2022; 20: 515-26.
 50. Park J, Tacam MJ, Chauhan G, Cohen EN, Gagliardi M, Iles LR, et al. Nonphosphorylatable PEA15 mutant inhibits epithelial-mesenchymal transition in triple-negative breast cancer partly through the regulation of IL-8 expression. *Breast Cancer Res Treat*. 2021; 189: 333-45.