Received: 11 August 2021 | Revised: 6 September 2021 | Accepted: 12 September 2021

DOI: 10.1002/jcla.24022

RESEARCH ARTICLE

WILEY

Identification of potential biomarkers associated with immune infiltration in papillary renal cell carcinoma

Ran Deng[1,2,3] | Jianpeng Li[1,2,3] | Hong Zhao[1,2,3] | Zhirui Zou[1,2,3] | Jiangwei Man[1,2,3] | Jinlong Cao[1,2,3] | Li Yang[1,2,3]

[1]Department of Urology, Lanzhou University Second Hospital, Lanzhou, China

[2]Key Laboratory of Gansu Province for Urological Diseases, Lanzhou, China

[3]Clinical Center of Gansu Province for Nephron-urology, Lanzhou, China

Correspondence
Li Yang, Department of Urology, Lanzhou University Second Hospital, Lanzhou, China.
Email: ery_yangli@lzu.edu.cn

Funding information
This study was supported by the (1) regulatory mechanism of AMPK in ischemic-reperfusion injury and fibrosis in renal transplantation (project id:CY2015-YJRC08); and (2) Gansu Provincial Education Department outstanding graduate "innovation star" project (project id:2021CXZX-154)

Abstract

Background: Immunotherapeutic approaches have recently emerged as effective treatment regimens against various types of cancer. However, the immune-mediated mechanisms surrounding papillary renal cell carcinoma (pRCC) remain unclear. This study aimed to investigate the tumor microenvironment (TME) and identify the potential immune-related biomarkers for pRCC.

Methods: The CIBERSORT algorithm was used to calculate the abundance ratio of immune cells in each pRCC samples. Univariate Cox analysis was used to select the prognostic-related tumor-infiltrating immune cells (TIICs). Multivariate Cox regression analysis was performed to develop a signature based on the selected prognostic-related TIICs. Then, these pRCC samples were divided into low- and high-risk groups according to the obtained signature. Analyses using Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), and Gene Set Enrichment Analysis (GSEA) were performed to investigate the biological function of the DEGs (differentially expressed genes) between the high- and low-risk groups. The hub genes were identified using a weighted gene co-expression network analysis (WGCNA) and a protein-protein interaction (PPI) analysis. The hub genes were subsequently validated by multiple clinical traits and databases.

Results: According to our analyses, nine immune cells play a vital role in the TME of pRCC. Our analyses also obtained nine potential immune-related biomarkers for pRCC, including *TOP2A*, *BUB1B*, *BUB1*, *TPX2*, *PBK*, *CEP55*, *ASPM*, *RRM2*, and *CENPF*.

Conclusion: In this study, our data revealed the crucial TIICs and potential immune-related biomarkers for pRCC and provided compelling insights into the pathogenesis and potential therapeutic targets for pRCC.

KEYWORDS
biomarkers, hub gene, papillary renal cell carcinoma, tumor microenvironment, tumor-infiltrating immune cells

Ran Deng and Jianpeng Li are contributed equally to this study.

1 | INTRODUCTION

Renal cell carcinoma (RCC) is a common urologic malignancy and has increased incidence in recent years.[1] Aside from clear cell renal cell carcinoma (ccRCC), papillary renal cell carcinoma (pRCC) is the most common subtype of RCC and accounts for 10%–15% of all RCC cases.[2] The histology of pRCC is characterized by a papillary fibrous vascular core of tumor cells and is markedly different from that of ccRCC.[3] There are two main histological subtypes of pRCC, depending on their characteristics. Compared with type 1 tumors, type 2 tumors have a worse prognosis. There was no significant difference in metastatic disease outcomes between pRCC and ccRCC[3]; however, the immune cell infiltration in the tumor microenvironment (TME) of pRCC is still unclear, and the elucidation of the different effects of immunotherapy is still in its infancy. Therefore, it is necessary to more comprehensively explore the pathogenesis of pRCC and its potential immune-related targets.

Due to its excellent efficacy, tumor immunotherapy has gradually become the fourth most commonly administered type of tumor treatment after surgery, radiotherapy, and chemotherapy.[4] Effective cancer immunotherapy requires overcoming the immunosuppressive TME.[5] Immune cells constitute a large fraction of the TME, and they function by modulating tumor progression and regulating antitumor immune responses.[6] Therefore, the elucidation of the mechanisms of immune cell infiltration in pRCC is crucial to understand the dynamics of its TME and to identify potential immunotherapeutic targets.

Recently, numerous studies have suggested that the TME components are associated with patient survival outcomes.[7,8] Specifically, immune cell composition in the TME is associated with survival in patients with RCC.[8] McDermott et al.[9] demonstrated that CD8+ T-cell responses had been associated with survival in patients with RCC treated with anti-PD-L1 antibodies. However, most studies on RCC are limited to ccRCC, so the specific mechanism of the immune microenvironment in pRCC remains unclear. In this study, we aim to explore the immune-related cellular and genetic targets for pRCC therapy and provide a new avenue for further research on immunotherapy.

2 | MATERIALS AND METHODS

2.1 | Data sources and technology roadmap

Clinical information and the gene expression information from RNA-seq data of patients with pRCC were downloaded from the UCSC Xena database (https://xenabrowser.net/datapages/), which includes TCGA data that underwent standardized processing.

The data obtained were normalized count data, including information from 291 tumor tissues and 32 adjacent normal tissues. The GEO database (https://www.ncbi.nlm.nih.gov/gds/) is a shared database for storing chips, second-generation sequencing, and other high-throughput sequencing data. The GSE26574 dataset (34 tumor samples and 15 normal samples) and the GSE7203 dataset (35 tumor samples and 12 normal samples) were downloaded from the GEO database. The technology roadmap is shown in Figure 1.

2.2 | Screening prognostic-related immune cells

CIBERSORT is an approach that primarily characterizes the immune cell composition of diverse tissue according to their gene expression profiles.[10] The CIBERSORT R source code and the LM22 signature matrix file, both of which obtain information from a total of 22 immune cell types, were downloaded from a website (https://cibersort.stanford.edu/). The R CIBERSORT algorithm was used to determine the abundance ratio of 22 Tumor-infiltrating immune cells (TIICs) (including 22 immune cells) in the 291 tumor samples obtained from the UCSC Xena database. The "barplot" and "pheatmap"[11] functions of R software were used to visualize the abundance ratio matrix from the data obtained in the previous algorithm. The Pearson correlation coefficients among these TIICs were then calculated using the "cor" function of the R package and was visualized using the "corrplot" package,[12] also in R. Finally, univariate Cox analysis was performed to screen the prognostic-related TIICs, with the threshold set at $p < 0.05$. The "forestplot" R package[13] was used to visualize the prognostic-related immune cells.

2.3 | Identifying the relationship between prognostic-related TIICs and clinical traits

A Wilcoxon test with a p-value <0.05 was used to analyze the relationship between prognostic-related TIICs and clinical traits. The "boxplot" function of R software was used to visualize the data obtained from this analysis. Kaplan–Meier (KM) survival analysis was then implemented using the R package "survival"[14] to show the relationship between the prognostic-related TIICs and OS (overall survival) in patients with pRCC.

2.4 | Defining the high- and low-risk groups based on the prognostic-related TIICs

The risk score for each pRCC sample was calculated using the regression coefficients derived from the multivariate Cox regression analysis for the prognostic-related TIICs. These samples were separated into high- and low-risk groups using the median risk score as a cutoff. A KM analysis via the log-rank test was performed to assess the survival differences between the high-risk and low-risk groups using the "survival"[14] package in R. A nomogram prediction model was then constructed for these TIICs using the R package "rms".[15] A time-dependent ROC curve and a corresponding calibration curve were derived to assess the predictive significance of the model by using the "rms"[15] and "timeROC"[16] packages in R.

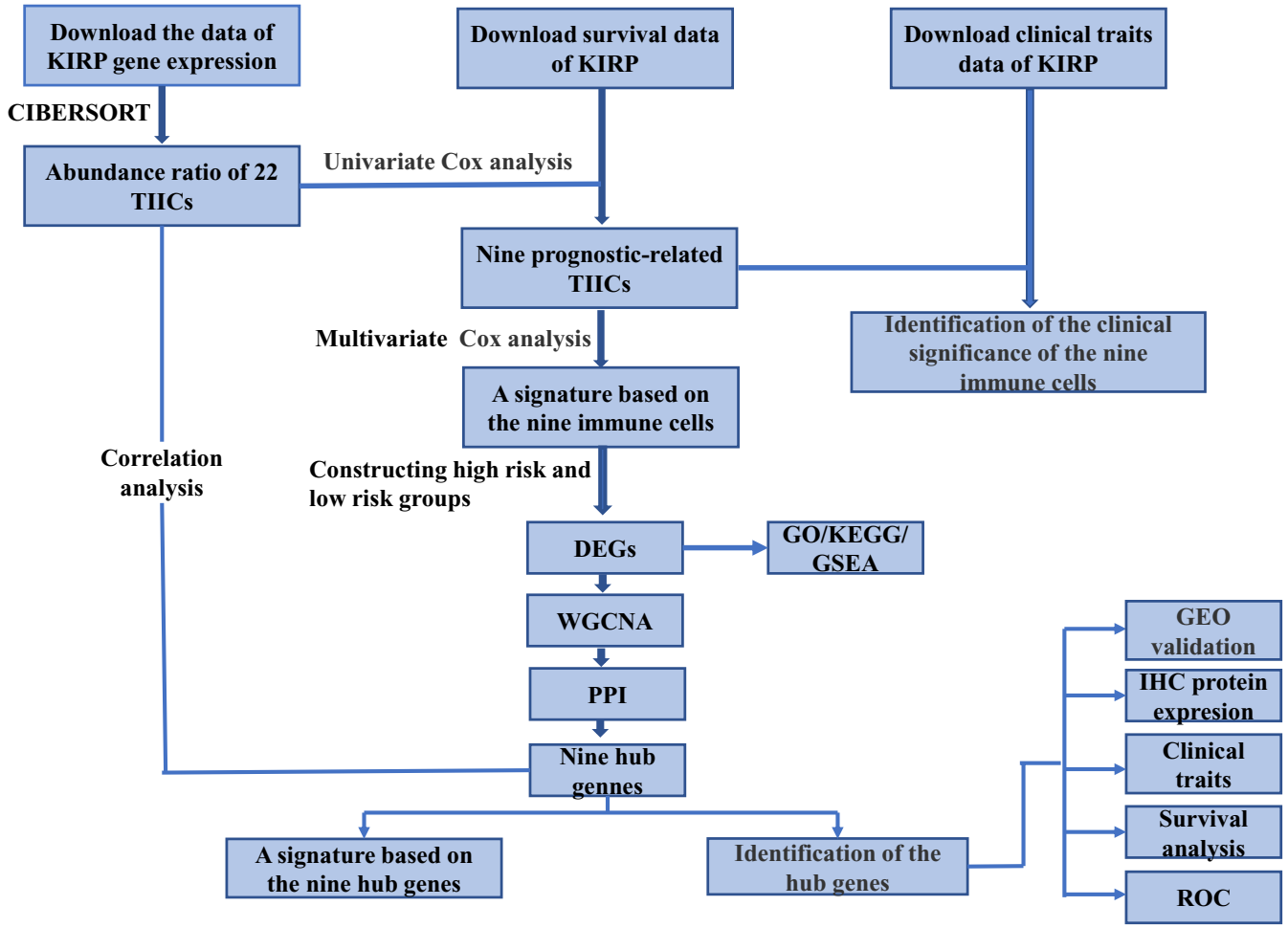

FIGURE 1 Workflow of the selection process for the eligible studies in the analysis. KIRP: kidney renal papillary cell carcinoma. CIBERSORT: an algorithm for analysis of immune cell composition of complex tissue from their gene expression. Abbreviations: DEGs, differentially expressed genes; GEO, Gene Expression Omnibus; GSEA, Gene Set Enrichment Analysis; IHC, immunohistochemical; PPI, protein-protein interaction; ROC, receiver operating characteristic; TCGA, The Cancer Genome Atlas; TIICs, tumor-infiltrating immune cells; WGCNA, weighted gene co-expression network analysis

2.5 | Identification and functional annotation of the differentially expressed genes (DEGs) between the high- and low-risk groups

Differential expression analysis between the high-risk and low-risk groups was conducted using the "DESeq2"[17] package in R, with *p*-values <0.05 and |fold change| >1 as the filters to determine whether these genes were significantly differentially expressed. Gene Ontology (GO, annotating gene products) and Kyoto Encyclopedia of Genes and Genomes (KEGG, introducing many metabolic pathways and the relationship between them) pathway analyses were performed using the "clusterProfiler"[18] R package and "org. Hs.eg. db" to reveal the potential functions of these DEGs, with a *p*-value <0.05 set as the cutoff. Gene Set Enrichment Analysis (GSEA) was then performed, also using the "clusterProfiler"[18] R package and immunologic signature gene set (C7 gene sets) obtained from the Molecular Signatures Database (MSigDB, https://www.gsea-msigdb.org/).

2.6 | Construction of a weighted gene co-expression network

A weighted gene co-expression network analysis (WGCNA) is a tool typically used for constructing gene co-expression networks. It aims to explore the correlation between gene networks and clinical phenotypes and identify the core genes in the network. We employed the R package "WGCNA" using a soft-thresholding power of 4 and a minimum module size of 30[19] to construct a co-expression network for the DEGs obtained. We then calculated the correlation between the modules obtained and clinical traits to identify the core modules.

2.7 | Constructing a protein-protein interaction (PPI) network and determining the hub genes

The STRING database (https://string-db.org/) searches for the interaction between known protein and predicted protein and

was used to construct the PPI network of the DEGs in the core modules, those with an interaction score >0.4. The five topological analysis methods in the "cytoHubba" app of Cytoscape v3.8.2, including Closeness, Maximal Clique Centrality (MCC), Maximum Neighborhood Component (MNC), Degree, and Edge Percolated Component (EPC), were utilized to screen the top 30 genes in the PPI network constructed. The hub genes were then identified by intersecting the top 30 genes obtained after running the various algorithms in the topological analyses mentioned above.

2.8 | Verification of the clinical significance of the hub genes

To further validate the clinical significance of the obtained hub genes, we analyzed the relationship between their expression profile and the pathological stage of the sample using the "ggpubr"[20] R package. A KM survival analysis and the log-rank test using the R package "survival"[14] were employed to analyze the relationship between the hub genes and OS. An ROC curve was obtained to evaluate the diagnostic value of the hub genes using the "pROC"[21] package in R. Moreover, a nomogram prediction model for the hub genes was constructed using the "rms"[15] package in R. A time-dependent ROC and calibration curve were also derived to assess the predictive significance of the model using the "rms"[15] and "timeROC" packages[16] in R. Finally, to confirm these results, the protein levels of these hub genes were verified using immunohistochemistry (IHC) data obtained from the Human Protein Atlas (HPA) database (https://www.proteinatlas.org/), which contains proteomics and transcriptomics data. The GSE26574 dataset and GSE7203 dataset served as the external validation dataset for these hub genes.

2.9 | Validation of the correlation between the hub genes and infiltrating immune cells

Spearman correlations between the infiltration of all 22 types of immune cells and the expression of hub genes were calculated via the "psych"[22] R package, and the results were visualized using the "pheatmap"[11] package in R. Meanwhile, the XCELL algorithm in Tumor Immune Estimation Resource 2.0 (TIMER2.0)[23] was also used to perform comprehensive correlation analyses between the immune cell signatures and the obtained hub genes.

2.10 | Statistical analyses

All statistical analyses were performed using R software (version 4.1.1). Spearman correlation was performed for correlation analysis. The Wilcoxon test was used to determine the relationship between immune cells, potential markers, and clinical traits. $p < 0.05$ was considered statistically significant.

3 | RESULTS

3.1 | Identifying the prognosis-related immune cells

The infiltration levels of 22 immune cells in the 291 pRCC samples obtained from patients are shown in Figure 2A and 2B. Macrophages M2, resting T-cell CD4 memory, and resting mast cells were of higher abundance than other immune cells. The correlation among these immune cells is shown in Figure 2C. Additionally, we screened the nine immune cells associated with OS via univariate Cox analysis, and the results were shown in Figure 2D. The immune cells associated with OS were follicular helper T cells, macrophages M1, activated dendritic cells, regulatory T cells (Tregs), B-cell memory, CD8 T cells, macrophages M2, naïve B cells, and CD4 memory-activated T cells.

3.2 | Clinical significance of prognostic-related immune cells

We also analyzed the association of the nine prognostic-related immune cells with various clinical traits, including TNM stage, tumor type, pathological stage, and OS. The results of this analysis are shown in Figure 3. The degree of infiltration of macrophages M1, macrophages M2, CD8T cells, and regulatory T cells (Tregs) differed between type 1 and type 2 pRCC (Figure 3A-3D). Macrophage M1, macrophages M2, and regulatory T cells (Tregs) were closely related to the pathological stage (Figure 3E-3G). Macrophage M1, macrophages M2, and regulatory T cells (Tregs) were associated with the TNM stage (Figure 3H-3L). Naïve B cells, macrophages M1, and CD4 memory-activated T cells were related to overall survival (Figure 3M-3O).

3.3 | Classification of patients into low- and high-risk groups and construction of a nomogram prediction model

To further investigate immune cell infiltration in pRCC, we constructed a risk grouping and prediction model based on the prognostic-related TIICs obtained. The risk score of each patient was calculated using the regression coefficients of the multivariate Cox regression of the nine prognosis-related TIICs. Furthermore, the median risk score cutoff for the patients was divided into a low-risk group and a high-risk group (Figure 4A). The analysis showed that more patients died in the high-risk group than in the low-risk group (Figure 4B). The degree of infiltration of the nine prognosis-related TIICs is shown in Figure 4C. The KM survival curves demonstrated that the OS of the high-risk group was shorter than that of the low-risk group ($p = 0.0028$), indicating that the grouping based on the risk score was reasonable (Figure 4D). A nomogram was then constructed based on a multivariate Cox regression analysis of the prognosis-related TIICs (Figure 4E). The calibration curve (Figure 4F) and the time-dependent ROC analysis (Figure 4G) of the nomogram were used to estimate the accuracy of the actual

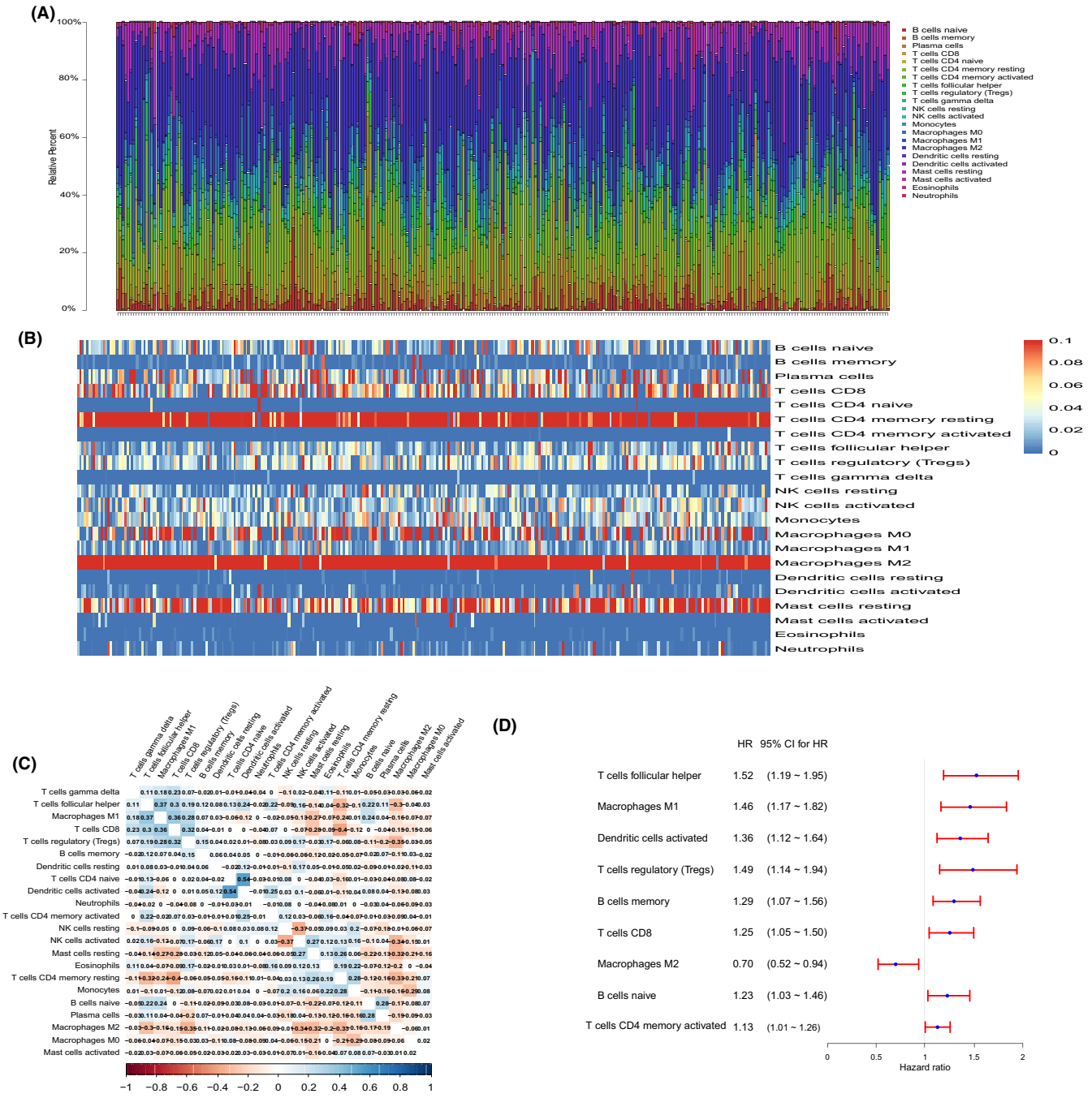

FIGURE 2 Identification of nine prognostic immune cells. (A) The landscape of tumor-infiltrating immune cells. (B) Heatmap of the fraction of tumor-infiltrating immune cells. (C) The correlation coefficient between the abundance ratios of the distinct immune cells. (D) The nine prognostic-related immune cells

observed rates with the predicted survival probability. Our results show that the nomogram has a good prognostic value, suggesting that the nine TIICs are closely associated with OS in pRCC.

3.4 | Identification and functional annotation of the DEGs

In this study, 1157 DEGs (1097 upregulated and 60 downregulated genes) were identified between the high-risk and low-risk groups. A volcano plot was obtained, and the results of the differential analysis are shown in Figure 5A. GO and KEGG enrichment analysis was then performed to investigate the biological functions associated with the obtained DEGs. The significantly enriched KEGG pathways are shown in Figure 5B (Supplementary Table). The pathways related to cytokine-cytokine receptor interactions and the cAMP signaling pathway, which play a key role in pRCC, were significantly enriched. The significant GO enrichment terms of the upregulated and downregulated DEGs are shown in Figure 5D and 5E, respectively. Enriched biological functions related to immunity include

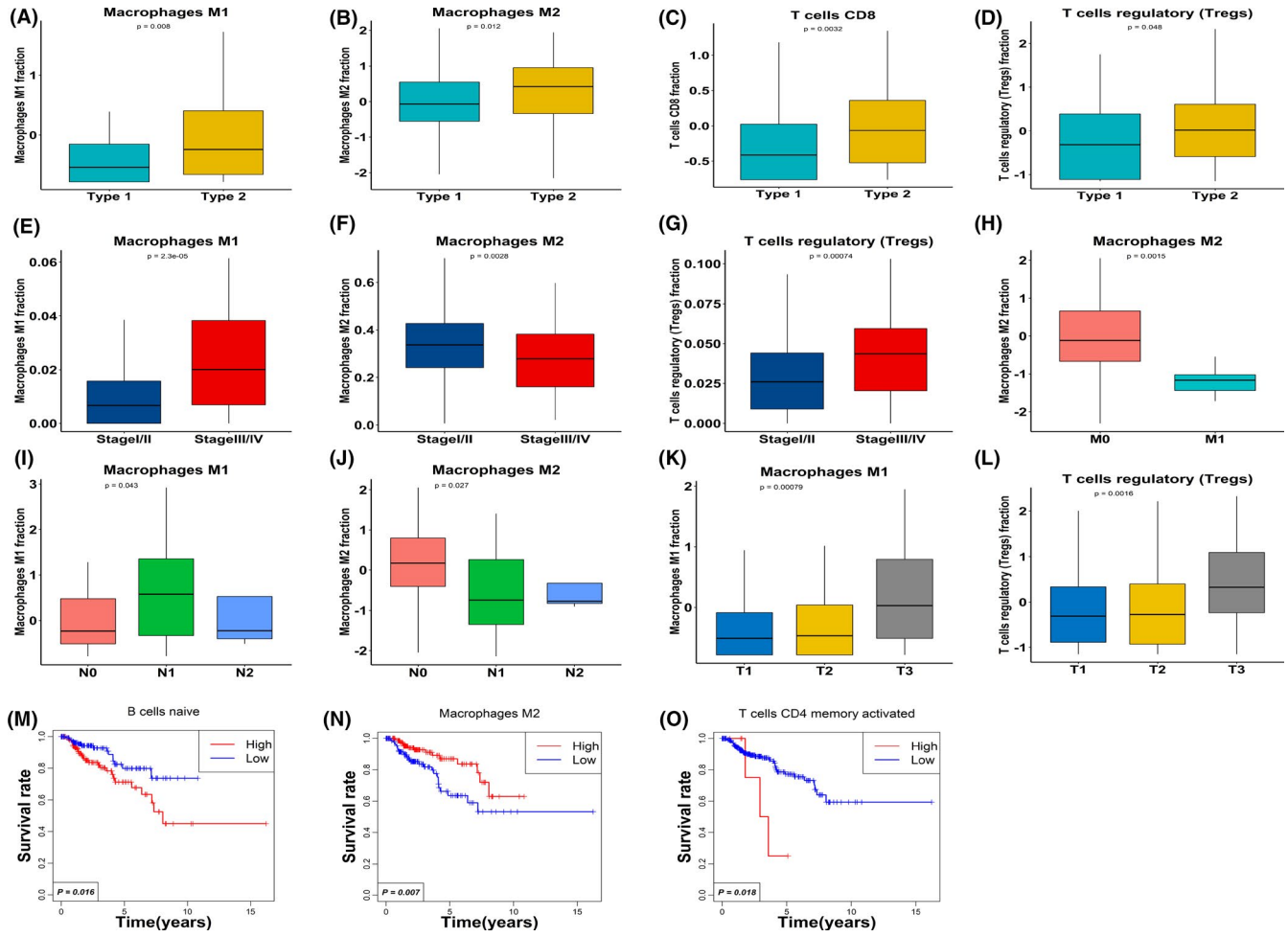

FIGURE 3 Relationship between the nine prognostic-related immune cells and clinical traits. (A)–(D) The relationship between immune cells and tumor type. (E)–(G) The relationship between immune cells and pathological stage. (H)–(L) The relationship between immune cells and TNM stage. (M)–(O) The relationship between immune cells and overall survival

the humoral immune response, IgG receptor activity, and T-cell chemotaxis.

Furthermore, GSEA was used to clarify the relationship between these DEGs and immunity. The results of the GSEA are shown in Figure 5E (Supplementary Table) and showed that regulatory T cells and CD8 T cells, which play a crucial role in cancer-related immunity, were significantly enriched.

3.5 | Identification of the core modules via WGCNA analysis

We also performed a complete WGCNA analysis on the 1157 DEGs, including screening the soft thresholds, constructing a dynamic tree cut, and plotting a network heatmap (Figure 6A-6C). The results showed that the 1157 DEGs could be divided into six modules (blue, brown, green, gray, turquoise, and yellow). There were 496 DEGs in the turquoise module, 313 in the blue module, 167 in the brown module, 135 in the yellow module, 95 in the green module, and 40 in the gray module (Figure 6B). Subsequently, we investigated the relationship between the six modules and clinical traits (pathologic TNM, pathologic stage, and tumor type) and found that the turquoise and blue modules were strongly associated with pathological stage and T stage (Figure 6D). Both gene significance and module membership were plotted for the blue module (Figure 6E-6F), and the results indicated that this module was significantly related to pathological stage (cor =0.91, p = 6e-121) and T stage (cor =0.89, p = 4.3e-108). Likewise, the gene significance of the turquoise module was also related to pathological stage (cor =0.83, p = 2.4e-127) and T stage (cor =0.81, p = 1.4e-116) (Figure 6G-6H). Hence, the blue and turquoise modules were used as core modules for further analyses.

3.6 | Identification of hub genes via multiple algorithms

The 496 genes in the turquoise module and the 313 genes in the blue module were analyzed using the STRING database, and a PPI network containing 706 nodes and 4261 edges was obtained. The top 30 genes in the PPI network were screened using DMNC, MCC, MNC, Degree,

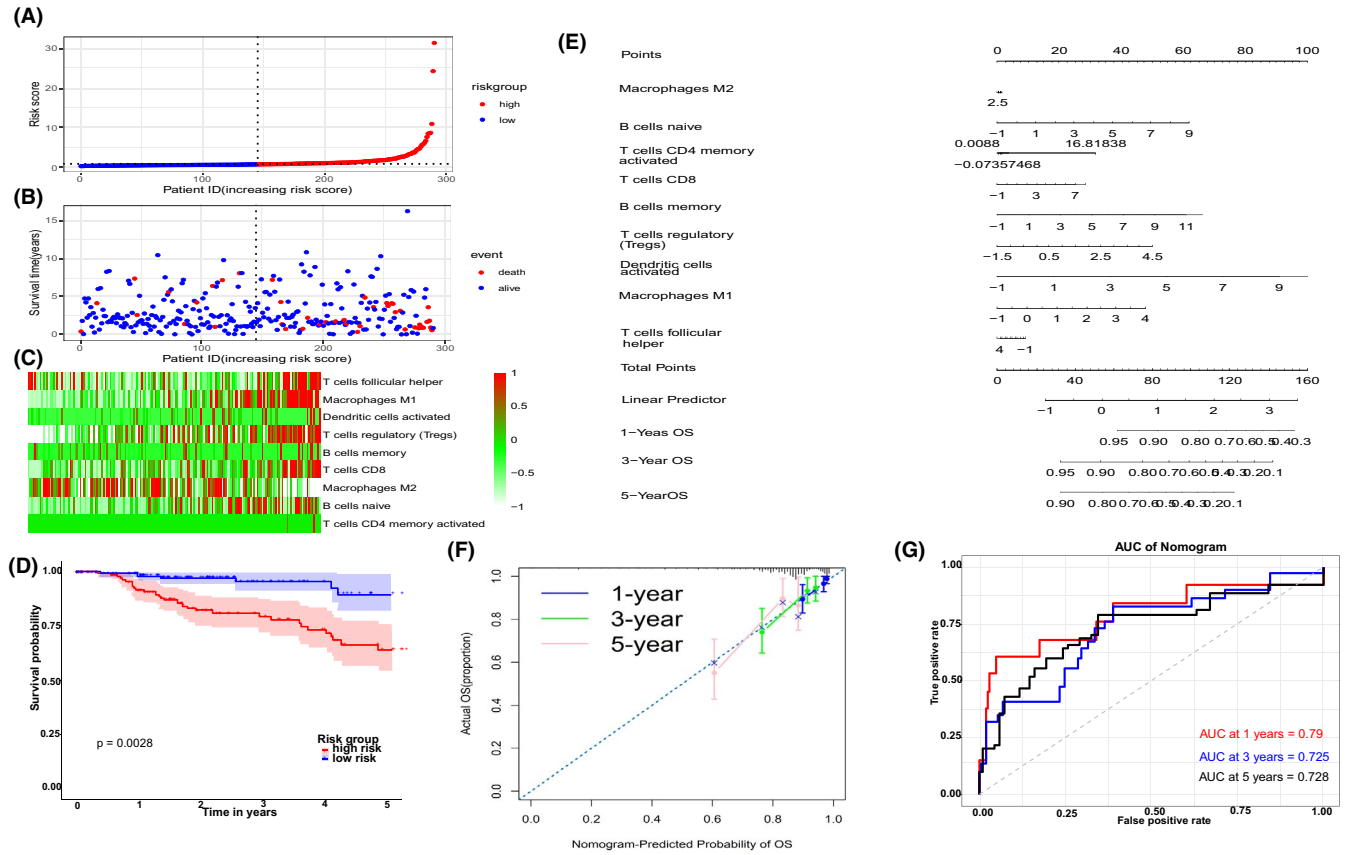

FIGURE 4 Constructing high- and low-risk groups and nomogram based on nine immune cells. (A) The curve of risk score. (B) Survival status of the patients. More dead patients in the high group. (C) Heatmap of the expression profiles of the nine prognostic immune cells in low- and high-risk groups. (D) Kaplan-Meier survival analysis of the low- and high-risk groups. (E) Nomogram integrated nine survival-related cells. (F) The calibration plot of the nomogram for agreement test between 1-, 3-, and 5-year OS prediction and actual outcome in TCGA dataset. (G) The time-dependent ROC curves of the nomogram. TCGA: The Cancer Genome Atlas; ROC: receiver operating characteristic

and EPC (Figure 7A-7E). Based on the five algorithms, the PPI network contained nine common hub genes (*TOP2A*, *BUB1B*, *BUB1*, *TPX2*, *PBK*, *CEP55*, *ASPM*, *RRM2*, and *CENPF*) (Figure 7F). The detailed information regarding these nine hub genes is shown in Table 1.

3.7 | Verification of the clinical significance of the hub genes

The nine hub genes obtained were overexpressed in tumor tissues compared with normal tissues (Figure 8A). The expression profile of these hub genes varied across different pathological stages and tended to increase with an increasing pathological stage (Figure 8B). In addition, we analyzed the relationship between the nine hub genes and the OS and found that the group with a lower expression of these genes had a better prognosis than the group with a high expression of these genes (Figure 8C). ROC curves were obtained to evaluate the diagnostic value of these genes in pRCC tumors and different tumor types (Figure 8D-8E). Results of the analysis indicate that all these genes had a certain accuracy in differentiating between type 1 from type 2 pRCC, except *TPOX2* (area under the curve [AUC] **<0**.5) (Figure 8D). The AUCs of all the hub genes were greater than 0.8, suggesting that they have good diagnostic value in differentiating normal tissues from pRCC (Figure 8E).

The results of univariate and multivariate Cox regression analyses of the hub genes are shown in Figure 9A and 9B. The univariate Cox regression analysis results showed that these genes were both significantly associated with a shorter OS (Figure 9A). In the multivariate Cox regression analysis, *ASPM*, *BUB1B*, and *TPX2* were found to be independent prognostic factors of patients with pRCC (Figure 9B). Based on the results of both the univariate and multivariate Cox regression analyses, we further constructed a nomogram model of these genes and assessed the predictive significance of the model via a time-dependent ROC and calibration curve (Figure 9C-9E). The nomogram achieved a good area under the ROC curve, between 0.787 and 0.893, and fitted well with the obtained calibration curves, indicating that the model provided good predictability.

The protein expression levels of the hub genes were verified through the HPA database, and the results are shown in Figure 10A-10G. *TPX2*, *TOP2A*, *CEP55*, and *CEBPF* were confirmed to be expressed in renal tumors. External validation for these hub genes found that they were consistent with the analyses in TCGA (Figure 10H-10J). It should be pointed out that *ASPM* and *CEP55* were not annotated successfully in GSE2703 dataset (Figure 10J).

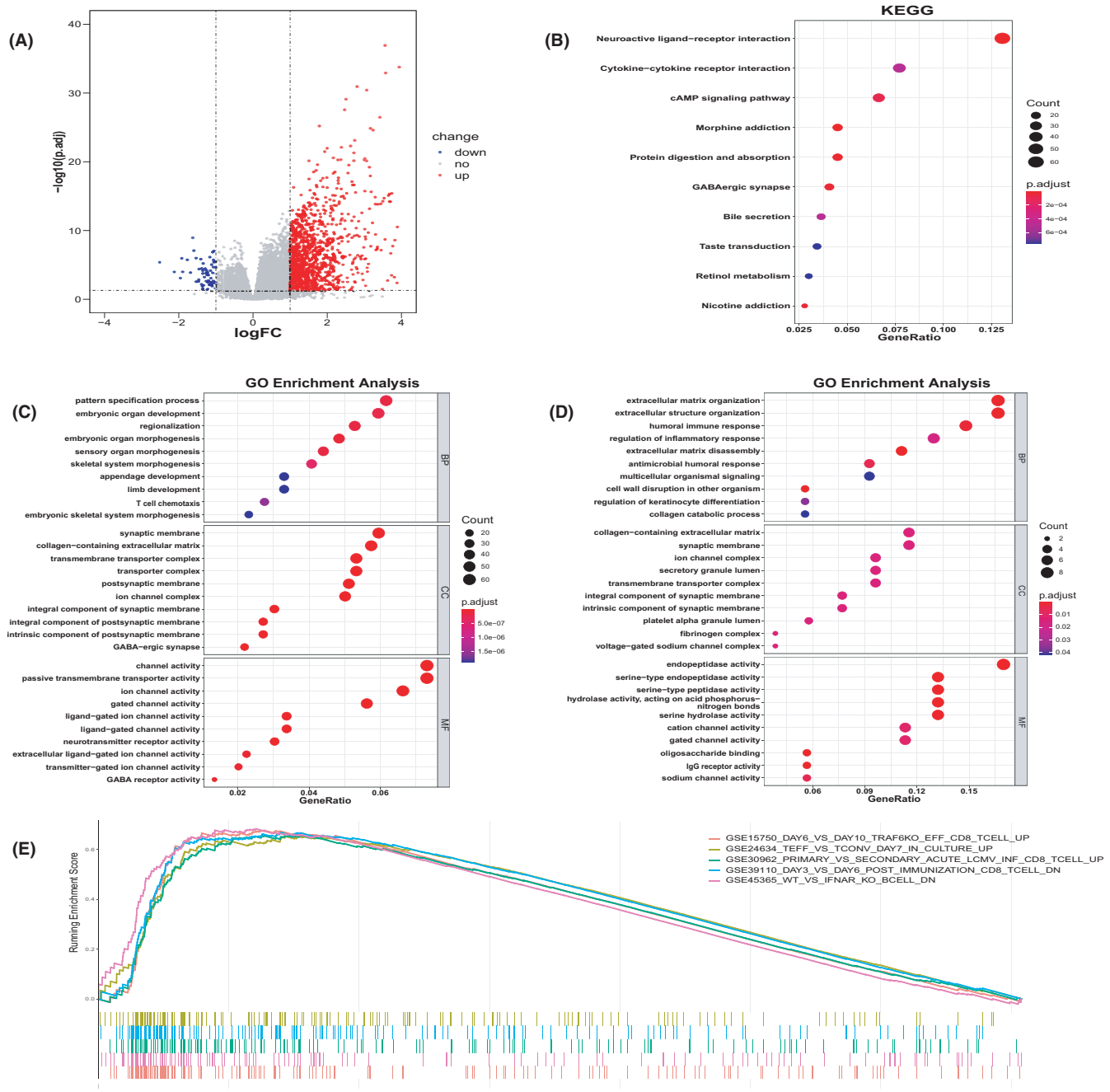

FIGURE 5 Identification and functional annotation of DEGs. (A) Volcano plot of the DEGs between low- and high-risk groups. Red nodes represent the significantly upregulated genes, and blue nodes represent the significantly downregulated genes. (B) Top 10 of KEGG enrichment analysis of DEGs. (C) The significant GO terms of upregulated genes. (D) The significant GO terms of downregulated genes. (E) Top 5 of GSEA enrichment analysis of DEGs. Abbreviations: DEGs, differentially expressed genes; GO, Gene Ontology; GSEA, gene set enrichment analysis; KEGG, Kyoto Encyclopedia of Genes and Genomes

3.8 | Validation of the correlation between the hub genes and the TIICs

The correlation of the nine hub genes with 22 TIICs was analyzed, and the results are shown in Figure 11A. It was observed that macrophages M1 and macrophages M2 were related to OS, and they were also significantly associated with the nine hub genes. Moreover, the XCELL algorithm was used to verify the correlation between the hub genes and immune infiltration of macrophages. The results of the correlation analysis as calculated via XCELL were consistent with the results obtained using the CIBERSORT algorithm (Figure 11B).

4 | DISCUSSION

Tumor-infiltrating immune cells are a critical part of the TME. These cells regulate tumor growth, invasion, and metastasis by altering the immune status of tumor cells. In this study, we evaluated the status

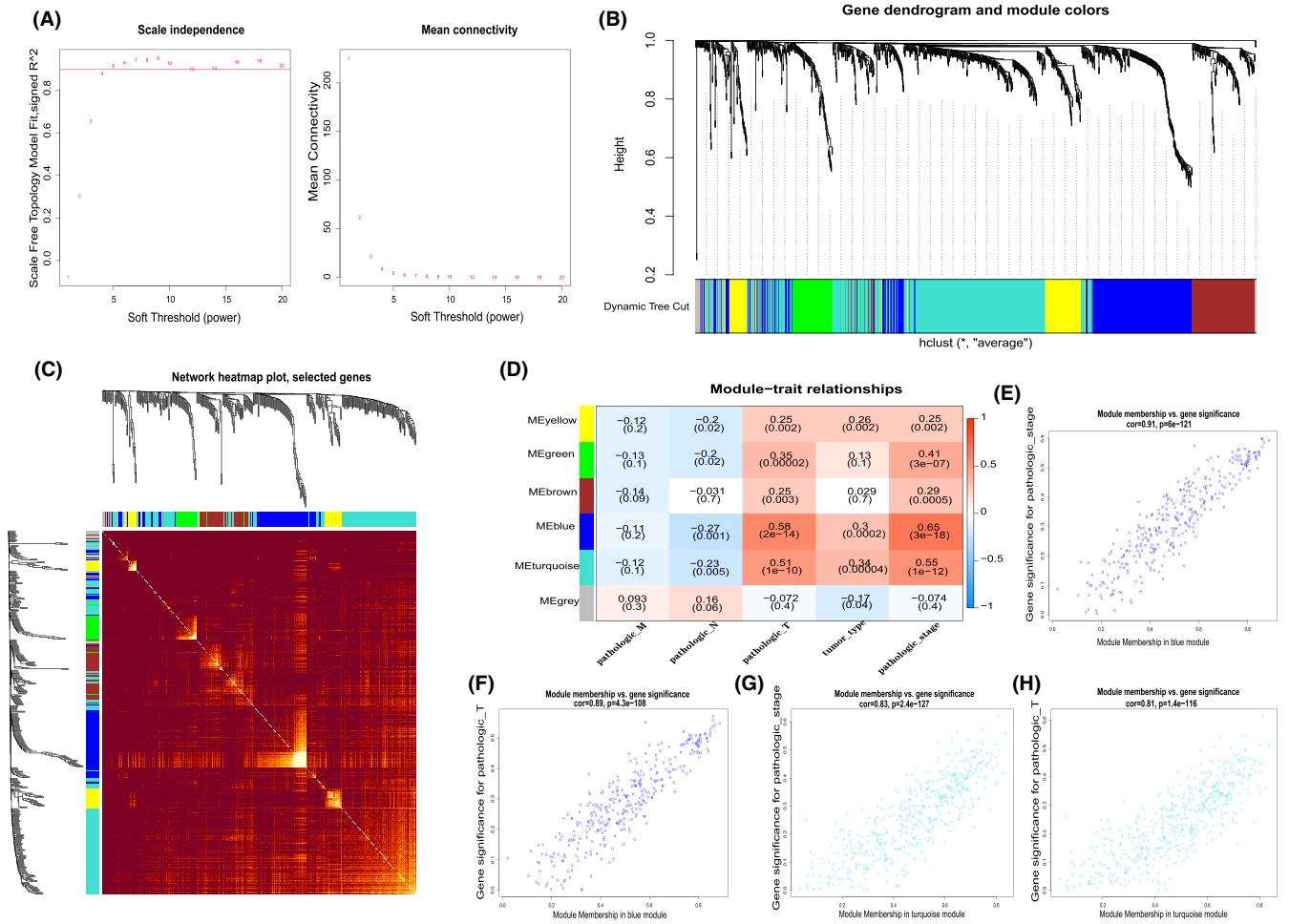

FIGURE 6 Identification of the core modules via constructing a weighted gene co-expression network. (A) Analysis of the scale-free topology model fit index for soft threshold powers (β). (B) A cluster dendrogram was built based on the dissimilarity of the topological overlap, which presents six gene co-expression modules in these DEGs, the gray module indicates no co-expression between the genes. (C) Heatmap of the weighted gene co-expression network. (D) Heatmap of the correlation between module eigengenes and clinical traits of pRCC. (E)–(F) The scatterplot of GS for pathologic stage and T stage vs MM in the blue module. (G)–(H) The scatterplot of GS for pathologic stage and T stage versus MM in the turquoise module. Abbreviations: DEGs, differentially expressed genes; GS, gene significance; MM, module membership; pRCC, papillary renal cell carcinoma

of the nine TIICs related to OS in 291 patients with pRCC and found that patients with high-risk scores had a shorter overall survival time. Nine immune-related markers for pRCC were screened out via multiple analyses by comparing the differences between the high-risk and low-risk groups.

In this study, nine TIICs related to OS, including follicular helper T cells, macrophages M1, activated dendritic cells, regulatory T cells (Tregs), B-cell memory, CD8 T cells, macrophages M2, naïve B cells, CD4 memory-activated T cells, and resting T-cell CD4 memory, were identified via univariate Cox analysis. Previous studies have shown that macrophages are one of the most abundant cell types in the tumor microenvironment, contributing to tumor progression.[24,25] In our study, correlation analysis with clinical traits showed that macrophages affected the tumor type, pathological stage, TNM stage, and prognosis of patients with pRCC. Moreover, we found that macrophages were also the most abundant cell types in pRCC and were significantly associated with the nine hub genes we obtained (*TOP2A*, *BUB1B*, *BUB1*, *TPX2*, *PBK*, *CEP55*, *ASPM*, *RRM2*, and *CENPF*). Therefore, the results of previous studies are consistent with the results obtained in this study.[24,25]

Functional enrichment analysis showed that the DEGs in the high- and low-risk groups were significantly associated with tumor cell infiltration-related pathways. Both the cytokine-cytokine receptor and cAMP signaling pathways were significantly enriched, based on our analyses. Cytokines play an important role in cancer-related immune responses and promote tumor angiogenesis, tumor cell invasion, and tumor cell metastasis.[26,27] IFN-γ is one of the cytokines necessary in immunomodulation and anticancer immunity, which also induces the expression of PD-L1 in most tumor cells.[28] The cAMP signaling pathway is a critical modulator of specific tumor cell properties such as proliferation, differentiation, and migration.[29] Dou A-X et al. showed that regulatory T cells might suppress the antitumor immune response through the intercellular transport of cAMP and the activation of the cAMP-protein kinase A signaling

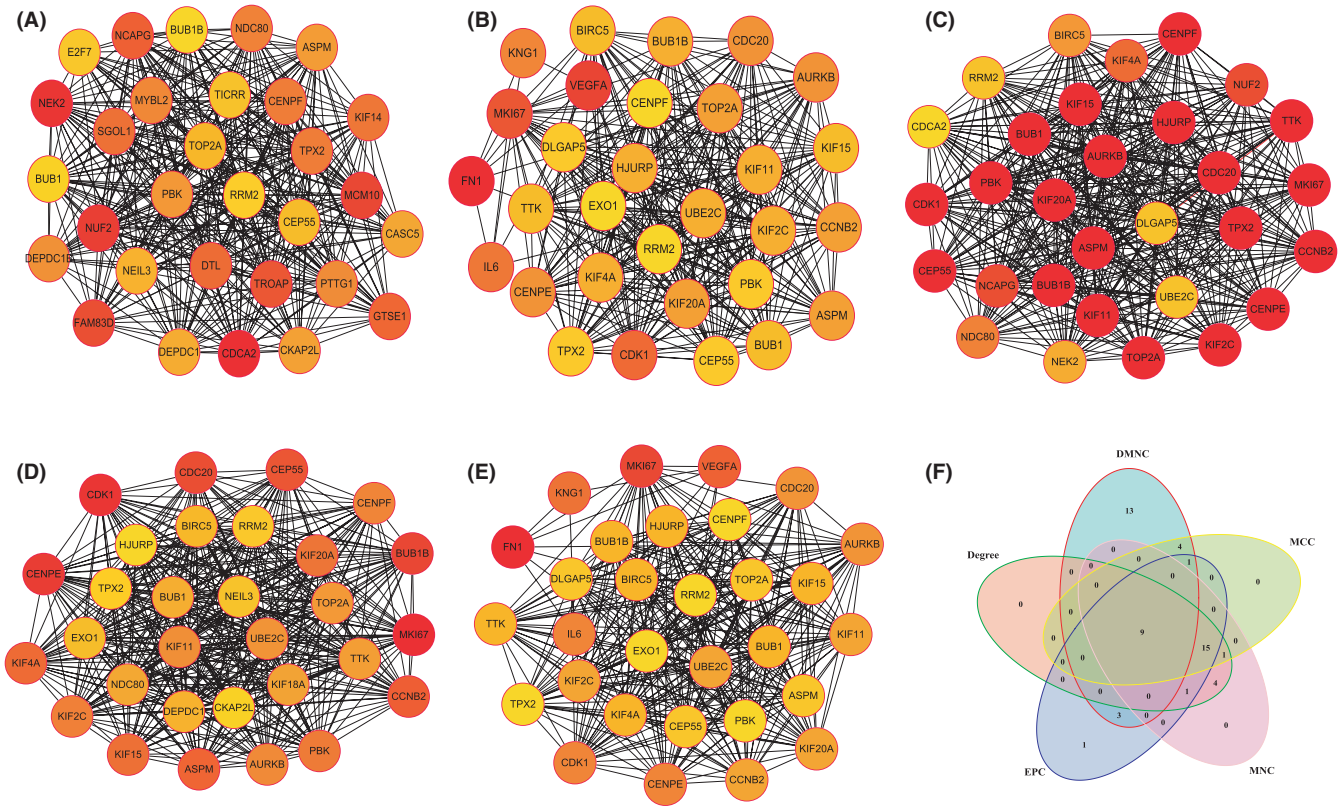

FIGURE 7 Identification of hub genes via multiple algorithms. (A)–(E) Screening the top 30 genes in the PPI network of DEGs as central genes using five algorithms including Closeness, MCC, MNC, Degree, and EPC. (F) Multiple algorithms obtain a Venn diagram of common genes. Abbreviations: DEGs: differentially expressed genes; EPC: edge percolated component; MCC: maximal clique centrality; MNC: maximum neighborhood component; PPI: protein-protein interaction

TABLE 1 The information of nine hub genes

Gene	Gene ID	Full name
TPX2	22974	*TPX2* microtubule nucleation factor
TOP2A	7153	DNA topoisomerase II alpha
BUB1B	701	*BUB1* mitotic checkpoint serine/threonine kinase B
BUB1	699	*BUB1* mitotic checkpoint serine/threonine kinase
CEP55	55156	centrosomal protein 55
RRM2	6241	ribonucleotide reductase regulatory subunit M2
ASPM	259266	assembly factor for spindle microtubules
PBK	55872	PDZ binding kinase
CENPF	1063	centromere protein F

pathway.[30] Furthermore, the GSEA found that regulatory T cells were significantly enriched in pRCC, which was consistent with the KEGG enrichment results.

The underlying mechanism of the hub genes remains unclear in pRCC. Currently, some studies have reported the effect of the expression of hub genes, such as *TPX2*, *TOP2A*, and *BUB1* in other tumors.[31-33] *TOP2A* cleavage is a broad DNA damage mechanism found in oncogenic translocations.[34] MiR-139–5 up-regulates the expression of its target gene *TOP2A* and promotes the progression of ccRCC.[35] Downregulation of *TPX2* inhibits the proliferation and invasion of endometrial cancer cells and promotes the apoptosis of EC-derived cells.[36] It also suppresses the growth of liver cancer by regulating the PI3K/AKT signaling pathway.[31] A dysregulated expression of *BUB1* leads to aberrant chromosomal replication and aneuploidy, contributing to the development of tumors.[32] Previous studies have demonstrated that these three genes play an important role in tumorigenesis and cancer development.

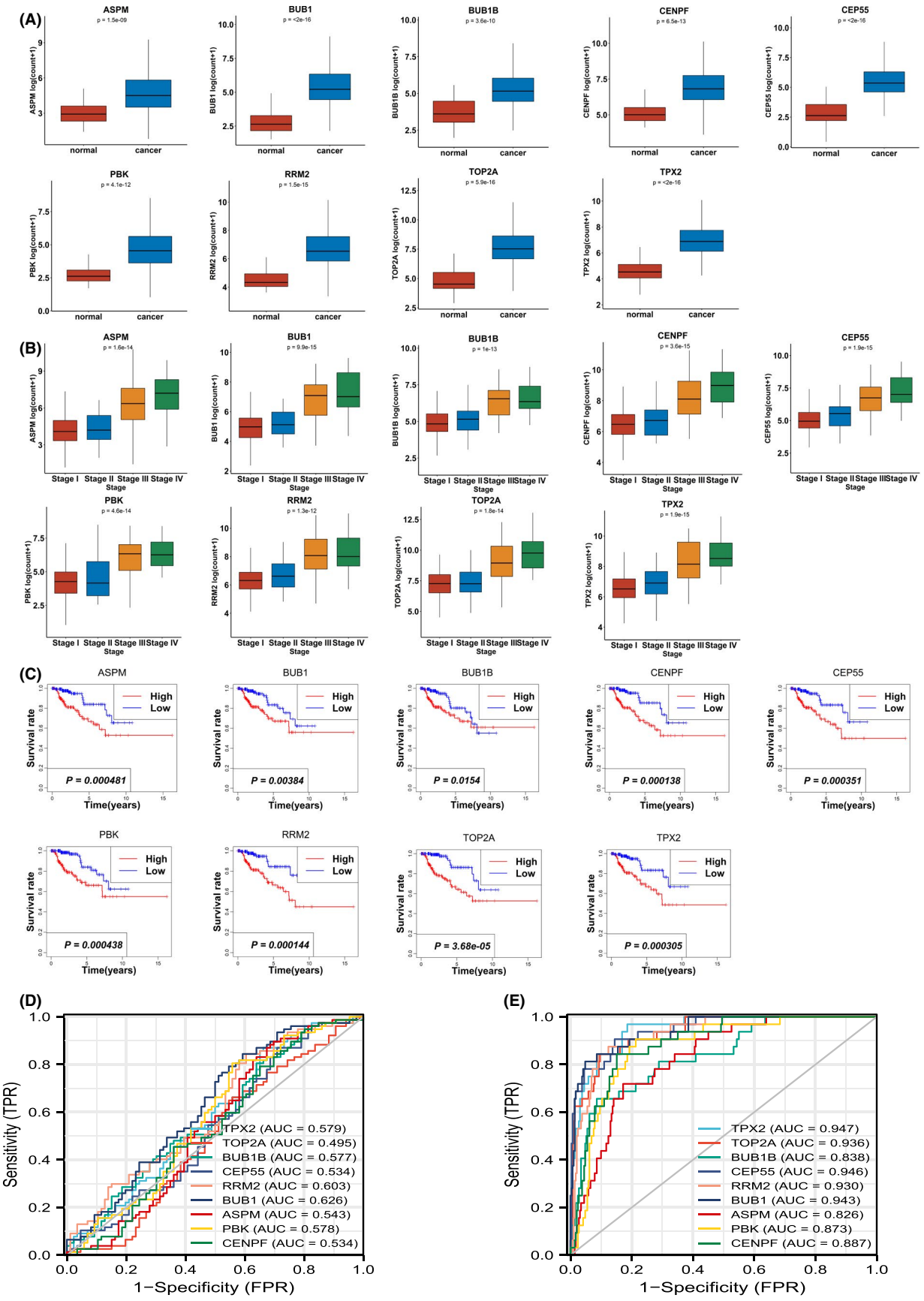

FIGURE 8 Validation of the hub genes in different clinical traits. (A) The expression difference in the nine hub genes between normal and tumor. (B) The relationship between the expression of the nine hub genes and pathologic stage. (C) Kaplan-Meier survival analysis of the nine hub genes. (D) Diagnostic value of the nine hub genes in different tumor types. (E) Diagnostic value of the nine hub genes in the tumor

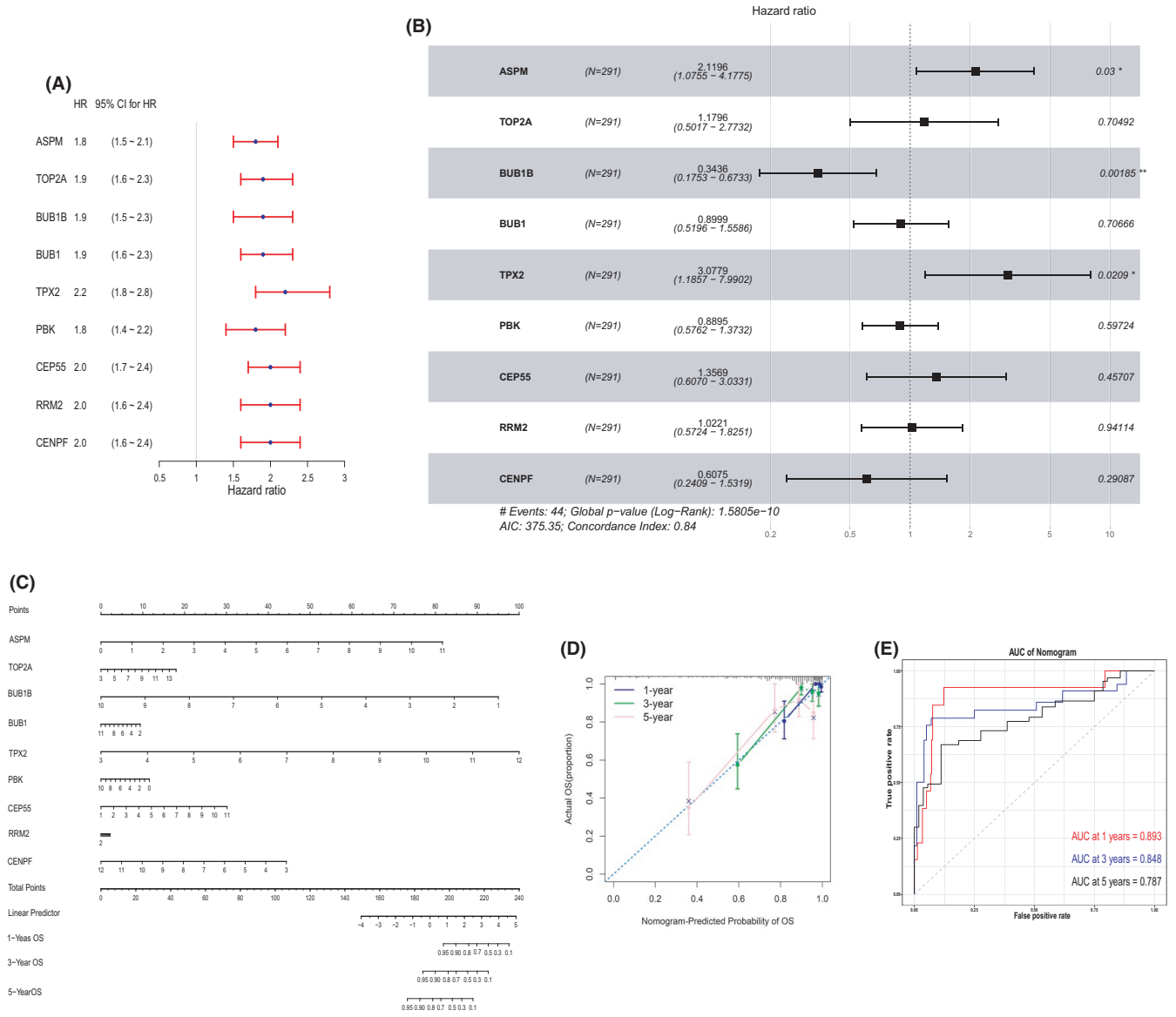

FIGURE 9 Construction of the prognostic signature based on hub genes. (A) The forest plot of the nine hub genes via univariate Cox regression analysis. (B) The forest plot of the nine hub genes multivariate Cox regression analysis. (C) Nomogram integrated the nine immune-related hub genes. (D) The calibration plot of the nomogram for agreement test between 1-, 3-, and 5-year OS prediction and actual outcome in TCGA dataset. (E) The time-dependent ROC curves of the nomogram in TCGA dataset. Abbreviations: ROC, receiver operating characteristic; TCGA, The Cancer Genome Atlas

In addition, we found that these hub genes such as *BUB1B*, *PBK*, *CEP55*, *ASPM*, *RRM2*, and *CENPF* strongly correlated with macrophages and had clinical significance in pathological staging, survival, and diagnostic value. A nomogram obtained containing these hub genes had good predictability, suggesting that these genes are significantly associated with prognosis. Therefore, these hub genes are potential immune-related biomarkers for pRCC. Additionally, our conclusions are supported by the results of previous studies. Mechanistically, it modulates the transcriptional activation of the mitotic checkpoint kinase *BUB1B*, which also promotes tumor growth and chemoresistance, leading to poor outcomes for patients with lung adenocarcinoma.[37] An *in vitro* study showed that the overexpression of *BUB1B* enhanced the proliferation, migration, and invasion of prostate cancer cell lines, while the removal of *BUB1B* did not affect these cell functions.[38] *PBK* is a novel serine-threonine kinase related to the mitogen-activated protein kinase (MAPK) family. It is an important link in many carcinogenic signaling pathways, including p38, extracellular signal-regulated kinase 1/2 (ERK1/2), and the FAK/Src-MMP signal pathways.[33] *CEP55* is a key regulatory factor of the cytoplasmic split and is associated with genomic instability; genomic instability is a hallmark of cancer.[39] Carcinogenic *CEP55* regulates the proliferation, migration, and invasion of tumor cells as mediated by the PI3K/AKT/mTOR pathway and is related to poor prognosis.[40] The potential role of *ASPM* on pRCC is still unclear; however, it was highly expressed in various cancers such as endometrial cancer and lung squamous cell carcinoma and is related to poor

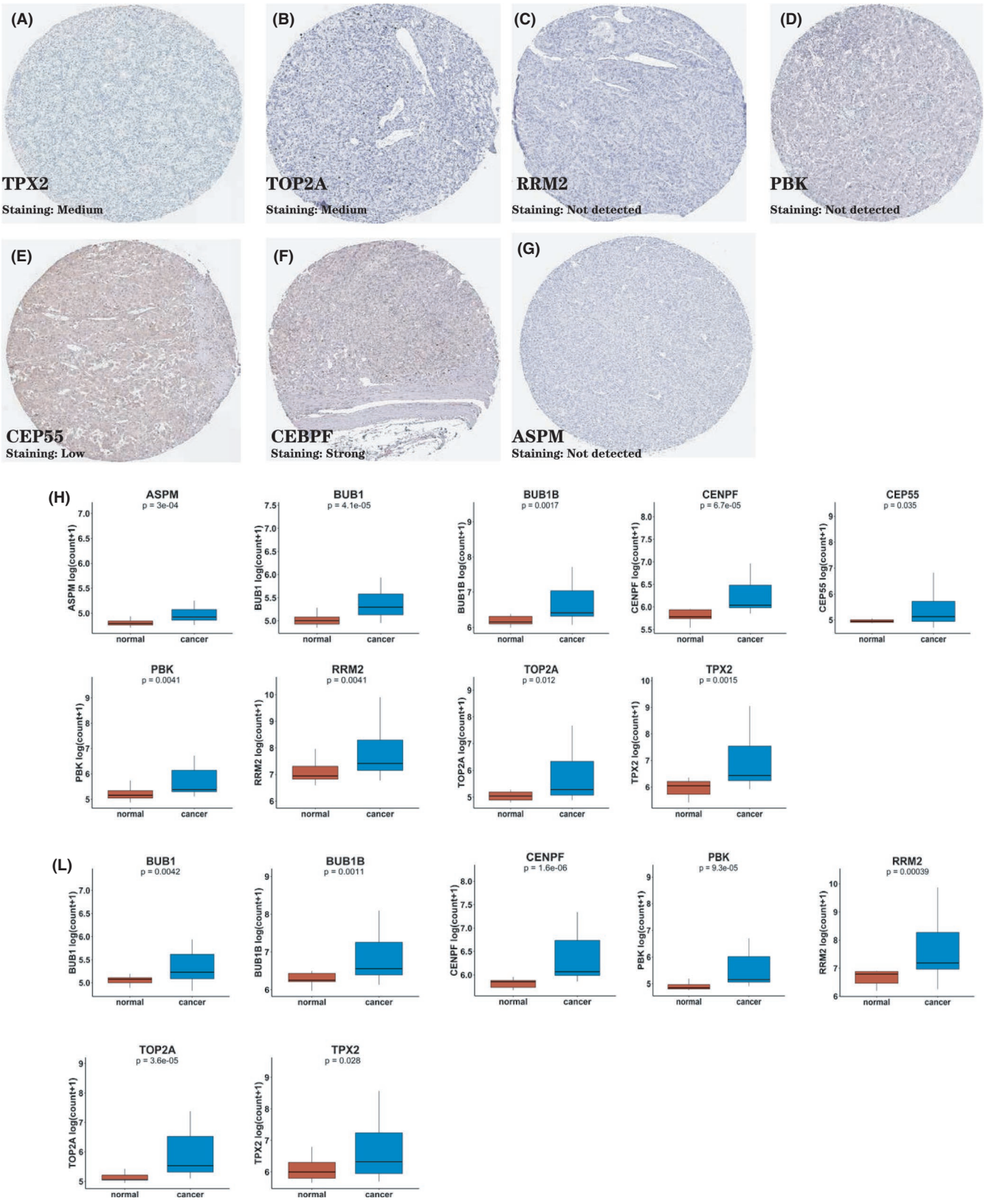

FIGURE 10 IHC of hub genes and validation of the hub genes in GEO dataset. (A)–(G) IHC of *TPX2*, *TOP2A*, *RRM2*, *PBK*, *CEP55*, *CEBPF*, and *ASPM* in renal carcinoma. (H) Independent dataset validation of the nine hub genes in GSE26574. (J) Independent dataset validation of these hub genes except *ASPM* and *CEP55* in GSE26574. Abbreviations: IHC, immunohistochemistry

clinical prognosis and an increased risk of cancer recurrence.[41,42] *RRM2* is essential for DNA synthesis and repair and is frequently overexpressed in various cancers.[43] The oncogenic role of *RRM2* has been linked to the promotion of epithelial-mesenchymal transition (EMT) and angiogenesis.[43] *CENPF* is a critical regulator of cancer metabolism, and the silencing of *CENPF* has been shown to increase the

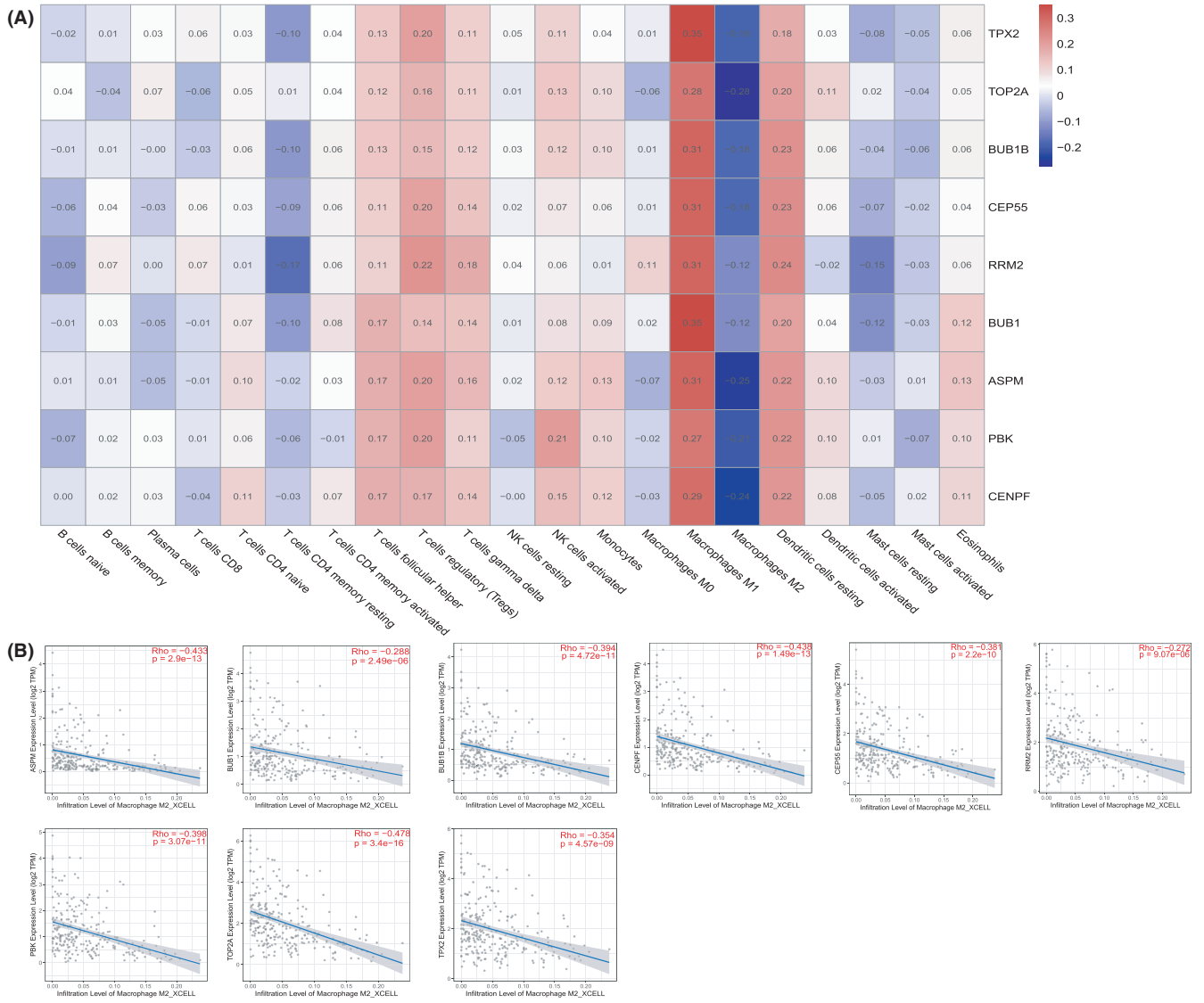

FIGURE 11 Correlation between the nine hub genes and immune cells. (A) The heatmap plot of the Spearman correlation coefficients between the nine hub genes and immune cells. (B) Validation of the correlation between the nine hub genes and macrophages via the XCELL algorithm

expression of inactive forms of pyruvate kinase M2, a rate-limiting enzyme indispensable for an irreversible reaction in glycolysis and reduces global bio-energetic capacity, acetyl-CoA production, histone acetylation, and lipid metabolism.[44] The above analysis mainly reveals that the hub genes obtained play an important role in tumor development and may be potential therapeutic targets in the treatment of pRCC.

In summary, nine prognostic-related immune cells and nine hub genes were identified, providing compelling insights into the pathogenesis of pRCC and may serve as potential therapeutic targets for pRCC. These nine types of immune cells may also provide important clues, so we can better understand the immune microenvironment in pRCC. The hub genes obtained can be considered as biomarkers for the prognosis of pRCC and may also serve as key targets for immunotherapy in pRCC. However, it is important to point out that the evidence presented in this study was obtained indirectly from bioinformatics analyses, which is its major limitation.

ACKNOWLEDGMENT

We are grateful for TCGA, UCSC, and GEO database. We are grateful to Bullet Edits Limited for polishing.

CONFLICT OF INTEREST

No potential conflict of interest was reported by the author(s).

AUTHOR CONTRIBUTIONS

Ran Deng and Jianpeng Li contributed to this article equally. RD and PL involved in the study conception and design.LY. RD, PL, HZ, ZZ, JM, and JC analyzed the data, authored or reviewed drafts of the paper, and approved the final draft.

ETHICAL APPROVAL

All data used in this work are publicly available.

DATA AVAILABILITY STATEMENT

All original data were loaded from the public database. These data are publicly available.

ORCID

Ran Deng https://orcid.org/0000-0002-9778-7561
Li Yang https://orcid.org/0000-0003-1151-5942

REFERENCES

1. Liu X, Mao YH, He XM, Zhang YJ, Sun Y. Analysis on inpatient health expenditures of renal cell carcinoma in a grade-A tertiary hospital in Beijing. *Chin Med J*. 2017;130(20):2447-2452.
2. Huang J, Huang D, Yan J, et al. Comprehensive subgroup analyses of survival outcomes between clear cell renal cell adenocarcinoma and papillary renal cell adenocarcinoma. *Cancer Med*. 2020;9(24):9409-9418.
3. Mendhiratta N, Muraki P, Sisk AE Jr, Shuch B. Papillary renal cell carcinoma: Review. *Urol Oncol*. 2021;39(6):327-337.
4. Gai W, Peng Z, Liu CH, Zhang L, Jiang H. Advances in cancer treatment by targeting the neddylation pathway. *Front Cell Dev Biol*. 2021;9:653882.
5. Marigo I, Zilio S, Desantis G, et al. T cell cancer therapy requires CD40-CD40L activation of tumor necrosis factor and inducible nitric-oxide-synthase-producing dendritic cells. *Cancer Cell*. 2016;30(3):377-390.
6. Shirota H, Klinman DM, Ito SE, Ito H, Kubo M, Ishioka C. IL4 from T follicular helper cells downregulates antitumor immunity. *Cancer Immunol Res*. 2017;5(1):61-71.
7. Hassan G, Seno M. Blood sand cancer: Cancer stem cells as origin of hematopoietic cells in solid tumor microenvironments. *Cells*. 2020;9(5):1293.
8. van Asten SD, de Groot R, van Loenen MM, et al. T cells expanded from renal cell carcinoma display tumor-specific CD137 expression but lack significant IFN-gamma, TNF-alpha or IL-2 production. *Oncoimmunology*. 2021;10(1):1860482.
9. McDermott DF, Huseni MA, Atkins MB, et al. Clinical activity and molecular correlates of response to atezolizumab alone or in combination with bevacizumab versus sunitinib in renal cell carcinoma. *Nat Med*. 2018;24(6):749-757.
10. Newman AM, Steen CB, Liu CL, et al. Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat Biotechnol*. 2019;37(7):773-782.
11. Kolde R. pheatmap: Pretty Heatmaps. 2019.
12. Simko} TWaV. R package "corrplot": Visualization of a Correlation Matrix. 2017.
13. Lumley MGaT. forestplot: Advanced Forest Plot Using 'grid' Graphics. 2020.
14. T T. A Package for Survival Analysis in R. 2020.
15. Jr FEH. rms: Regression Modeling Strategies. 2020.
16. Blanche P, Dartigues JF, Jacqmin-Gadda H. Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. *Stat Med*. 2013;32(30):5381-5397.
17. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):550.
18. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. 2012;16(5):284-287.
19. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008;9:559.
20. Kassambara A, ggpubr: 'ggplot2' Based Publication Ready Plots. 2020.
21. Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*. 2011;12:77.
22. Revelle W. psych: Procedures for Psychological, Psychometric, and Personality Research. 2020.
23. Li T, Fu J, Zeng Z, et al. TIMER2.0 for analysis of tumor-infiltrating immune cells. *Nucleic Acids Res*. 2020;48(W1):W509-W514.
24. Terashima Y, Toda E, Itakura M, et al. Targeting FROUNT with disulfiram suppresses macrophage accumulation and its tumor-promoting properties. *Nat Commun*. 2020;11(1):609.
25. Li R, Serrano JC, Xing H, et al. Interstitial flow promotes macrophage polarization toward an M2 phenotype. *Mol Biol Cell*. 2018;29(16):1927-1940.
26. Qin F, Liu X, Chen J, et al. Anti-TGF-beta attenuates tumor growth via polarization of tumor associated neutrophils towards an anti-tumor phenotype in colorectal cancer. *J Cancer*. 2020;11(9):2580-2592.
27. Xu R, Li Y, Liu Y, et al. How are MCPIP1 and cytokines mutually regulated in cancer-related immunity? *Protein Cell*. 2020;11(12):881-893.
28. Nishino M, Ramaiya NH, Hatabu H, Hodi FS. Monitoring immune-checkpoint blockade: response evaluation and biomarker development. *Nat Rev Clin Oncol*. 2017;14(11):655-668.
29. Schonrath K, Pan W, Klein-Szanto AJ, Braunewell KH. Involvement of VILIP-1 (visinin-like protein) and opposite roles of cyclic AMP and GMP signaling in in vitro cell migration of murine skin squamous cell carcinoma. *Mol Carcinog*. 2011;50(5):319-333.
30. Dou AX, Feng LL, Liu XQ, Wang X. Cyclic adenosine monophosphate involvement in low-dose cyclophosphamide-reversed immune evasion in a mouse lymphoma model. *Cell Mol Immunol*. 2012;9(6):482-488.
31. Huang DH, Jian J, Li S, Zhang Y, Liu LZ. TPX2 silencing exerts antitumor effects on hepatocellular carcinoma by regulating the PI3K/AKT signaling pathway. *Int J Mol Med*. 2019;44(6):2113-2122.
32. Sun Z, Xiao B, Jha HC, Lu J, Banerjee S, Robertson ES. Kaposi's sarcoma-associated herpesvirus-encoded LANA can induce chromosomal instability through targeted degradation of the mitotic checkpoint kinase Bub1. *J Virol*. 2014;88(13):7367-7378.
33. Mao P, Bao G, Wang YC, et al. PDZ-binding kinase-dependent transcriptional regulation of CCNB2 promotes tumorigenesis and radioresistance in glioblastoma. *Transl Oncol*. 2020;13(2):287-294.
34. Yu X, Davenport JW, Urtishak KA, et al. Genome-wide TOP2A DNA cleavage is biased toward translocated and highly transcribed loci. *Genome Res*. 2017;27(7):1238-1249.
35. Zhang C, Qu Y, Xiao H, et al. LncRNA SNHG3 promotes clear cell renal cell carcinoma proliferation and migration by upregulating TOP2A. *Exp Cell Res*. 2019;384(1):111595.
36. Jiang T, Sui D, You D, et al. MiR-29a-5p inhibits proliferation and invasion and induces apoptosis in endometrial carcinoma via targeting TPX2. *Cell Cycle*. 2018;17(10):1268-1278.
37. Shih JH, Chen HY, Lin SC, et al. Integrative analyses of non-coding RNAs reveal the potential mechanisms augmenting tumor malignancy in lung adenocarcinoma. *Nucleic Acids Res*. 2020;48(3):1175-1191.
38. Fu X, Chen G, Cai ZD, et al. Overexpression of BUB1B contributes to progression of prostate cancer and predicts poor outcome in patients with prostate cancer. *Onco Targets Ther*. 2016;9:2211-2220.
39. Kalimutho M, Sinha D, Jeffery J, et al. CEP55 is a determinant of cell fate during perturbed mitosis in breast cancer. *EMBO Mol Med*. 2018;10(9):e8566.
40. Jia Y, Xiao Z, Gongsun X, et al. CEP55 promotes the proliferation, migration and invasion of esophageal squamous cell carcinoma via the PI3K/Akt pathway. *Onco Targets Ther*. 2018;11:4221-4232.

41. Yuan YJ, Sun Y, Gao R, Yin ZZ, Yuan ZY, Xu LM. Abnormal spindle-like microcephaly-associated protein (ASPM) contributes to the progression of Lung Squamous Cell Carcinoma (LSCC) by regulating CDK4. *J Cancer*. 2020;11(18):5413-5423.
42. Zhou J, Wang H, Sun W, Han N, Chen L. ASPM is a predictor of overall survival and has therapeutic potential in endometrial cancer. *Am J Transl Res*. 2020;12(5):1942-1953.
43. Mazzu YZ, Armenia J, Chakraborty G, et al. A novel mechanism driving poor-prognosis prostate cancer: Overexpression of the DNA repair gene, ribonucleotide reductase small subunit M2 (RRM2). *Clin Cancer Res*. 2019;25(14):4480-4492.
44. Shahid M, Lee MY, Piplani H, et al. Centromere protein F (CENPF), a microtubule binding protein, modulates cancer metabolism by regulating pyruvate kinase M2 phosphorylation signaling. *Cell Cycle*. 2018;17(24):2802-2818.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Deng R, Li J, Zhao H, et al. Identification of potential biomarkers associated with immune infiltration in papillary renal cell carcinoma. *J Clin Lab Anal*. 2021;35:e24022. https://doi.org/10.1002/jcla.24022