



Identification of a Novel Protein-Based Signature to Improve Prognosis Prediction in Renal Clear Cell Carcinoma

Guangdi Chu^{1†}, Ting Xu^{2†}, Guanqun Zhu¹, Shuaihong Liu¹, Haitao Niu^{1*} and Mingxin Zhang^{1*}

OPEN ACCESS

Edited by:

Guiting Lin,
University of California,
San Francisco, United States

Reviewed by:

Yuan Yuan Zhang,
Wake Forest Baptist Medical Center,
United States
Ruili Guan,
Peking University First Hospital, China

*Correspondence:

Haitao Niu
niuht0532@126.com
Mingxin Zhang
dr.zmx@foxmail.com

[†]These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Molecular Diagnostics
and Therapeutics,
a section of the journal
Frontiers in Molecular Biosciences

Received: 11 November 2020

Accepted: 08 March 2021

Published: 25 March 2021

Citation:

Chu G, Xu T, Zhu G, Liu S, Niu H
and Zhang M (2021) Identification of a
Novel Protein-Based Signature
to Improve Prognosis Prediction
in Renal Clear Cell Carcinoma.
Front. Mol. Biosci. 8:623120.
doi: 10.3389/fmolb.2021.623120

¹ Department of Urology, The Affiliated Hospital of Qingdao University, Qingdao, China, ² Department of Geratology, The 971th Hospital of PLA Navy, Qingdao, China

Background: Clear cell renal cell carcinoma (ccRCC) is one of the most common types of malignant adult kidney cancer, and its incidence and mortality are not optimistic. It is well known that tumor-related protein markers play an important role in cancer detection, prognosis prediction, or treatment selection, such as carcinoembryonic antigen (CEA), programmed cell death 1 (PD-1), programmed cell death 1 ligand 1 (PD-L1), and cytotoxic T lymphocyte antigen 4 (CTLA-4), so a comprehensive analysis was performed in this study to explore the prognostic value of protein expression in patients with ccRCC.

Materials and Methods: Protein expression data were obtained from The Cancer Proteome Atlas (TCPA), and clinical information were downloaded from The Cancer Genome Atlas (TCGA). We selected 445 patients with complete information and then separated them into a training set and testing set. We performed univariate, least absolute shrinkage and selection operator (LASSO) Cox analyses to find prognosis-related proteins (PRPs) and constructed a protein signature. Then, we used stratified analysis to fully verify the prognostic significance of the prognostic-related protein signature score (PRPscore). Besides, we also explored the differences in immunotherapy response and immune cell infiltration level in high and low score groups. The consensus clustering analysis was also performed to identify potential cancer subgroups.

Results: From the training set, a total of 233 PRPs were selected, and a seven-protein signature was constructed, including ACC1, AR, MAPK, PDK1, PEA15, SYK, and BRAF. Based on the PRPscore, patients could be divided into two groups with significantly different overall survival rates. Univariate and multivariate Cox regression analyses proved that this signature was an independent prognostic factor for patients ($P < 0.001$). Moreover, the signature showed a high ability to distinguish prognostic outcomes among

subgroups, and the low score group had a better prognosis ($P < 0.001$) and better immunotherapy response ($P = 0.003$) than the high score group.

Conclusion: We constructed a novel protein signature with robust predictive power and high clinical value. This will help to guide the disease management and individualized treatment of ccRCC patients.

Keywords: clear cell renal cell carcinoma, proteomics, immunotherapy, TCGA, TCPA

INTRODUCTION

Renal cell carcinoma (RCC) is one of the most common cancers originating from the renal epithelium (Wang X. M. et al., 2020), with an estimated 73,750 new cases and 14,830 deaths in America in 2020 (Siegel et al., 2020). The ccRCC is the most frequent form of RCC, affecting 80–90% of patients (Ljungberg et al., 2019). The high heterogeneity and delayed detection of ccRCC have been proven to be obstacles to treatment and maybe major factors in recurrence (Fendler et al., 2020). Therefore, the identification of reliable tools for early detection and prediction of clinical outcomes is critical to the improvement of patient treatment and prognosis.

Proteins are important for maintaining the normal functions of the human body. Protein stability requires proper production, degradation, folding, and activity of proteins, which are essential for any cellular function (Bartelt and Widenmaier, 2019). Moreover, proteins are also closely associated with tumorigenesis, with some proteins being identified to be involved in cancer and the pathogenesis of carcinoma (Torresano et al., 2020). Based on this, studies focused on proteomics have sprung up as a result (Menschaert and Fenyo, 2017). The proteomic studies set their focus on the study of proteins at a large scale (Wu and Yang, 2020; Wu et al., 2020). They could be widely combined with genomics, epigenomics, transcriptomics, and other novels multi-omics analysis to reveal novel therapeutic targets or potential biomarkers, which could drive new strategies for diagnosis and treatment and many proteins have been widely used in clinical therapy as important drug targets (Menschaert et al., 2010), such as PD-1, PD-L1 and CTLA-4 (Aggen et al., 2020; Gulati and Vaishampayan, 2020; Xu et al., 2020). However, the overall performance of one biomarker has been unsatisfactory in terms of sensitivity and specificity. Therefore, the signature consisting of a variety of components has attracted people's attention.

The TCPA database is an open-access bioinformatics data repository that can access high-quality Reverse-phase protein arrays (RPPAs), which represents an advanced proteomic technology that can quantitatively evaluate a large number of

protein markers in thousands of samples in a cost-effective, sensitive and high-throughput way (Nishizuka et al., 2003; Tibes et al., 2006; Hennessy et al., 2010; Li et al., 2017a). And this quantitative antibody-based assay has been widely used to explore the molecular events that drive tumorigenesis or progression and to evaluate the biomarkers of cancer treatment sensitivity and drug resistance (Sheehan et al., 2005; Spurrier et al., 2008). Besides, the TCPA database includes more than 8,000 patient samples of 32 cancer types from The TCGA database, more than 650 independent cancer cell lines of 19 cell lineages (Li et al., 2013, 2017b; Hoadley et al., 2018), and other multi-omics datasets such as mRNA expression, miRNA expression, somatic copy-number alterations, somatic mutations, and DNA methylation (Chen et al., 2019). All of these data of TCGA were collected from TCGA Pan-Cancer Atlas¹ and the TCGA marker publications (Li et al., 2015). The TCPA database enables us to overcome the computational obstacles of complex RPPA data and makes it convenient for us to download and follow-up analysis of the data. Therefore, we chose to use the TCPA database and the TCGA database for our study.

In this study, we further screened and constructed a prognostic-related protein signature which composed of 7 proteins by identifying 233 protein markers from the TCPA database and TCGA database. This protein signature could be identified as an independent prognostic factor for patients with ccRCC. It could accurately predict the prognosis of patients and distinguish the people who benefit from immunotherapy. In conclusion, it would play an important role in the personalized precision therapy of patients with ccRCC in the future.

MATERIALS AND METHODS

Dataset Sources

The protein expression data of ccRCC patients was downloaded from TCPA (Chen et al., 2019), and the clinical information and transcriptome expression data were downloaded from TCGA (Linehan and Ricketts, 2019). We combined the protein expression data of patients with the clinical information based on the ID number. Any patient with incomplete protein expression data or clinical information was excluded, and we ultimately obtained complete data from 445 patients for further analysis. These patients were randomized into a training set and a testing set with the help of the R package “caret,” and the ratio of grouping was 1:1. The training set was used to construct the

Abbreviations: ccRCC, clear cell renal cell carcinoma; RCC, renal cell carcinoma; TCPA, the Cancer Proteome Atlas; TCGA, The Cancer Genome Atlas; HPA, The Human Protein Atlas; PRPs, prognosis-related proteins; LASSO, least absolute shrinkage and selection operator; ROC curves, receiver operating characteristic curves; DCA curves, decision curve analysis curves; AUC, area under the curve; OS, overall survival; PD-1, programmed cell death 1; PD-L1, programmed cell death 1 ligand 1; CTLA-4, cytotoxic T lymphocyte antigen 4; IHC, immunohistochemical; PCA, principal component analysis; GSEA, gene set enrichment analysis; FDR, false discovery rate; CC, consensus clustering; AJCC, American joint committee on cancer.

¹<https://gdc.cancer.gov/about-data/publications/pancanatlas>

risk protein signature via LASSO-Cox regression, the testing set and the whole dataset were used to test the performance of the signature. The expression data (TPM: transcripts per kilobase million) of TCGA Pan-Cancer was download from “TOIL RSEM tpm ($n = 10,535$) UCSC Toil RNAseq Recompute” of the UCSC Xena database² and the TPM expression data of GTEx database was download from “TOIL RSEM tpm ($n = 7,862$) UCSC Toil RNAseq Recompute” of the UCSC Xena database. These two cohorts were processed by the TOIL process to avoid the computational batch effects.

Construction of the Protein Signature and Calculation of Risk Scores

Possible proteins related to the prognosis of patients with ccRCC were identified by univariate Cox analysis, in which $P < 0.001$ was set as the cutoff value. Next, the LASSO analysis was used to select potential risk proteins from the significant proteins identified by the univariate Cox analysis and eliminate overfit proteins in the signature. Finally, the prognosis-related proteins identified from the LASSO algorithm were further analyzed by multivariate Cox proportional hazards regression to construct the protein signature.

To calculate the PRPscore for each patient, we used the regression coefficients calculated in the multivariate Cox proportional hazards regression analysis for proteins in the signature to weigh their values. This analysis adopted the following formula:

$$PRPscore = \sum_{i=1}^n \text{Coef } i * \text{Expr } i$$

Coef i is the coefficient of protein i in the multivariate Cox analysis; “Expr i ” is the expression value of the protein selected from the signature. The regression coefficients and the expression of the protein selected can be found in the “protein i ” parameter. We calculated and summed the results for each protein in the signature, and this sum was the PRPscore of each patient. This score was determined for each patient in the training set, and the median PRPscore of the training set was used as the cutoff value. Patients with ccRCC in the training set were divided into a high score group and a low score group by the median PRPscore. Kaplan-Meier analysis was used to test the effect of this signature on the prognosis of ccRCC patients. With the help of the “survivalROC” R package, time-dependent ROC curves were plotted to evaluate the accuracy, and risk curves were used to classify patients based on this signature.

Validation of the Prognosis-Related Protein Signature

The testing set and the whole data set were used to verify the predictive power and applicability of the multiprotein signature. Patients were divided into a high score group and a low score group by the median PRPscore of the training set. Kaplan-Meier survival curves, Receiver operating characteristic (ROC) curves, and risk curves were plotted to validate the reliability

of the protein signature constructed in the training set. Also, the risk-stratified analysis was performed in the whole dataset to validate the predictive power of the signature in more specific subgroups. And we also used the pROC package (Robin et al., 2011) to evaluate and compare the sensitivity and specificity of each protein and PRPscore. Besides, to further verify the clinical value of PRPscore, we performed Kaplan-Meier survival analysis to compare the survival difference of progression-free interval (PFI) (Liu et al., 2018) between high and low PRPscore groups, log-rank $P < 0.05$ was set as the cutoff value.

Construction and Validation of the Novel Prognostic Nomogram

Univariate and multivariate Cox regression analyses were performed in the training set, testing set, and the whole dataset to determine whether the PRPscore was an independent prognostic factor for patients with ccRCC. And $P < 0.05$ was set as the cutoff value for significance. The independent prognostic factors screened by the independent prognostic analysis were subsequently incorporated into the construction of the nomogram. And the calibration plots were plotted to test the utility of the nomogram for predicting 1, 3, and 5 years outcomes in patients. Nomograms incorporating PRPscore and clinical variables for predicting OS and PFI of patients were both plotted.

Further Analysis of Proteins in the Prognosis Related Protein Signature

Firstly, Kaplan-Meier curves of survival data were used to analyze whether the proteins in the signature affected the prognosis of patients with ccRCC via the TCGA database. $P < 0.05$ was the cutoff value. Secondly, we used the “Datasets” modules in the TCGA database to analyze the seven proteins in PRPS at the pan-cancer level (Li et al., 2017a). Thirdly, corresponding encoding genes of the proteins in the signature were found in the TCGA database, and we used the data of TCGA and GTEx datasets to analyze their differential expression level between tumor tissues and normal tissues at the pan-cancer mRNA level. Then, immunohistochemical (IHC) staining data of these proteins from normal and tumor tissues were retrieved from the Human Protein Atlas (HPA³) as external validation. Next, the Kruskal-Wallis test and Wilcoxon sign-rank test were performed to evaluate whether proteins in the signature were significantly correlated with clinical characteristics. $P < 0.05$ was also set as the cutoff value. Finally, coexpression network analysis was performed to identify more potential proteins that may affect the prognosis of ccRCC patients. $P < 0.001$ and coexpression score < 0.4 were set as the cutoff values. And the correlation of these proteins was also displayed in the chord diagram.

Exploration of Characteristic Differences Between High and Low PRPscore Groups

To further analyze the difference between the high PRPscore group and low PRPscore group identified by the risk protein

²<https://xenabrowser.net/datapages/>

³<https://www.proteinatlas.org/>

signature, The GSEA algorithm, a computational method to evaluate whether a predefined set of genes has statistically significant and consistent differences between two biological states (Subramanian et al., 2005), was performed via GSEA software from the Broad Institute⁴ to reveal positively and negatively affected pathways between the two groups, which helped us understand the mechanism more deeply. A false discovery rate (FDR) < 25% and nominal $P < 0.05$ were used as cutoff values. Besides, we used the CIBERSORT (Linehan and Ricketts, 2019) and LM22 gene signatures to quantify the proportion of immune cells in ccRCC. The algorithm can sensitively and specifically distinguish 22 human immune cell phenotypes. It is a deconvolution algorithm based on support vector regression, which uses a set of minimum reference gene expression values corresponding to each cell type to infer the proportion of cell types in large tumor sample data of mixed cell types. $P < 0.05$ indicates that the inferred cell composition is reliable.

Then, the clinical value of PRPscore was judged by clinical correlation analysis. The finding of the significant correlation between PRPscore and some clinical features will help it to play a greater role in the accurate management of patients' diseases. Finally, to evaluate the possibility of individual response to immunotherapy, the Tumour Immune Dysfunction and Exclusion (TIDE) algorithm (Jiang et al., 2018) was conducted. And the subclass mapping algorithm (Hoshida et al., 2007) was also explored in response to anti-PD-1 or anti-CTAL-4 therapy based on another published data set that included 47 melanoma patients who responded to immunotherapy (Roh et al., 2017).

The Molecular Subtypes of Clear Cell Renal Cell Cancer

Unsupervised class discovery is a useful method to categorize groups of individuals with similar biological characteristics. With the help of the R package ConsensusClusterPlus (Monti et al., 2003), consensus clustering (CC) can provide sufficient evidence to identify unsupervised classes in a dataset. In this research, we used CC to identify subgroups of ccRCC according to the expression level of proteins in the signature. The maximum number of subgroups was set as 10, and 1000 permutations were performed to ensure the stability of the classification. Kaplan-Meier analysis with the log-rank test was performed in these subgroups, and log-rank $P < 0.05$ was the cutoff value.

RESULTS

Data Processing and Construction of the Training Set and Testing Set

The flowchart of our study is shown in **Figure 1**. To fully verify the accuracy of our signature, we divided the whole dataset into a training set ($n = 224$) and a testing set ($n = 221$). The training set was used to construct the prognosis protein signature, and the testing set and whole dataset ($n = 445$) were used for validation.

⁴<https://www.gsea-msigdb.org/gsea/>

The clinical information of samples in this study could be found in **Supplementary Table 1**.

Determination of Prognosis-Related Proteins and the Protein Signature

To determine possible prognosis-related proteins, we performed univariate Cox regression analysis on expression data for each protein in the training set. A total of 233 proteins were found to be significantly connected to the overall survival (OS) of patients with ccRCC ($P < 0.001$) (**Figure 2A**). Then, LASSO analysis was used to remove proteins that were excessively associated with each other, and twelve proteins were identified as significant proteins via the LASSO analysis (**Figures 2B,C**). Next, we performed multivariate Cox proportional risk regression analysis (forward selection and backward selection) to further refine the list of identified proteins. Finally, we obtained seven important proteins: ACC1, AR, MAPK (pT202Y204), PDK1 (pS241), PEA15, SYK, and BRAF (pS445). The proteins associated with a high-risk phenotype were ACC1, PEA15, and SYK. AR, MAPK, PDK1, and BRAF were associated with a low-risk phenotype (**Figure 2D**).

To investigate the utility of the risk proteins in predicting the prognosis of ccRCC patients, we used the expression levels and estimated regression coefficients of the risk proteins to calculate the PRPscore of each patient. The formula is as follows:

$$\text{Prognostic-related protein signature score} = (0.741451282 \times \text{expression of ACC1}) + (-0.48465864 \times \text{expression of AR}) + (-0.329458826 \times \text{expression of MAPK}) + (-1.188753725 \times \text{expression of PDK1}) + (1.527500839 \times \text{expression of PEA15}) + (0.536422579 \times \text{expression of SYK}) + (-1.441692539 \times \text{expression of BRAF}).$$

According to the median PRPscore, we divided the patients in the training set into a high PRPscore group ($n = 112$) and a low PRPscore group ($n = 112$). Kaplan-Meier curves showed that the prognosis of the high PRPscore group was worse than that of the low PRPscore group ($P < 0.001$) (**Figure 3A**). Our calculations revealed that the 1-, 3- and 5-year OS rates of the high-score group in the training set were 78.6%, 63.6%, and 31.8%, respectively, while those of the low-score group in the training set were 99.1%, 93.4%, and 80.6%. Using risk curves, we analyzed the distribution of patients in the training set ranked by the PRPscore, determined the survival status of each patient, and described the expression patterns of risk proteins in the high PRPscore and low PRPscore groups. In low PRPscore patients, four proteins (AR, MAPK, PDK1, and BRAF) were upregulated while ACC1, PEA15, and SYK were downregulated. In patients with high PRPscore scores, these risk proteins showed the opposite expression patterns (**Figure 3B**). Besides, the AUC value of the PRPscore according to ROC curves was 0.860 at 1 year, 0.793 at 3 years, and 0.788 at 5 years (**Figure 3C**).

The Protein Signature Was Verified to Exhibit Excellent Performance

We used the testing set ($n = 221$) and the entire TCGA dataset ($n = 445$) to confirm the accuracy of the signature. The PRPscore of each patient in the testing set and the entire

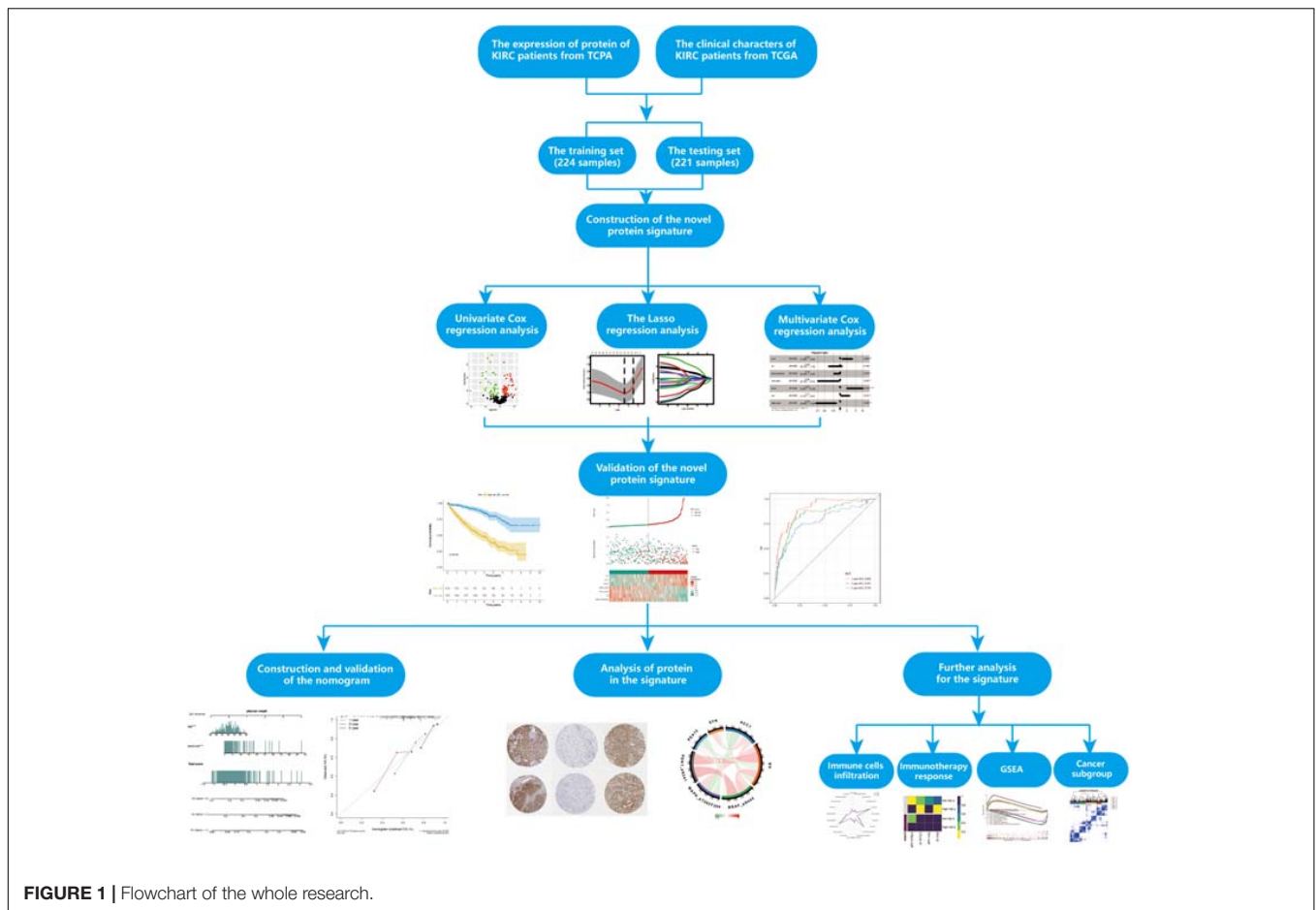
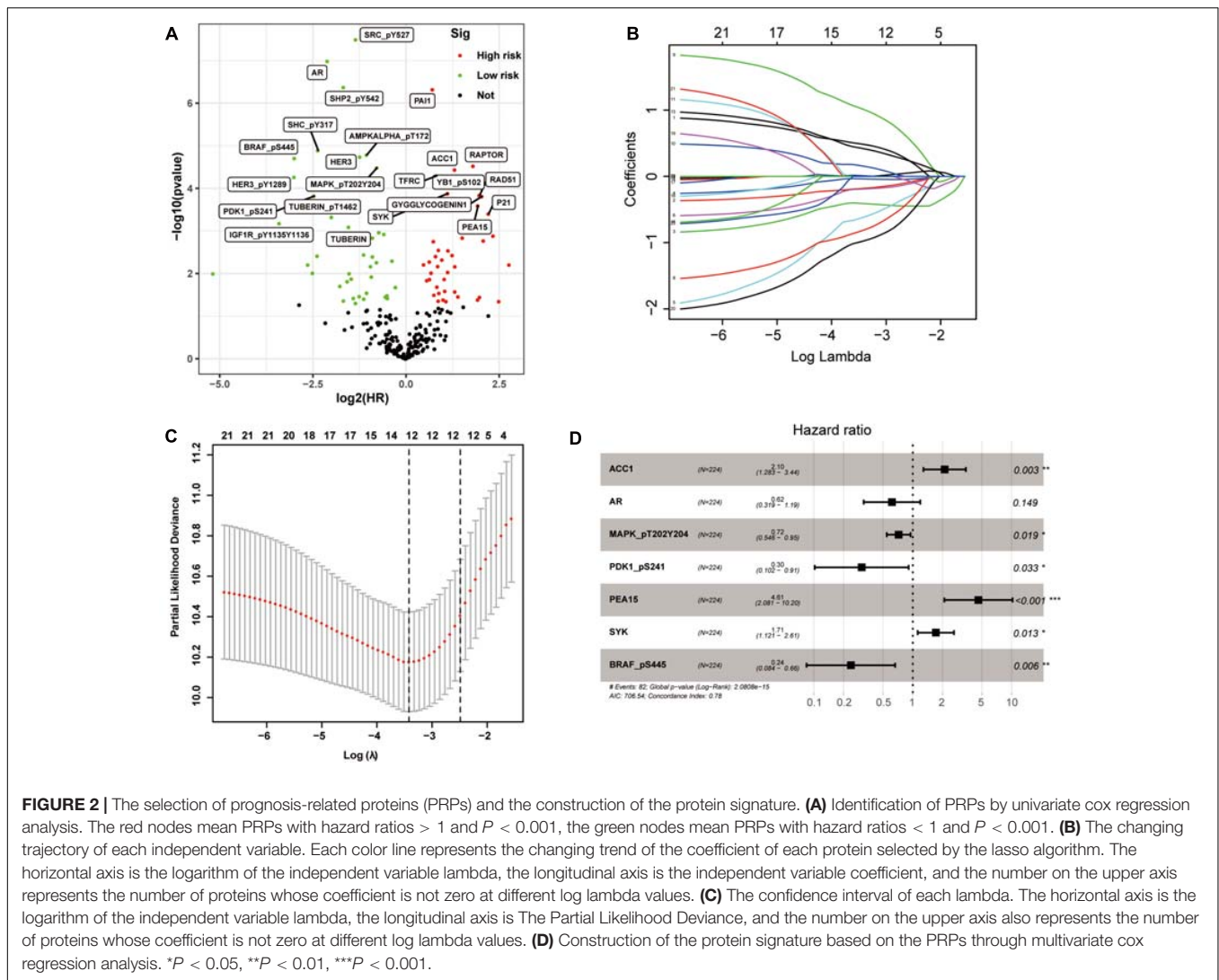


FIGURE 1 | Flowchart of the whole research.

TCGA dataset was calculated using the expression levels of the seven risk proteins (ACC1, AR, MAPK, PDK1, PEA15, SYK, and BRAF). Patients were then divided into two groups based on the cutoff value. In the testing set, 107 patients were classified as high PRPscore, and 114 were low PRPscore. In the entire TCGA dataset, 219 patients were classified as high PRPscore, and 226 were low PRPscore. The Kaplan-Meier survival curves showed that the two groups in the testing set were significantly different ($P < 0.001$) (Figure 3D). In the testing set, the 1-, 3-, and 5-year survival rates of the high PRPscore group were 81.6%, 59.6%, and 44.8%, respectively, while the 1-, 3-, and 5-year survival rates of the low PRPscore group were 95.5%, 91.2%, and 79%. The risk score distribution, survival statuses, and risk protein expression heatmap of the testing set are shown in Figure 3E. And the ROC analysis showed that the 1-, 3-, and 5-year AUC values were 0.774, 0.773, and 0.744 in the testing set (Figure 3F). The significant survival differences between high and low PRPscore groups were also verified in the whole TCGA dataset ($P < 0.001$) (Figure 3G). In the whole TCGA dataset, the 1-, 3-, and 5-year survival rates of the high PRPscore group were 80.3%, 56.8%, and 39.2%, respectively. In the low PRPscore group, the 1-, 3-, and 5-year survival rates were 97.2%, 90.4%, and 79.3%, respectively. The risk curve showed similar characters to that of the training set and the

testing set (Figure 3H) and the ROC analysis showed that the 1-, 3-, and 5-year AUC values were 0.818, 0.784, and 0.769, respectively (Figure 3I).

Next, we explored the risk-stratified analysis. To ensure the sum of the number of subgroups and prevent statistical errors, we examined the clinical data of each sample in detail and excluded the samples with incomplete clinical information from our analysis. After the examination, we excluded two samples with unknown grade information and two samples with unclear AJCC-stage information (Table 1). A total of 441 samples with complete clinical information were included in the stratified analysis. All ccRCC patients were divided into several subgroups: age ≤ 65 subgroup ($n = 293$), age > 65 subgroup ($n = 148$), female subgroup ($n = 147$), male subgroup ($n = 294$), tumor grade I-II subgroup ($n = 195$), tumor grade III-IV subgroup ($n = 246$), AJCC stage I-II subgroup ($n = 259$), AJCC stage III-IV subgroup ($n = 182$), T stage I-II subgroup ($n = 274$), T stage III-IV subgroup ($n = 167$), M stage M0 subgroup ($n = 368$) and M stage M1 subgroup ($n = 73$). All of these subgroups could be stratified as high risk or low risk based on the PRPscore, and the different PRPscore groups had a significant difference in OS rate ($P < 0.05$), which indicated that the protein signature performed well in the stratified analysis and demonstrated the robustness of prognostic value of the protein signature (Figure 4).



Then, compared with a single protein, the PRPscore shows a more accurate prediction ability (Supplementary Figure 1), and the comparison between these markers can be seen in Supplementary Table 2. Besides, a pan-cancer study of TCGA (Liu et al., 2018) shows that PFI is also of great significance in studying the prognosis of patients. The short-term clinical follow-up interval is conducive to the outcome analysis of aggressive cancer types, which can complement each other with OS and improve the reliability of the follow-up study. So, we studied the difference in PFI outcome between high and low score groups and found that the prognosis of the high score group was worse (log-rank $P < 0.001$) (Supplementary Figure 2), which was consistent with the results of the OS study.

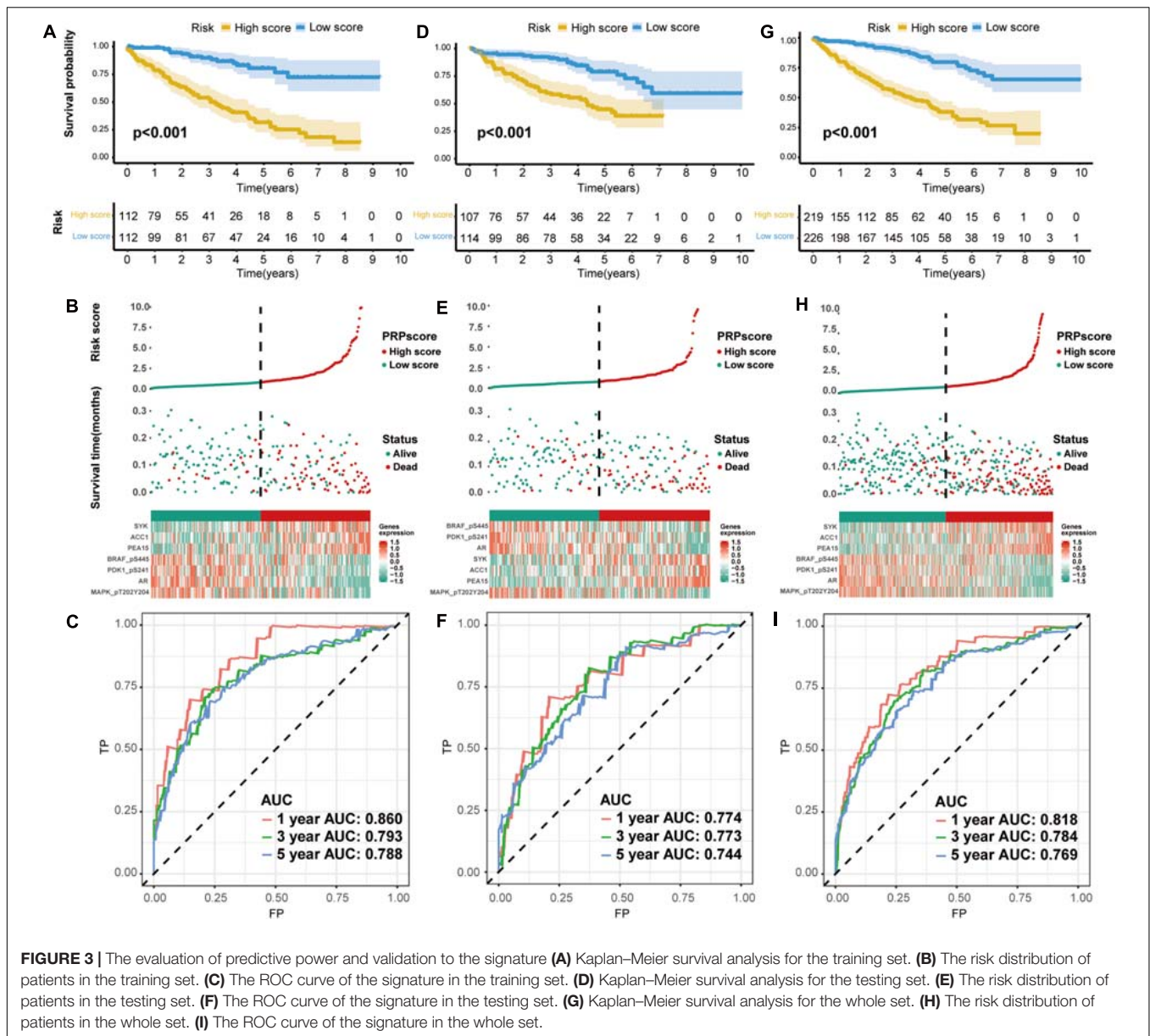
Identification of Independent Prognostic Factors and Construction of Nomogram

Univariate and multivariate Cox regression analyses were performed in the training set (Supplementary Figure 3A), testing set (Supplementary Figure 3B), and whole TCGA

dataset (Supplementary Figure 3C) to explore the independent prognostic factors of ccRCC patients, and the age and PRPscore were identified eventually ($P < 0.05$). To complete our study, we also analyzed the independent prognostic factors for PFI and identified the stage and PRPscore ($P < 0.05$) (Supplementary Figure 3D). To construct a more suitable and accurate tool for clinical practice, the compound nomogram for OS (Figure 5A) and PFI (Figure 5B) were established, which included the independent prognostic factors we found above. The calibration curves corresponding to each nomogram are also drawn separately (OS: Figure 5C, PFI: Figure 5D). And the curves showed that the predicted result did not deviate excessively from the actual result, which means the signature has high accuracy.

Multi-Dimensional Analysis Revealed the Clinical Value of Proteins in the Signature

Firstly, Kaplan-Meier survival analysis was performed using the TCGA database, and the results indicated that all proteins in the signature had a significant effect on the prognosis of



patients with ccRCC (ACCA1: log-rank $P = 2.7796e-04$, AR: log-rank $P = 1.7907e-06$, BRAF: log-rank $P = 4.2255e-02$, MAPK: log-rank $P = 1.1684e-04$, PDK1: log-rank $P = 3.863e-03$, PEA15: log-rank $P = 1.055e-04$, SYK: log-rank $P = 1.5203e-03$) (Figure 6A), and high expression level of ACC1, PEA15 and SYK were related to poor prognosis, the high expression level of AR, MAPK, PDK1 and BRAF were associated with better survival outcome. Secondly, we used the "Datasets" modules in the TCPA database to analyze these seven proteins at the pan-cancer level (Li et al., 2017a) and found that their expression levels were different among different cancer types (Supplementary Figure 4). Thirdly, the coding genes of these proteins were analyzed by mRNA differential expression at the pan-cancer level. The ACCA1, AR, MAPK, PDK1, PEA15, and SYK showed significant differential expression between normal tissues and

tumor tissues ($P < 0.0001$). But BRAF did not show a significant difference (Supplementary Figure 5). Then, the images of IHC staining for these proteins were obtained from the HPA database, and we found that ACCA1 (the gene encoding ACC1), MAPK1, PEA15, and BRAF had strong staining in a high proportion of cells. AR and PDK1 had a medium staining level in a medium proportion of cells, and SYK showed weak staining in a low proportion of cells (Figure 6B).

Next, the clinical correlation analysis showed that the expression of these proteins was related to clinical characteristics in ccRCC patients ($P < 0.05$) (Table 2). MAPK, PEA15, and PDK1 were closely associated with cancer metastasis. The high expression level of MAPK and PDK1 were associated with M0, the survival curves also showed patients with high expression levels had better survival outcomes. And PEA15 exhibited the

opposite features. All the results of the correlation analysis could be found in **Supplementary Figure 6**. Finally, we performed protein coexpression analysis for all seven proteins and identified 56 proteins significantly coexpressed ($P < 0.001$, coexpression score < 0.4). These 56 coexpressed proteins may be key for future clinical management and improving prognosis for ccRCC patients (**Figure 6C**). And the correlation of these proteins in the signature was shown in **Figure 6D**, in which PDK1 and BRAF showed the strongest positive correlation, while ACC1 and MAPK showed the strongest negative correlation.

Different Immune Infiltration, Immunotherapy Response, and Other Characters in High and Low PRPscore Groups

With the help of the PCA algorithm, we found that the protein signature could better categorize patients into two groups than all proteins, thus confirming the rationality of our choice to include these proteins in the prediction risk signature (**Figures 7A,B**). GSEA of these two groups of patients revealed that the glycosaminoglycan biosynthesis chondroitin sulfate pathway, cytokine-cytokine receptor interaction pathway, hematopoietic cell lineage pathway, complement, and coagulation cascade pathway, and prion disease pathway were significantly enriched in the high PRPscore group. In contrast, the fatty acid metabolism pathway, histidine metabolism pathway, PPAR signaling pathway, valine leucine, and isoleucine degradation pathway, and propanoate metabolism pathway were especially enriched in the low PRPscore group (**Figure 7C**). Besides, the

landscape of immune cell infiltration of these two groups as shown in **Figure 7D**, and the content of plasma cells, CD8 T cells, CD4 memory-activated T cells, follicular helper T cells, regulatory T cells (Tregs), M0 macrophages, M1 macrophages, M2 macrophages, resting dendritic cells, activated dendritic cells and resting mast cells et al. were significantly different between the two groups ($P < 0.05$) (**Figure 7E**). By analyzing the correlation between PRPscore and clinical features, we found the PRPscore was significantly increased in high-grade tumor patients and patients with distant metastasis ($P < 0.05$) (**Figure 7F**), which suggested that the score might be related to poor prognosis. It was also consistent with our previous survival analysis and multivariate Cox regression analysis.

Then, we further assessed the potential immunotherapy response in each patient by the TIDE algorithm and observed that patients in the low PRPscore group (43.56%, 98/225) were more likely to respond to immunotherapy than patients in the high PRPscore group (29.77%, 64/215) (**Figure 7G**). The immunotherapy response has a significant correlation with the PRPscore ($P = 0.003$). Subsequently, we based on previous studies to analyze the potential response of anti-CTLA-4 and anti-PD-1 therapy, however, we could not find the difference in response between these two groups by comparison (**Figure 7H**).

Identification of PRP-Based Subgroups and Survival Analysis

The protein signature was proven to be a robust tool for predicting the prognosis of ccRCC patients. To further explore the importance of this protein signature in ccRCC, consensus clustering analysis was performed to identify subgroups based on these proteins' expression (**Figures 8A–C**). The optimal number of categories was limited by many factors. First, the area of the cumulative distribution function curve needed to be stable. Second, the correlation between the categories could not be too strong. Ultimately, we determined that 6 was the best number of subgroups. Then, we performed a Kaplan-Meier survival analysis. The results revealed significant differences between these subgroups (log-rank $P < 0.0001$), and cluster 2 had the worst prognosis, while cluster 4 and cluster 6 had better outcomes (**Figure 8D**). Clinical characteristics in these subgroups were also analyzed; cluster 2 had a higher percentage of patients with the advanced grade, clusters 4 and 6 had a higher percentage of patients with lower AJCC stage, clusters 2 and 3 had a higher rate of patients with higher T stage, cluster 2 had a higher rate of patients with higher N stage, and cluster 6 had the lowest rate of metastasis (**Figure 8E**). The exploration of subgroups also indicated the utility of the protein signature we constructed in ccRCC patients.

DISCUSSION

Clear cell renal cell carcinoma is the most common type of renal cancer. The prognosis of patients with ccRCC is worse than that of those with kidney renal papillary cell carcinoma or kidney chromophobe, other types of renal cell cancer (Ljungberg et al., 2019). Therefore, detection and treatment in the early

TABLE 1 | Clinical information of ccRCC patients with complete protein expression data.

Clinical characters	Variable	Number (total = 441)	Percentages (%)
Age	< = 65	293	66.44%
	> 65	148	33.56%
Gender	FEMALE	147	33.33%
	MALE	294	66.67%
Tumor grade	G1	9	2.04%
	G2	186	42.18%
	G3	174	39.46%
	G4	72	16.33%
AJCC-stage	I	215	48.75%
	II	44	9.98%
	III	107	24.26%
	IV	75	17.01%
T-stage	T1	220	49.89%
	T2	54	12.24%
	T3	157	35.60%
	T4	10	2.27%
N-stage	N0	202	45.80%
	N1	13	2.95%
	NX	226	51.25%
M-stage	M0	368	83.45%
	M1	73	16.55%

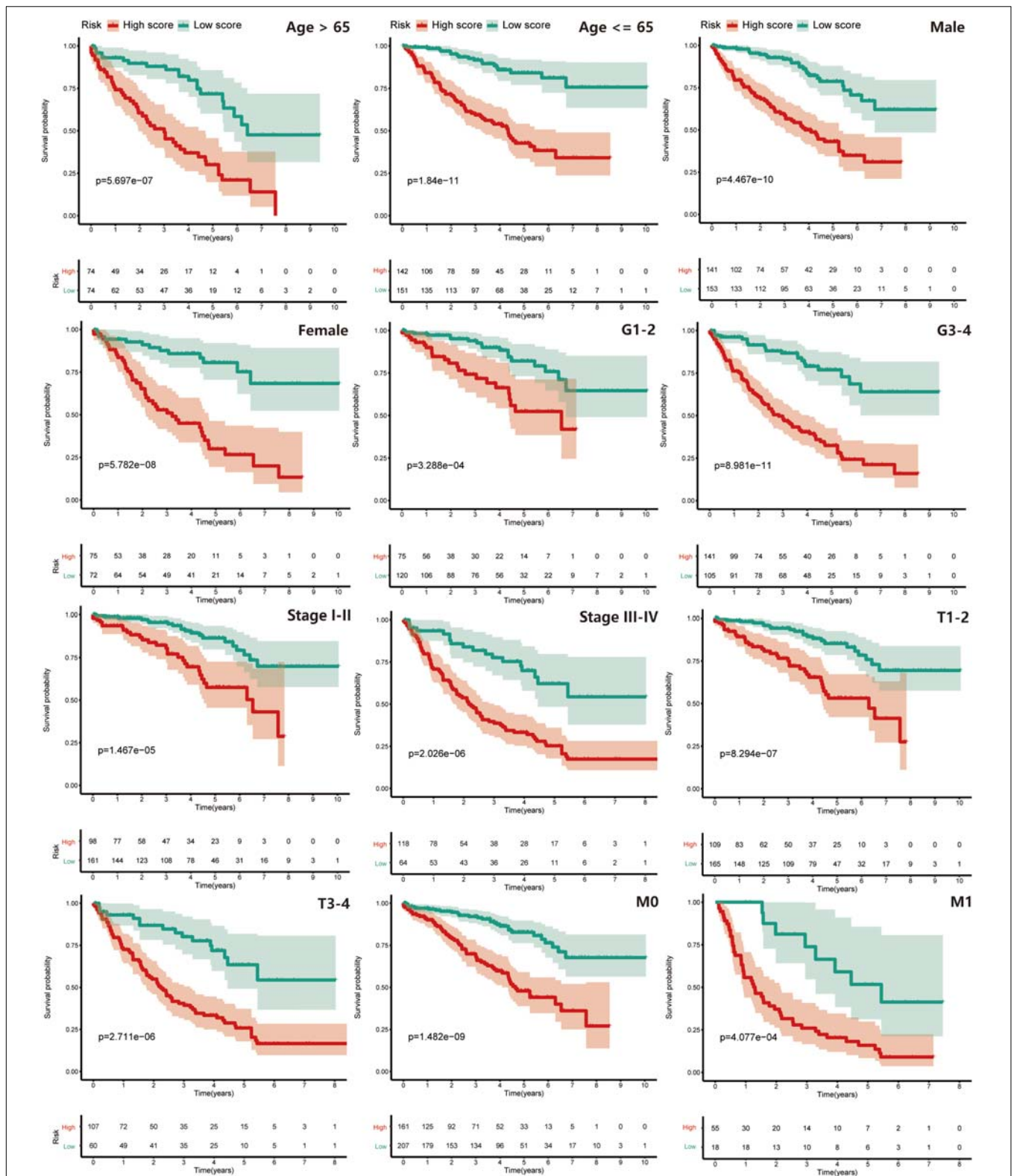
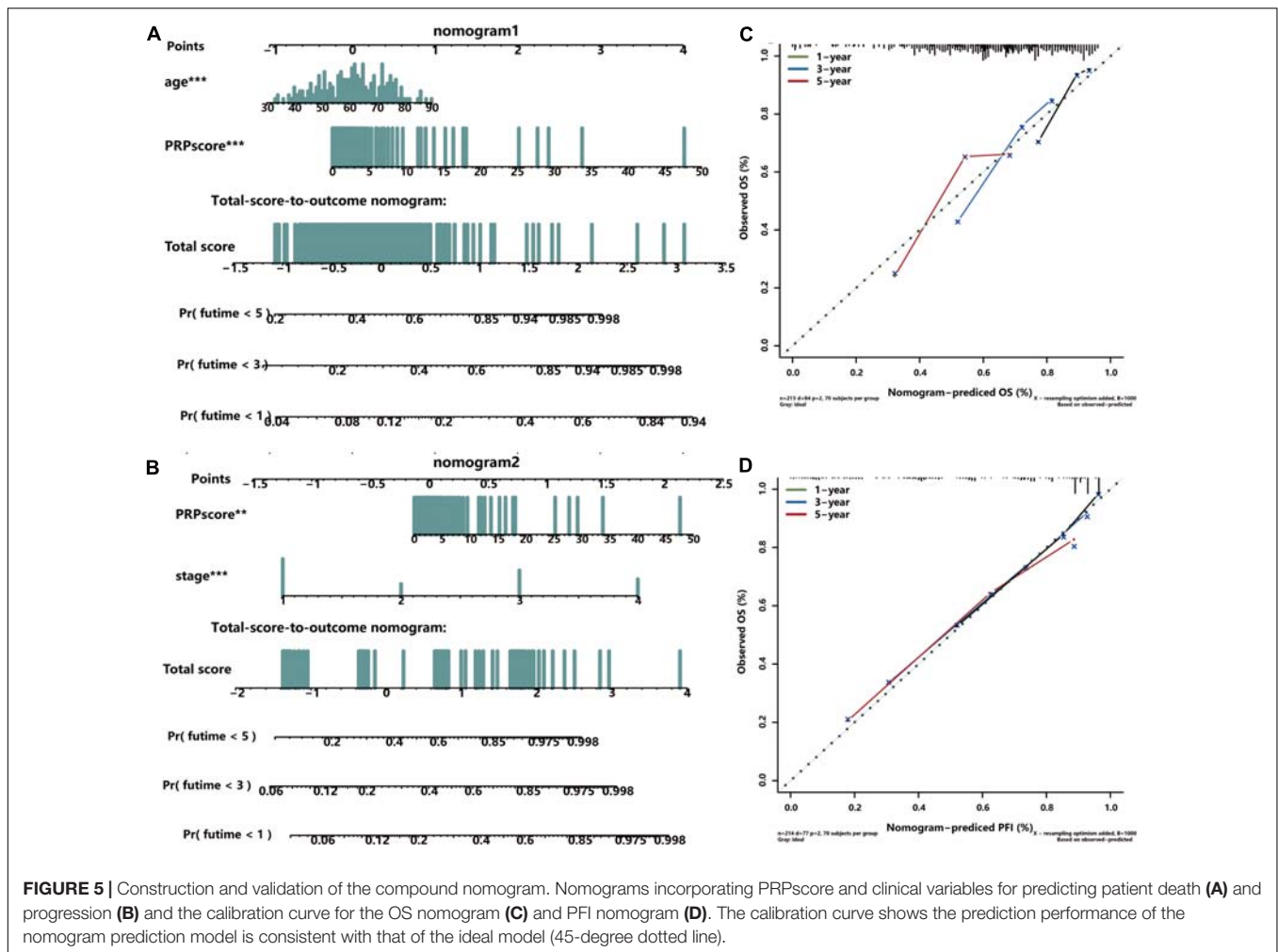


FIGURE 4 | The stratified analysis for the signature. Patients were stratified into Age > 65 subgroup, Age ≤ 65 subgroup, Male subgroup, Female subgroup, G1-2 subgroup, G3-4 subgroup, Stage I-II subgroup, Stage III-IV subgroup, T1-2 subgroup, T3-4 subgroup, M0 subgroup, and M1 subgroup to verify the clinical value of the signature, Kaplan-Meier survival analysis was performed and $p < 0.05$ was the cut-off value. Red means the high PRPscore patients and green means the low PRPscore patients.



stage are vital for patients with ccRCC. To achieve this goal, researchers have identified many potential biomarkers and prediction signatures, especially those related to immunity (Wan et al., 2019). However, it is rare to find a signature based on protein expression. Therefore, we constructed a novel protein signature with data from the TCPA database. Patients could be stratified according to the PRPscore calculated by the signature, and the score was verified as an independent prognosis factor of ccRCC patients, and the predictive capacity was proven to be very accurate.

Undeniably, compared with the recent explosive growth of next-generation sequencing data at the DNA and RNA levels, large-scale cancer proteomics data are relatively limited (Li et al., 2017a). Even in our study, we could find that the number of samples using TCGA data analysis alone is significantly larger than that of the merged TCPA and TCGA databases. However, we should recognize that most genes eventually need to play a role at the protein level, and the correlation between DNA and RNA levels and protein levels is low, especially with post-translationally modified proteins, so it is necessary to evaluate protein levels directly (Li et al., 2017a). Although the progress of next-generation sequencing (NGS) makes the

analysis of RNA level more and more popular, the application of NGS is limited because of its cost, repeatability, and data analysis. In comparison, tumor protein biomarkers are more practical and reliable (Russell et al., 2019; Sun N. et al., 2020). In addition, protein signature has been studied in bladder urothelial carcinoma (Luo and Zhang, 2020), colorectal cancer (Yue et al., 2020), hepatocellular carcinoma (Wu and Yang, 2020), head and neck squamous cell carcinoma (Wu et al., 2020), lung squamous cell carcinoma (Fang et al., 2020), esophageal squamous cell carcinoma (Wang et al., 2018), and pancreatic ductal adenocarcinoma (Burki, 2017), but in patients with ccRCC, related studies are rare. Therefore, our research will be enlightening and valuable to other researchers.

To further clarify the advantages of our signature and the direction of further verification and exploration in the future. We searched several tumor markers of ccRCC recognized by a wide range of scholars and made a brief review to compare them with the protein signature we constructed (Supplementary Table 3). We found that most of the widely recognized tumor markers of ccRCC have studied the mechanism in-depth (Gossage et al., 2015; Zhang et al., 2018), and many targeted drugs have been developed or even approved for use

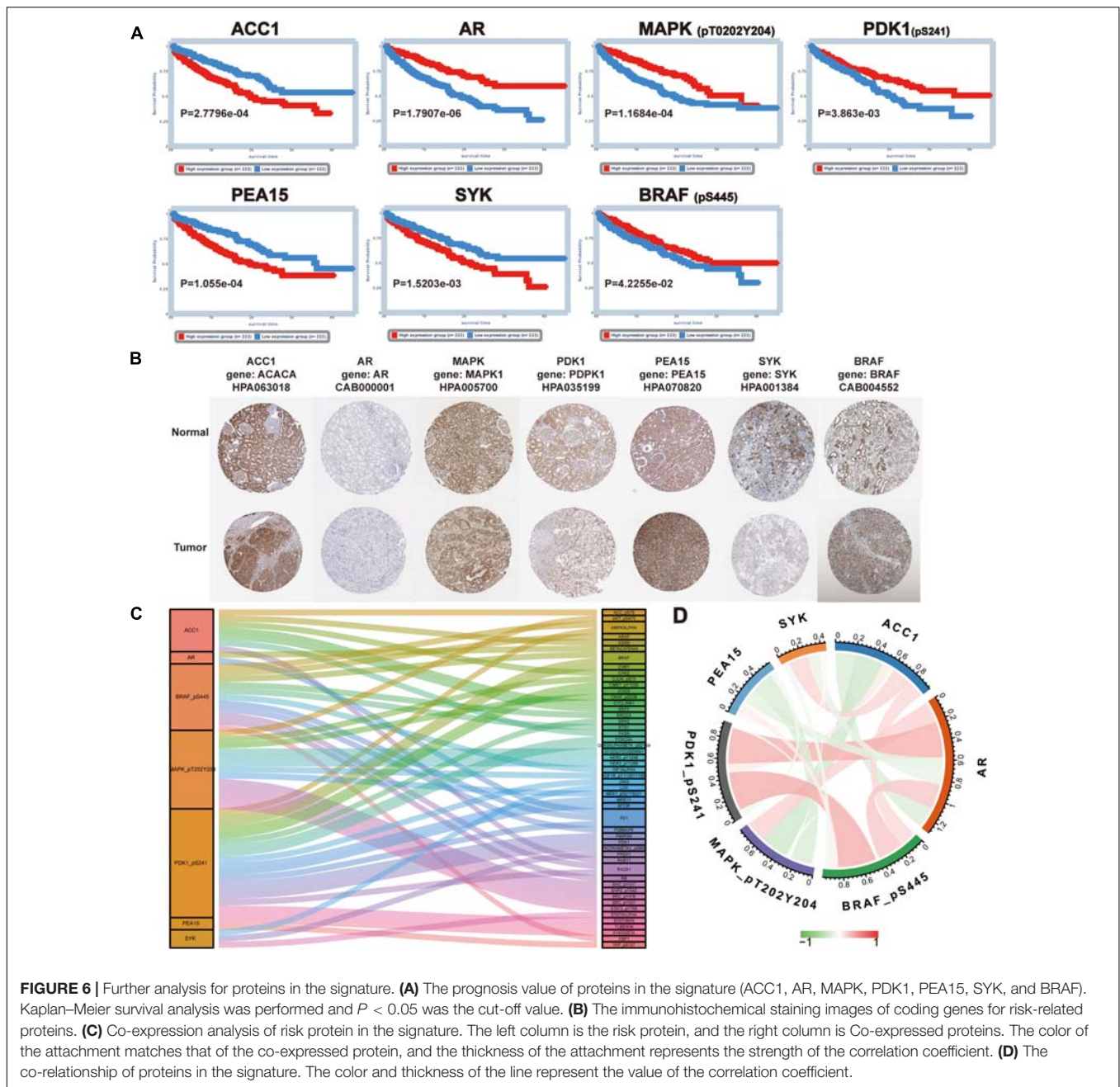
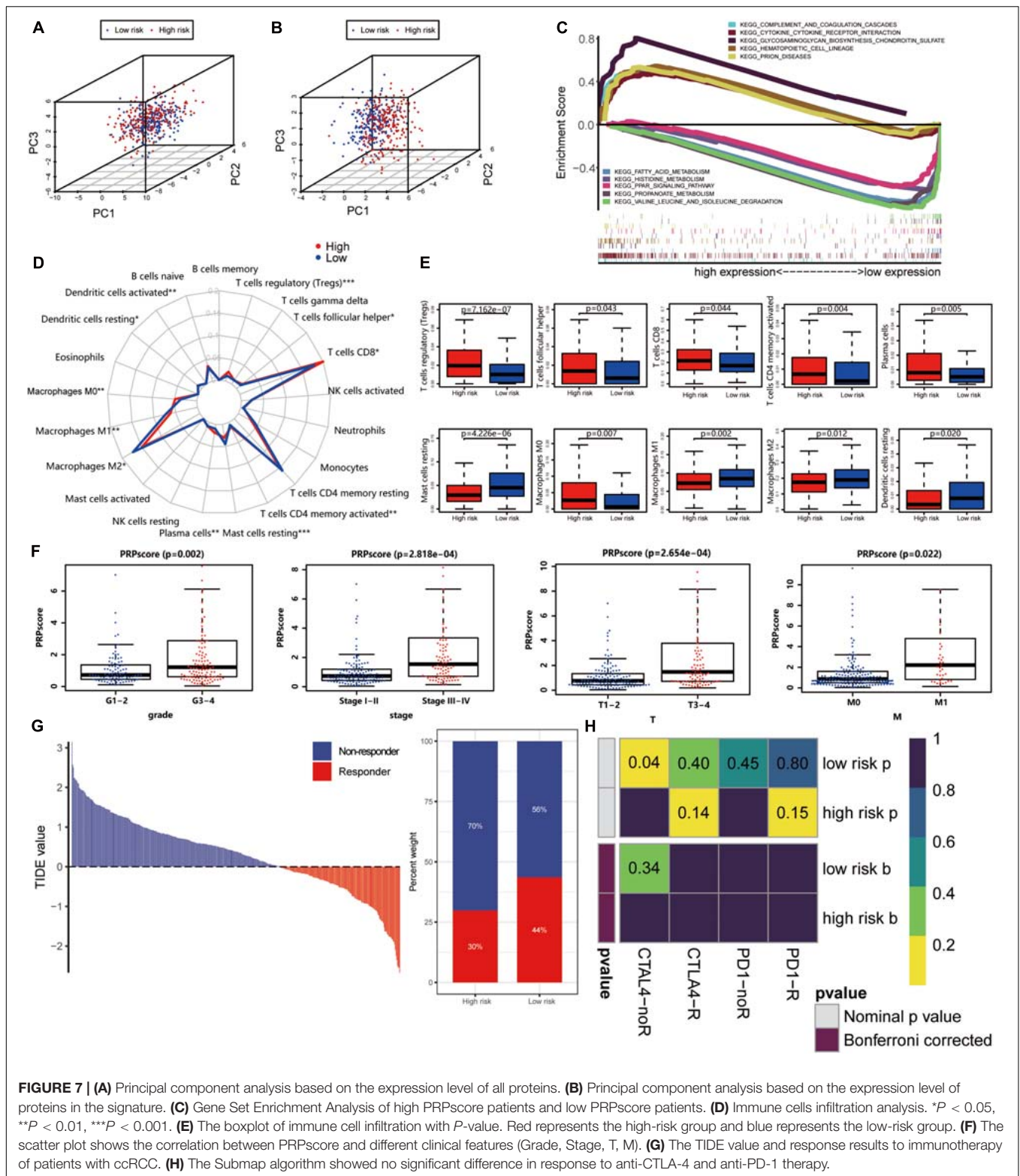


FIGURE 6 | Further analysis for proteins in the signature. **(A)** The prognosis value of proteins in the signature (ACC1, AR, MAPK, PDK1, PEA15, SYK, and BRAF). Kaplan–Meier survival analysis was performed and $P < 0.05$ was the cut-off value. **(B)** The immunohistochemical staining images of coding genes for risk-related proteins. **(C)** Co-expression analysis of risk protein in the signature. The left column is the risk protein, and the right column is Co-expressed proteins. The color of the attachment matches that of the co-expressed protein, and the thickness of the attachment represents the strength of the correlation coefficient. **(D)** The co-relationship of proteins in the signature. The color and thickness of the line represent the value of the correlation coefficient.

TABLE 2 | The relationship between protein expression and clinicopathological characteristics.

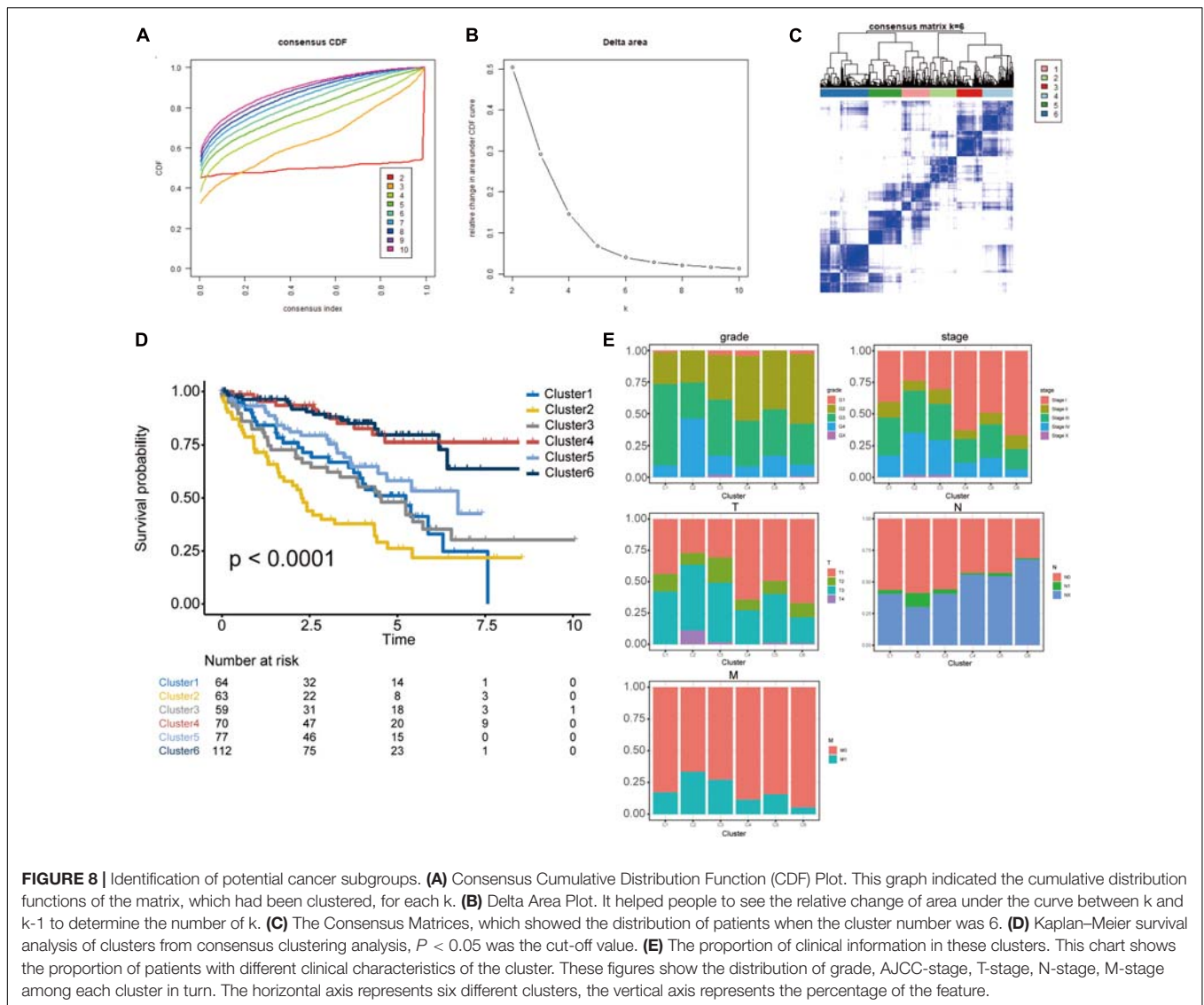
ID	Age (> 65 ≤ 65)	Gender (male female)	Grade (GI&II GIII&IV)	Stage (I& II III&IV)	T (TI&II TIII&IV)	N (NO N1)	M (MO M1)
ACC1	0.194(0.847)	-0.709 (0.479)	-3.054 (0.003)	-2.98 (0.003)	-3.242 (0.001)	-1.493 (0.159)	-0.923 (0.360)
AR	0.15 (0.881)	-2.677 (0.008)	2.661 (0.008)	3.404 (8.019e-04)	3.891 (1.403e-04)	2.791 (0.015)	1.628 (0.110)
MAPK	0.336 (0.738)	-0.195 (0.846)	2.239 (0.026)	2.051 (0.041)	1.564 (0.120)	2.35 (0.034)	2.572 (0.013)
PDK11	-0.355 (0.723)	-0.77 (0.443)	1.929 (0.055)	3.42 (7.543e-04)	3.933 (1.174e-04)	2.266 (0.042)	2.177 (0.033)
PEA15	-0.131 (0.896)	2.189 (0.030)	-0.993 (0.322)	-3.606 (3.903e-04)	-3.228 (0.001)	-3.165 (0.007)	-3.659 (5.613e-04)
SYK	1.314 (0.191)	-1.012 (0.313)	-1.258 (0.210)	-1.48 (0.140)	-1.589 (0.114)	-1.667 (0.119)	-1.365 (0.177)
BRAF	-0.993 (0.322)	2.401 (0.017)	0.925 (0.356)	1.338 (0.182)	1.209 (0.228)	0.499 (0.625)	1.543 (0.128)

t: *t* value from Student's *t*-test; *p*: *p*-value from Student's *t*-test. Among these seven proteins, the expression of six proteins showed a significant correlation with the clinical characteristics of the patients. But the expression of SYK did not show any significant correlation with clinical characters. The specific value of *t*(*p*) was shown in the table.



(Ricketts et al., 2016; Hsieh et al., 2017). However, it is difficult to distinguish the beneficiaries of some targeted drugs, and some markers as independent prognostic factors are still controversial (Leibovich et al., 2007). Our signature has a good ability to

predict the prognosis and distinguish the people who benefit from immunotherapy and could be used as an independent prognostic factor, so our signature may contribute to the accurate treatment of diseases in the future. In addition to



comparing with a single biomarker, we also compared our study with other previous studies at the signature level, such as the long non-coding RNA signature (Qu et al., 2018; Sun Z. et al., 2020), glucose metabolism-related signature (Wang S. et al., 2020), and immune-related risk signature (Hua et al., 2020), etc. By comparison, we found that the previous scholars' researches are also relatively sufficient, and the modeling method they use is mainly based on the LASSO-Cox algorithm, which is similar to ours. However, most of their studies are limited to the RNA level and have little analysis on the protein level. Meanwhile, their studies on the association between the signature and specific treatment are rare, while our study discussed in detail the relationship between the protein signature and immunotherapy response. Besides, some studies are multi-center and large sample sizes studies (Qu et al., 2018), which makes their result with high confidence. We would learn from them and verify our findings in more independent cohorts.

In this study, we used the LASSO-Cox algorithm to construct the signature and calculated the PRPscore for each patient. Based on the score, we divided the patients into high and low-score groups. In the training set, testing set, and the whole data set, we found there were significant differences in prognosis between the two groups (log-rank $P < 0.05$), and the prognosis in the low score group was better than that in the high score group, which indicated that there was a close relationship between PRPscore and the prognosis of patients and the PRPscore may be a poor prognostic factor. Next, time-dependent ROC revealed the robustness of PRPscore in predicting the prognosis of patients. Generally speaking, $AUC > 0.60$ indicates that the prediction ability of the model is acceptable, and $AUC > 0.75$ indicates that the prediction ability of the model is excellent. However, the accuracy of our signature prediction for 1, 3, and 5 years in three cohorts are almost more than 0.75, which strongly proves its accurate prediction ability. The excellent performance of the signature in the subsequent risk-stratified analysis

once again verified the potential value of this signature and the TIDE algorithm defines the great potential of PRPscore in screening people who benefit from immunotherapy, which further broadens the scope of use of this signature.

Most proteins in our signature have been reported to be connected with the prognosis of patients with cancer. ACC1 (gene: ACACA, acetyl-CoA carboxylase alpha) is a cytosolic enzyme with carboxyltransferase and biotin carboxylase activity involved in *de novo* fatty acid synthesis (Raimondo et al., 2018). It plays an important role in the pathogenesis and development of ccRCC (Han et al., 2017). Our study found that a higher expression level of ACC1 was related to higher tumor grade, AJCC stage, and T stage, which are known to lead to a lower OS rate of ccRCC patients. AR (gene: AR, androgen receptor), a transcriptional regulator involved in many cellular functions, has been proven to be strongly associated with cancer development and patient survival (Huang et al., 2017; Wang et al., 2017; Hu et al., 2020). We also found that higher expression of AR was related to lower AJCC stage, T stage, and N stage and higher OS rate. MAPK-pT202Y204 (gene: MAPK1, mitogen-activated protein kinase 1), a potential drug target in ccRCC, has been proven to participate in the regulation of ccRCC and other kinds of cancers (Gupta et al., 2017; Zununi Vahed et al., 2017; Gong et al., 2019). Higher expression of this protein was related to a lower metastasis rate and better clinical outcomes, which indicated that this protein may be protective in ccRCC patients. PDK1-pS241 is produced by the gene PDK1 according to the TCPA database and is overexpressed in many types of cancer (Wang et al., 2019). Previous reports have suggested that PDK1 plays an important role in cancer (Wang and Sun, 2018), and it is often modified by microRNAs or used as a drug target (Zhou et al., 2017; Wang and Sun, 2018; Wang et al., 2019). The utility of PDK1-pS241 was similar to that of MAPK-pT202Y204, showing beneficial effects for ccRCC patients. PEA-15 consists of 130 amino acid residues and has a special role in the regulation of the ERK/MAPK signaling pathway (Sulzmaier et al., 2012; Tang et al., 2019) and the progression of a variety of tumors (Dong et al., 2019; Jiang et al., 2019; Luo et al., 2020). Our study found that high expression of this protein may lead to tumor metastasis and high grade. As a nonreceptor cytoplasmic tyrosine kinase, SYK (spleen tyrosine kinase) is a key mediator in a variety of inflammatory cell and immune signaling pathways and has been proven to be a useful drug target for many cancers (Shinde et al., 2019; Cremer and Stegmaier, 2020; Yang et al., 2020). BRAF-pS445 is encoded by BRAF, a famous oncogene (Post et al., 2020), and its mechanism has been explored by many studies (Maruschke et al., 2018; Xue et al., 2018). By reviewing previous studies, we found that the function of AR in ccRCC is still controversial, and some researchers have reached conclusions that conflict with ours (Chang et al., 2014). There are some possible reasons for these differences. First, the total dataset in our study only included 445 patients, a relatively small sample size, which may lead to incorrect results. Second, this protein may play a different role in different stages of ccRCC, but the detailed mechanism requires further exploration.

Although the signature we constructed has sufficient accuracy, our study still has some limitations. For example, more independent cohorts with a higher number of samples should be used to verify the prognostic power of this signature. And the interactions and mechanisms between these proteins should be evaluated by *in vitro* and *in vivo* experiments.

CONCLUSION

Our study constructed a novel prognosis-related protein signature to predict the prognosis of patients with ccRCC based on TCPA and TCGA data. Through the PRPscore calculated by the signature, patients could be divided into high and low PRPscore groups. The low score group has a better prognosis and could benefit from immunotherapy. Moreover, our signature also provided many potential protein biomarkers for the improvement of treatment and prognosis in patients with ccRCC. However, further experiments and validation in more independent cohorts are necessary to verify the conclusions of this research.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

GC and TX designed the study and drafted the manuscript. GZ and SL analyzed the data. HN and MZ supervised the research. All authors read and approved the final manuscript.

FUNDING

This work was supported by the National Natural Science Foundation of China under grant numbers 81772713, 81472411, 81981260351, and 81972378; Taishan Scholar Program of Shandong Province under grant number tsqn20161077; Natural Science Foundation of Shandong Province under grant number ZR2016HQ18; and Key Research and Development Program of Shandong Province under grant number 2018GSF118197.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmolb.2021.623120/full#supplementary-material>

Supplementary Figure 1 | The ROC curve of PRPscore and protein in the signature predicting the prognosis of patients.

Supplementary Figure 2 | The prognosis outcome of PFI between the high PRPscore group and low PRPscore group had a significant difference based on

the Kaplan–Meier survival analysis ($p = 1.05e-09$). The yellow represented the high PRPscore group, the blue represented the low PRPscore group, $p < 0.05$ was the cut-off value.

Supplementary Figure 3 | The univariate Cox regression analysis and multivariate cox regression analysis of PRPscore and other clinical characters. (A) The training set (B) The testing set (C) The whole data set. (D) The independent prognosis analysis of patients with PFI outcome. The result of univariate cox analysis was shown in the left column and the right column represented the result of multivariate cox analysis. $P < 0.05$ was the cut-off value.

Supplementary Figure 4 | The expression level of protein at the pan-cancer level. The horizontal axis represents different types of tumors, and the vertical axis

refers to RPPA protein abundances. The name of each protein is shown on the left side of the picture.

Supplementary Figure 5 | The expression level of protein signature's coding gene at the pan-cancer level. The horizontal axis represents different types of tumors, and the vertical axis refers to the expression level of genes. The name of each gene is shown on the left side of the picture. The blue represents normal tissue, the yellow represents tumor tissue. ****: $p < 0.0001$, ***: $p < 0.001$, **: $p < 0.01$, *: $p < 0.05$, ns: $p > 0.05$.

Supplementary Figure 6 | The relationship between the expression level of proteins in the signature and clinical characters of ccRCC patients. $P < 0.05$ was the cut-off value.

REFERENCES

Aggen, D. H., Drake, C. G., and Rini, B. I. (2020). Targeting PD-1 or PD-L1 in metastatic kidney cancer – combination therapy in the first line setting. *Clin. Cancer Res.* 122, 1175–1184. doi: 10.1158/1078-0432.Ccr-19-3323

Bartelt, A., and Widenmaier, S. B. (2019). Proteostasis in thermogenesis and obesity. *Biol. Chem.* 401, 1019–1030. doi: 10.1515/hsz-2019-0427

Burki, T. (2017). Protein biomarker for pancreatic ductal adenocarcinoma. *Lancet Oncol.* 18:e511. doi: 10.1016/S1470-2045(17)30564-8

Chang, C., Lee, S. O., Yeh, S., and Chang, T. M. (2014). Androgen receptor (AR) differential roles in hormone-related tumors including prostate, bladder, kidney, lung, breast and liver. *Oncogene* 33, 3225–3234. doi: 10.1038/onc.2013.274

Chen, M. M., Li, J., Wang, Y., Akbani, R., Lu, Y., Mills, G. B., et al. (2019). TCGA v3.0: an integrative platform to explore the pan-cancer analysis of functional proteomic data. *Mol. Cell Proteom.* 18(8 Suppl. 1), S15–S25.

Cremer, A., and Stegmaier, K. (2020). Targeting DUBs to degrade oncogenic proteins. *Br. J. Cancer* 122, 1121–1123. doi: 10.1038/s41416-020-0728-7

Dong, F., Yang, Q., Wu, Z., Hu, X., Shi, D., Feng, M., et al. (2019). Identification of survival-related predictors in hepatocellular carcinoma through integrated genomic, transcriptomic, and proteomic analyses. *Biomed. Pharmacother.* 114:108856. doi: 10.1016/j.biopha.2019.108856

Fang, X., Liu, X., Weng, C., Wu, Y., Li, B., Mao, H., et al. (2020). Construction and validation of a protein prognostic model for lung squamous cell carcinoma. *Int. J. Med. Sci.* 17, 2718–2727. doi: 10.7150/ijms.47224

Fendler, A., Bauer, D., Busch, J., Jung, K., Wulf-Goldenberg, A., Kunz, S., et al. (2020). Inhibiting WNT and NOTCH in renal cancer stem cells and the implications for human patients. *Nat. Commun.* 11:929. doi: 10.1038/s41467-020-14700-7

Gong, D., Zhang, J., Chen, Y., Xu, Y., Ma, J., Hu, G., et al. (2019). The m(6)A-suppressed P2RX6 activation promotes renal cancer cells migration and invasion through ATP-induced Ca(2+) influx modulating ERK1/2 phosphorylation and MMP9 signaling pathway. *J. Exp. Clin. Cancer Res.* 38:233.

Gossage, L., Eisen, T., and Maher, E. R. (2015). VHL, the story of a tumour suppressor gene. *Nat. Rev. Cancer* 15, 55–64. doi: 10.1038/nrc3844

Gulati, S., and Vaishampayan, U. (2020). Current state of systemic therapies for advanced renal cell carcinoma. *Curr. Oncol. Rep.* 22:26. doi: 10.1007/s11912-020-0892-1

Gupta, P. P., Bastikar, V. A., Kuciauskas, D., Kothari, S. L., Cicenas, J., and Valius, M. (2017). Molecular modeling and structure-based drug discovery approach reveals protein kinases as off-targets for novel anticancer drug RH1. *Med. Oncol.* 34:176.

Han, G., Zhao, W., Song, X., Kwok-Shing Ng, P., Karam, J. A., Jonasch, E., et al. (2017). Unique protein expression signatures of survival time in kidney renal clear cell carcinoma through a pan-cancer screening. *BMC Genom.* 18(Suppl. 6):678. doi: 10.1186/s12864-017-4026-6

Hennessy, B. T., Lu, Y., Gonzalez-Angulo, A. M., Carey, M. S., Myhre, S., Ju, Z., et al. (2010). A Technical assessment of the utility of reverse phase protein arrays for the study of the functional proteome in non-microdissected human breast cancers. *Clin. Proteom.* 6, 129–151. doi: 10.1007/s12014-010-9055-y

Hoadley, K. A., Yau, C., Hinoue, T., Wolf, D. M., Lazar, A. J., Drill, E., et al. (2018). Cell-of-origin patterns dominate the molecular classification of 10,000 tumors from 33 types of cancer. *Cell* 173, 291.e–304.e. doi: 10.1016/j.cell.2018.03.022

Hoshida, Y., Brunet, J.-P., Tamayo, P., Golub, T. R., and Mesirov, J. P. (2007). Subclass mapping: identifying common subtypes in independent disease data sets. *PLoS One* 2:e1195. doi: 10.1371/journal.pone.0001195

Hsieh, J. J., Purdue, M. P., Signoretti, S., Swanton, C., Albiges, L., Schmidinger, M., et al. (2017). Renal cell carcinoma. *Nat. Rev. Dis. Primers* 3:17009. doi: 10.1038/nrdp.2017.9

Hu, C., Fang, D., Xu, H., Wang, Q., and Xia, H. (2020). The androgen receptor expression and association with patient's survival in different cancers. *Genomics* 112, 1926–1940. doi: 10.1016/j.ygeno.2019.11.005

Hua, X., Chen, J., Su, Y., and Liang, C. (2020). Identification of an immune-related risk signature for predicting prognosis in clear cell renal cell carcinoma. *Aging* 12, 2302–2332. doi: 10.18632/aging.102746

Huang, Q., Sun, Y., Ma, X., Gao, Y., Li, X., Niu, Y., et al. (2017). Androgen receptor increases hematogenous metastasis yet decreases lymphatic metastasis of renal cell carcinoma. *Nat. Commun.* 8:918.

Jiang, P., Gu, S., Pan, D., Fu, J., Sahu, A., Hu, X., et al. (2018). Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nat. Med.* 24, 1550–1558. doi: 10.1038/s41591-018-0136-1

Jiang, X., Zhang, C., Li, W., Jiang, D., Wei, Z., Lv, M., et al. (2019). PEA15 contributes to the clinicopathology and AKT-regulated cisplatin resistance in gastric cancer. *Oncol. Rep.* 41, 1949–1959. doi: 10.3892/or.2018.6934

Leibovich, B. C., Sheinin, Y., Lohse, C. M., Thompson, R. H., Chevillet, J. C., Zavada, J., et al. (2007). Carbonic anhydrase IX is not an independent predictor of outcome for patients with clear cell renal cell carcinoma. *J. Clin. Oncol.* 25, 4757–4764. doi: 10.1200/jco.2007.12.1087

Li, J., Akbani, R., Zhao, W., Lu, Y., Weinstein, J. N., Mills, G. B., et al. (2017a). Explore, visualize, and analyze functional cancer proteomic data using the cancer proteome Atlas. *Cancer Res.* 77, e51–e54. doi: 10.1158/0008-5472.CAN-17-0369

Li, J., Han, L., Roebuck, P., Diao, L., Liu, L., Yuan, Y., et al. (2015). TANRIC: an interactive open platform to explore the function of lncRNAs in cancer. *Cancer Res.* 75, 3728–3737. doi: 10.1158/0008-5472.CAN-15-0273

Li, J., Lu, Y., Akbani, R., Ju, Z., Roebuck, P. L., Liu, W., et al. (2013). TCGA: a resource for cancer functional proteomics data. *Nat. Methods* 10, 1046–1047. doi: 10.1038/nmeth.2650

Li, J., Zhao, W., Akbani, R., Liu, W., Ju, Z., Ling, S., et al. (2017b). Characterization of human cancer cell lines by reverse-phase protein arrays. *Cancer Cell* 31, 225–239. doi: 10.1016/j.ccell.2017.01.005

Linehan, W. M., and Ricketts, C. J. (2019). The Cancer genome Atlas of renal cell carcinoma: findings and clinical implications. *Nat. Rev. Urol.* 16, 539–552. doi: 10.1038/s41585-019-0211-5

Liu, J., Lichtenberg, T., Hoadley, K. A., Poisson, L. M., Lazar, A. J., Cherniack, A. D., et al. (2018). An integrated TCGA Pan-cancer clinical data resource to drive high-quality survival outcome analytics. *Cell* 173, 400–416.e11. doi: 10.1016/j.cell.2018.02.052

Ljungberg, B., Albiges, L., Abu-Ghanem, Y., Bensalah, K., Dabestani, S., Fernandez-Pello, S., et al. (2019). European association of urology guidelines on renal cell carcinoma: the 2019 update. *Eur. Urol.* 75, 799–810. doi: 10.1016/j.eururo.2019.02.011

- Luo, Q., and Zhang, X. (2020). Construction of protein-related risk score model in Bladder Urothelial carcinoma. *Biomed. Res. Int.* 2020: 7147824. doi: 10.1155/2020/7147824
- Luo, Y., Fang, C., Jin, L., Ding, H., Lyu, Y., and Ni, G. (2020). The microRNA212 regulated PEA15 promotes ovarian cancer progression by inhibiting of apoptosis. *J. Cancer* 11, 1424–1435. doi: 10.7150/jca.32886
- Maruschke, M., Koczan, D., Ziems, B., and Hakenberg, O. W. (2018). Copy number alterations with prognostic potential in clear cell renal cell carcinoma. *Urol. Int.* 101, 417–424.
- Menschaert, G., and Fenyö, D. (2017). Proteogenomics from a bioinformatics angle: a growing field. *Mass Spectrom. Rev.* 36, 584–599. doi: 10.1002/mas.21483
- Menschaert, G., Vandekerckhove, T. T. M., Baggerman, G., Schoofs, L., Luyten, W., and Van Crielinge, W. (2010). Peptidomics coming of age: a review of contributions from a bioinformatics angle. *J. Proteome Res.* 9, 2051–2061. doi: 10.1021/pr900929m
- Monti, S., Tamayo, P., Mesirov, J., and Golub, T. (2003). Consensus clustering: a resampling-based method for class discovery and visualization of gene expression microarray data. *Mach. Learn.* 52, 91–118. doi: 10.1023/A:1023949509487
- Nishizuka, S., Charboneau, L., Young, L., Major, S., Reinhold, W. C., Waltham, M., et al. (2003). Proteomic profiling of the NCI-60 cancer cell lines using new high-density reverse-phase lysate microarrays. *Proc. Natl. Acad. Sci. U.S.A.* 100, 14229–14234. doi: 10.1073/pnas.2331323100
- Post, J. B., Roodhart, J. M. L., and Snippet, H. J. G. (2020). Colorectal cancer modeling with organoids: discriminating between oncogenic RAS and BRAF variants. *Trends Cancer* 6, 111–129. doi: 10.1016/j.trecan.2019.12.005
- Qu, L., Wang, Z.-L., Chen, Q., Li, Y.-M., He, H.-W., Hsieh, J. J., et al. (2018). Prognostic value of a long non-coding RNA signature in localized clear cell renal cell carcinoma. *Eur. Urol.* 74, 756–763. doi: 10.1016/j.eururo.2018.07.032
- Raimondo, S., Saieva, L., Cristaldi, M., Monteleone, F., Fontana, S., and Alessandro, R. (2018). Label-free quantitative proteomic profiling of colon cancer cells identifies acetyl-CoA carboxylase alpha as antitumor target of *Citrus limon*-derived nanovesicles. *J. Proteom.* 173, 1–11. doi: 10.1016/j.jprot.2017.11.017
- Ricketts, C. J., Crooks, D. R., and Linehan, W. M. (2016). Targeting HIF2 α in clear-cell renal cell carcinoma. *Cancer Cell* 30, 515–517. doi: 10.1016/j.ccell.2016.09.016
- Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., et al. (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinform.* 12:77. doi: 10.1186/1471-2105-12-77
- Roh, W., Chen, P.-L., Reuben, A., Spencer, C. N., Prieto, P. A., Miller, J. P., et al. (2017). Integrated molecular analysis of tumor biopsies on sequential CTLA-4 and PD-1 blockade reveals markers of response and resistance. *Sci. Transl. Med.* 9:eah3560. doi: 10.1126/scitranslmed.ah3560
- Russell, M. R., Graham, C., D'Amato, A., Gentry-Maharaj, A., Ryan, A., Kalsi, J. K., et al. (2019). Diagnosis of epithelial ovarian cancer using a combined protein biomarker panel. *Br. J. Cancer* 121, 483–489. doi: 10.1038/s41416-019-0544-0
- Sheehan, K. M., Calvert, V. S., Kay, E. W., Lu, Y., Fishman, D., Espina, V., et al. (2005). Use of reverse phase protein microarrays and reference standard development for molecular network analysis of metastatic ovarian carcinoma. *Mol. Cell Proteom.* 4, 346–355. doi: 10.1074/mcp.t500003-mcp200
- Shinde, A., Hardy, S. D., Kim, D., Akhand, S. S., Jolly, M. K., Wang, W. H., et al. (2019). Spleen tyrosine kinase-mediated autophagy is required for epithelial-mesenchymal plasticity and metastasis in breast cancer. *Cancer Res.* 79, 1831–1843. doi: 10.1158/0008-5472.CAN-18-2636
- Siegel, R. L., Miller, K. D., and Jemal, A. (2020). Cancer statistics, 2020. *CA Cancer J. Clin.* 70, 7–30. doi: 10.3322/caac.21590
- Spurrier, B., Ramalingam, S., and Nishizuka, S. (2008). Reverse-phase protein lysate microarrays for cell signaling analysis. *Nat. Protoc.* 3, 1796–1808. doi: 10.1038/nprot.2008.179
- Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U.S.A.* 102, 15545–15550. doi: 10.1073/pnas.0506580102
- Sulzmaier, F. J., Valmiki, M. K., Nelson, D. A., Caliva, M. J., Geerts, D., Matter, M. L., et al. (2012). PEA-15 potentiates H-Ras-mediated epithelial cell transformation through phospholipase D. *Oncogene* 31, 3547–3560. doi: 10.1038/ncr.2011.514
- Sun, N., Sun, S., Gao, Y., Li, Y., Lu, Z., Yuan, Z., et al. (2020). Utility of isocitrate dehydrogenase 1 as a serum protein biomarker for the early detection of non-small-cell lung cancer: a multicenter in vitro diagnostic clinical trial. *Cancer Sci.* 111, 1739–1749. doi: 10.1111/cas.14387
- Sun, Z., Jing, C., Xiao, C., and Li, T. (2020). Long non-coding RNA profile study identifies an immune-related lncRNA prognostic signature for kidney renal clear cell carcinoma. *Front. Oncol.* 10:1430. doi: 10.3389/fonc.2020.01430
- Tang, B., Liang, W., Liao, Y., Li, Z., Wang, Y., and Yan, C. (2019). PEA15 promotes liver metastasis of colorectal cancer by upregulating the ERK/MAPK signaling pathway. *Oncol. Rep.* 41, 43–56. doi: 10.3892/or.2018.6825
- Tibes, R., Qiu, Y., Lu, Y., Hennessy, B., Andreoff, M., Mills, G. B., et al. (2006). Reverse phase protein array: validation of a novel proteomic technology and utility for analysis of primary leukemia specimens and hematopoietic stem cells. *Mol. Cancer Ther.* 5, 2512–2521. doi: 10.1158/1535-7163.mct-06-0334
- Torresano, L., Nuevo-Tapióles, C., Santacatterina, F., and Cuezva, J. M. (2020). Metabolic reprogramming and disease progression in cancer patients. *Biochim. Biophys. Mol. Basis Dis.* 1866:165721. doi: 10.1016/j.bbdis.2020.165721
- Wan, B., Liu, B., Huang, Y., Yu, G., and Lv, C. (2019). Prognostic value of immune-related genes in clear cell renal cell carcinoma. *Aging (Albany NY)* 11, 11474–11489. doi: 10.18632/aging.10548
- Wang, J., and Sun, X. (2018). MicroRNA-375 inhibits the proliferation, migration and invasion of kidney cancer cells by triggering apoptosis and modulation of PDK1 expression. *Environ. Toxicol. Pharmacol.* 62, 227–233. doi: 10.1016/j.etap.2018.08.002
- Wang, K., Sun, Y., Tao, W., Fei, X., and Chang, C. (2017). Androgen receptor (AR) promotes clear cell renal cell carcinoma (ccRCC) migration and invasion via altering the circHIAT1/miR-195-5p/29a-3p/29c-3p/CDC42 signals. *Cancer Lett.* 394, 1–12. doi: 10.1016/j.canlet.2016.12.036
- Wang, M., Smith, J. S., and Wei, W.-Q. (2018). Tissue protein biomarker candidates to predict progression of esophageal squamous cell carcinoma and precancerous lesions. *Ann. N. Y. Acad. Sci.* 1434, 59–69. doi: 10.1111/nyas.13863
- Wang, S., Zhang, L., Yu, Z., Chai, K., and Chen, J. (2020). Identification of a glucose metabolism-related signature for prediction of clinical prognosis in clear cell renal cell carcinoma. *J. Cancer* 11, 4996–5006. doi: 10.7150/jca.45296
- Wang, X. M., Lu, Y., Song, Y. M., Dong, J., Li, R. Y., Wang, G. L., et al. (2020). Integrative genomic study of Chinese clear cell renal cell carcinoma reveals features associated with thrombus. *Nat. Commun.* 11:739. doi: 10.1038/s41467-020-14601-9
- Wang, Y., He, Y., Bai, H., Dang, Y., Gao, J., and Lv, P. (2019). Phosphoinositide-dependent kinase 1-associated glycolysis is regulated by miR-409-3p in clear cell renal cell carcinoma. *J. Cell Biochem.* 120, 126–134. doi: 10.1002/jcb.27152
- Wu, Z.-H., and Yang, D.-L. (2020). Identification of a protein signature for predicting overall survival of hepatocellular carcinoma: a study based on data mining. *BMC Cancer* 20:720. doi: 10.1186/s12885-020-07229-x
- Wu, Z.-H., Yun, T., and Cheng, Q. (2020). Data mining identifies six proteins that can act as prognostic markers for head and neck squamous cell carcinoma. *Cell Transpl.* 29. doi: 10.1177/0963689720929308
- Xu, W., Atkins, M. B., and McDermott, D. F. (2020). Checkpoint inhibitor immunotherapy in kidney cancer. *Nat. Rev. Urol.* 17, 137–150. doi: 10.1038/s41585-020-0282-3
- Xue, S., Jiang, S. Q., Li, Q. W., Wang, S., Li, J., Yang, S., et al. (2018). Decreased expression of BRAF-activated long non-coding RNA is associated with the proliferation of clear cell renal cell carcinoma. *BMC Urol.* 18:79. doi: 10.1186/s12894-018-0395-7
- Yang, J., Meng, C., Weisberg, E., Case, A., Lamberto, I., Magin, R. S., et al. (2020). Inhibition of the deubiquitinase USP10 induces degradation of SYK. *Br. J. Cancer* 122, 1175–1184. doi: 10.1038/s41416-020-0731-z
- Yue, T., Liu, C., Zhu, J., Huang, Z., Guo, S., Zhang, Y., et al. (2020). Identification of 6 hub proteins and protein risk signature of colorectal cancer. *Biomed. Res. Int.* 2020:6135060. doi: 10.1155/2020/6135060
- Zhang, J., Wu, T., Simon, J., Takada, M., Saito, R., Fan, C., et al. (2018). VHL substrate transcription factor ZHX2 as an oncogenic driver in clear cell renal cell carcinoma. *Science* 361, 290–295. doi: 10.1126/science.aap8411

- Zhou, J., Yun, E. J., Chen, W., Ding, Y., Wu, K., Wang, B., et al. (2017). Targeting 3-phosphoinositide-dependent protein kinase 1 associated with drug-resistant renal cell carcinoma using new oridonin analogs. *Cell Death Dis.* 8:e2701. doi: 10.1038/cddis.2017.121
- Zununi Vahed, S., Barzegari, A., Rahbar Saadat, Y., Goreyshi, A., and Omid, Y. (2017). Leuconostoc mesenteroides-derived anticancer pharmaceuticals hinder inflammation and cell survival in colon cancer cells by modulating NF-kappaB/AKT/PTEN/MAPK pathways. *Biomed. Pharmacother.* 94, 1094–1100. doi: 10.1016/j.biopha.2017.08.033

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Chu, Xu, Zhu, Liu, Niu and Zhang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.