

Integrated analysis of the functions of RNA binding proteins in clear cell renal cell carcinoma

Xiaoliang Hua

Anhui Medical University

Juan Chen

University of Chinese Academy of Science

Shengdong Ge

Anhui Medical University

Haibing Xiao

Anhui Medical University

Li Zhang

Anhui Medical University

Chaozhao Liang (✉ liang_chaozhao@ahmu.edu.cn)

<https://orcid.org/0000-0003-2317-1323>

Primary research

Keywords: clear cell renal cell carcinoma, RNA binding proteins, prognostic model, biomarker, bioinformatics

Posted Date: May 18th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-28762/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: RNA binding proteins (RBPs) dysregulation is involved in the processes of various tumors. However, the roles of RBPs in clear cell renal cell carcinoma (ccRCC) remain poorly understood. Systematic exploration of the roles of RBPs in ccRCC may provide new insights for the treatments of ccRCC.

Methods: Expression data of RBPs was obtained from The Cancer Genome Atlas database. The roles of RBPs in ccRCC were systematically investigated using consensus clustering methods. Differentially expressed RBPs between normal and tumor tissues were obtained. Protein-protein interaction (PPI) network was constructed using “STRING” software. The expression levels of hub genes were validated in The Human Protein Atlas (HPA) database and receiver operating characteristic (ROC) curves were used to evaluate diagnostic value. Univariate and Lasso Cox regression and Kaplan–Meier curves were used to screen the most useful prognostic genes. Multivariate Cox regression was performed to construct a risk score model. The efficiency of the model was evaluated using time-dependent ROC and Kaplan–Meier curves, and validated in E-MTAB-3267 set.

Results: Two clusters were identified based on the expression similarity of RBPs, and the cluster 2 was closely correlated with the malignancy of ccRCC. Several oncogenic pathways, including epithelial mesenchymal transition, G2M checkpoint, KRAS signaling and IL6 JAK STAT3 signaling were enriched in cluster 2. In addition, we obtained 115 differently expressed RBPs in ccRCC, comprising 71 up-regulated and 44 down-regulated ones. Ten hub RBPs with good diagnostic value were obtained from PPI network and validated in HPA database. Ten RBPs were identified as survival-related genes and used to construct a risk score model. The model could be used to stratify patients with different prognosis. We found high-risk patients tended to be advanced stage, high grade, high pathological T staging and could be an independent risk factor for overall survival of ccRCC patients.

Conclusions: We identified ten RBPs with diagnostic value, which might be the potential diagnostic biomarkers for ccRCC. A risk score model was established to stratify patients and could be used as a complementation for clinical factors to guide clinical practice in the future.

Background

Renal cell carcinoma (RCC) is one of the most frequently diagnosed malignancies of the urinary system, and clear cell renal cell carcinoma (ccRCC) is the most common histologic subtype, that accounts for approximately 80–90% RCC[1]. The incidence rate and mortality of ccRCC have ascended steadily over the past few decades[2, 3], posing a great burden on health care systems. At present, the diagnostic methods of ccRCC mainly depend on image examination, histopathological examination and molecular diagnosis. Many molecular biomarkers have been uncovered based on next-generation sequencing technologies, and can be used to guide diagnosis and treatment of ccRCC. CcRCC has been considered as a heterogeneous disease with different biological behaviors and prognosis[4]. Some patients will

experience relapse or metastasis in spite of surgical resection, which is the major cause of death[4, 5]. Therefore, identification of sensitive and specific biomarkers to predict tumor relapse and progression and construction of a more powerful stratification model to guide patients' treatment are urgently needed.

RNA binding proteins (RBPs) are recognized as pleiotropic proteins which can interact with a wide variety of RNAs, including snoRNAs, snRNAs, ncRNAs, mRNAs, miRNAs, tRNAs and rRNAs. They can interact with target RNAs and form ribonucleoprotein complexes, regulating gene expressions at the post-transcriptional level[6]. Previous studies have shown that RBPs play vital roles in regulating RNA metabolism, RNA stability, alternative splicing, modification, location and translation[7]. In ccRCC, RBPs play vital roles in tumorigenesis and tumour progression, and the roles of some RBPs have been revealed. For example, QKI suppresses tumorigenesis and proliferation of ccRCC by regulating the expression of HIF-1 α and inhibiting the activation of Ras-MAPK signaling pathway[8, 9]. PTBP1 promotes proliferation, migration, and invasion of ccRCC by mediating alternative splicing of PKM[10]. CIPP is up-regulated in ccRCC and CIPP deficiency can inhibit proliferation and enhance the chemosensitivity to gemcitabine in ccRCC[11]. RBM38 acts as a tumor suppressor in ccRCC, which can inhibit tumor proliferation, migration and invasion through influencing their target RNA stability[12]. However, most of the RBPs that involved in tumorigenesis and progression of ccRCC remain undetermined. Therefore, a systematic analysis of RBPs will not only help to uncover their potential roles in ccRCC, but also contribute to discover diagnostic and prognostic biomarkers.

Taking into consideration of the important roles of RBPs in ccRCC, we used bioinformatics methods to systematically explore potential biomarkers for ccRCC patients and construct a risk score model to stratify patients. We firstly explored the roles of these RBPs in estimating prognosis and stratification of ccRCC patients using consensus clustering. Then, we obtained differentially expressed RBPs and performed systematic analysis to reveal the potential biomarkers of ccRCC patients and its underlying molecular functions. In order to better guide clinical practice, we used these RBPs with excellent prognostic value to calculate the risk score for each patient, and established a risk score model to predicting survival of ccRCC patients. The model showed excellent performance in stratifying patients and could be used as supplement for clinical factors to guide clinical practice in the future.

Methods

1. Database

Database used in this study included The Cancer Genome Atlas (TCGA) and ArrayExpress database. For TCGA dataset, we downloaded from UCSC Xena (<https://xenabrowser.net/>), and the expression data was performed $\log_2(x + 1)$ transformed RSEM normalization. Inclusion and exclusion criteria was implemented as follows: (1) Duplicate ccRCC samples are deleted. (2) CcRCC patients with follow-up time less than one week were excluded. A total of 72 normal samples and 522 ccRCC samples were included in this study. For ArrayExpress database, we used the expression data and clinical information of E-MTAB-3267 dataset. A total of 53 ccRCC patients were included. All genes were uniformly

normalized for these datasets. For repeated genes, their expression levels were calculated as the mean expressions. The sheet of RBPs was obtained from previous study[6], and a total of 1524 RBP genes were included.

2. Consensus Clustering And Differentially Expressed Analysis

To investigate the functions of RBPs in ccRCC, we performed consensus clustering to determine cluster numbers using the ConsensusClusterPlus R package after we obtained the expressions of RBPs[13]. The cluster numbers were confirmed using principal component analysis (PCA) using the “ggplot2” package. Differentially expressed analysis were performed between ccRCC and normal samples using “limma” package with criterion of $FDR < 0.05$ and $|FC| > 2$.

3. Functional Enrichment Analysis

The Database for Annotation, Visualization and Integrated Discovery platform (<https://david.ncifcrf.gov/>) was used to perform gene ontology (GO) enrichment analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis. $FDR < 0.05$ was set as cutoff value. Moreover, gene set enrichment analysis (GSEA) was also performed to determine the functional mechanisms. The `h.all.v7.1.symbols.gmt` file was selected as the reference gene set file. $FDR < 0.25$ and normalized P value < 0.05 were set as threshold value.

4. Protein-protein Interaction Network Construction

These differentially expressed RBPs were submitted to the STRING database (<https://string-db.org/>) to construct protein-protein interaction (PPI) network. Genes with interaction score of 0.4 were extracted and visualized using Cytoscape v3.7.1 software (<http://cytoscape.org/>). In addition, the key modules were screened from the PPI network with the K-Core value = 4 using the MCODE plug-in in Cytoscape. Hub RBPs were selected using the cytoHubba plug-in using the MNC algorithm.

5. Hub RBPs Validation And Efficacy Evaluation

The expression levels of these hub RBPs were validated in The Human Protein Atlas (HPA) database[14]. The diagnostic efficiency of these hub RBPs was evaluated using receiver operating characteristic (ROC) curves and the area under the curve (AUC) was calculated using “pROC” package[15]. Moreover, the mutation and copy number alteration (CNA) changes of these RBPs were evaluated using segmentation analysis and the GISTIC algorithm in cBioPortal[16].

6. Selection Of Survival-related RBP Genes

To select survival-related RBP genes, we performed univariate Cox regression analysis for these differentially expressed RBPs using the “survival” R package. These survival-related RBP genes were submitted to perform LASSO cox regression analysis to select the most useful prognostic genes[17]. Then, Kaplan–Meier curves and log-rank test were used to confirm the prognostic value of these genes from LASSO analysis. A P value < 0.05 was considered as cutoff value.

7. Construction Of A Risk Score Model

Based on these survival-related RBP genes, we performed multiple Cox regression analysis to obtain the coefficients. Risk score model was constructed and the risk score was calculated using following formula: $Risk\ score = \beta_1 * Exp_1 + \beta_2 * Exp_2 + \beta_i * Exp_i$. The β was the coefficients from multiple Cox regression analysis, and Exp represented expression levels of selected genes. The Kaplan–Meier curves and ROC curves were implemented to evaluate the efficiency of the risk score model.

8. Statistical analysis

Pearson’s chi-square test were used to compare the distribution of clinical information. Differences among variables were test using t test. Univariate and multivariate Cox regression analysis were performed to evaluate the prognostic value. Kaplan–Meier curves and log-rank test were used to evaluate the survival difference. All these procedures were conducted on R software.

Results:

Mapping of RNA binding proteins and consensus clustering

In this study, we performed a systematic analysis of RNA binding proteins in ccRCC patients. The flow chart of the study design is shown in Fig. 1. A total of 522 ccRCC patients met inclusion criteria and were included for subsequent analysis. We firstly mapped these RBPs genes on TCGA dataset and performed consensus clustering using the ConsensusClusterPlus package[13]. Based on the expression similarity of RBPs genes, $k = 3$ seemed to be a more stable value from $k = 2$ to 10 in the TCGA dataset (Figs. 2B,C). In addition, we also performed PCA analysis to evaluated the reliability of the consensus clustering. After we divided these patients into 3 clusters, PCA analysis showed that the cluster 1 and cluster 3 had high similarity and gathered together (Figs. 2E). Therefore, we divided patients into 2 clusters (Figs. 2A). PCA analysis showed that cluster classifications showed different distribution between two clusters. Patients in cluster 1 gathered together and cluster 2 gathered together (Figs. 2D). These results determined the reliability of two different clusters.

2. Cluster Classification Was Correlated With The Malignancy Of ccRCC

To evaluate the association between two clusters and clinical factors, we performed survival analysis and evaluated the distribution characteristics for the two clusters. Kaplan–Meier curve showed that patients in cluster 2 had poorer prognosis compared with cluster 1 (Figs. 2F). In addition, we found that distributions of clinical factors between two clusters, including gender, stage, grade, pathological T staging and pathological M staging were significantly different (Table 1). Patients in cluster 2 had relatively higher proportions of advanced stage (Fig. 3A), high grade (Fig. 3B), high pathological T staging (Fig. 3C), lymph node metastasis (Fig. 3D) and distant metastasis (Fig. 3E) than cluster 1. These results showed patients in cluster 2 were closely correlated with the malignancy of ccRCC. GSEA analysis revealed that cluster 2 significantly enriched several oncogenic pathways (Fig. 3F), including epithelial mesenchymal transition (normalized enrichment score (NES) = 1.578, size = 194), G2M checkpoint (NES = 1.532, size = 184), KRAS signaling (NES = 1.539, size = 188) and IL6 JAK STAT3 signaling (NES = 1.360, size = 87). All these results indicated that cluster classifications identified by consensus clustering are correlated with the malignancy of ccRCC.

3. Screening Of Differentially Expressed RBPs In ccRCC

A total of 522 ccRCC and 72 normal tissue samples were included to obtain differentially expressed genes using “limma” package, and intersected with RBPs. A total of 115 genes, comprising 71 up-regulated and 44 down-regulated RBPs, were obtained. The results of differentially expressed analysis were showed in Table S1. Then, we performed functional enrichment analysis of these differentially expressed RBPs, and the results meeting the criteria were showed in Table S2. For up-regulated RBPs, the most significantly enriched terms were “SRP-dependent cotranslational protein targeting to membrane” for biological processes, “ribosome” for cellular component, “RNA binding” for molecular function and “Ribosome” for KEGG pathway. For down-regulated RBPs, the most significantly enriched terms were “mRNA processing” for biological processes and “nucleotide binding” for molecular function.

4. PPI Network Construction And Key Module Screening

To further evaluate the roles of these differentially expressed RBPs in ccRCC, we submitted these genes to STRING database and constructed a PPI network (Fig. 4A). We obtained 113 PPI nodes, 398 PPI edges and a PPI enrichment p-value < 1.0e-16. Meanwhile, the functional enrichments for the PPI network were obtained. For biological processes, the mainly enriched terms were mRNA metabolic process, nucleobase-containing compound catabolic process, cellular nitrogen compound catabolic process, RNA catabolic process and cellular nitrogen compound metabolic process. For molecular function, the terms including RNA binding, nucleic acid binding, heterocyclic compound binding, organic cyclic compound binding and structural constituent of ribosome were enriched. For cellular component, the PPI network was located in

ribonucleoprotein complex. They also located in cytosolic ribosome, ribosome, ribosomal subunit and ribonucleoprotein granule. For KEGG pathway,

The mainly enriched results were nonsense mediated decay independent of the exon junction complex, eukaryotic translation termination, selenocysteine synthesis, selenoamino acid metabolism and viral mRNA translation. Then, we screen hub genes using cytoHubba plug-in with MNC algorithm, and ten hub genes were selected: RPLP0, RPS14, OASL, RPS20, RPL35, RPS2, RPL10, RPL30, RPS15 and RPL18. We further selected import subnetworks with MCODE plug-in (Fig. 4B, C). The ClueGO was used to explore the functional mechanism. The module 1 was mainly enriched in cytosolic large ribosomal subunit, cytosolic ribosome, ribosome assembly and ribosome biogenesis (Fig. 4B), whereas the module 2 was mainly enriched in 2'-5'-oligoadenylate synthetase activity, negative regulation of viral genome replication and type I interferon signaling pathway (Fig. 4C).

5. Evaluation Of Hub Genes

These 10 hub genes from PPI network were further evaluated in HPA database. The results showed that these genes (RPLP0, RPS14, OASL, RPS20, RPL35, RPS2, RPL10, RPL30, RPS15 and RPL18) were highly expressed in ccRCC samples compared with normal samples (Fig. 5). In addition, we used ROC curves to evaluate the efficiency of these genes to differentiate tumor samples from normal samples using “pROC” package[15]. These genes showed good performance of diagnostic accuracy for ccRCC (Figure S1). Moreover, the genetic changes (mutation and copy number alteration for TCGA) of these genes were evaluated using the cBioPortal online tool, and 77 samples out of 448 ccRCC samples were found altered (Figure S1K). The mutation frequencies of these 10 genes were low except RPS14 gene. The amplification of RPS14 gene was the most frequent copy-number alteration.

6. Identification Of RBPs With Prognostic Value

The prognostic values of these differentially expressed RBPs were evaluated in ccRCC patients. We performed univariate Cox regression analysis to select survival-related RBPs in TCGA dataset. The results showed that 77 out of 115 differentially expressed RBPs were correlated with patients' overall survival (Table S3). Then, we performed LASSO Cox regression analysis based on survival-related RBPs in TCGA dataset to select the most prognostic genes, and 12 genes were selected, including PABPC1L, IGF2BP3, PAIP2B, RBM47, RPL22L1, IGF2BP2, RNASE2, CSDA, CNP, DDX25, RALYL and NYNRIN (Figure S2A,B). To further determine the prognostic values of these twelve genes, Kaplan–Meier curves and log-rank test for overall survival were used, and ten RBP genes showed good prognostic values between high and low expressions, including CNP (Figure S2C), CSDA (Figure S2D), DDX25 (Figure S2E), IGF2BP2 (Figure S2F), IGF2BP3 (Figure S2G), PABPC1L (Figure S2H), PAIP2B (Figure S2I), RBM47 (Figure S2J), RNASE2 (Figure S2K) and RPL22L1 (Figure S2L). While, NYNRIN (Figure S2M) and RALYL (Figure S2N) showed no significant difference in survival.

7. Construction And Validation Of A Risk Score Model

These ten survival-related RBP genes were performed multiple Cox regression analysis to construct a risk score model. The risk score for each ccRCC patient was calculated using the genes' expressions multiplies the coefficients obtained from multiple cox regression analysis. The detailed formula was showed as follow: Risk score = $(0.22403 * PABPC1L) + (0.08177 * IGF2BP3) + (-0.01128 * PAIP2B) + (-0.28659 * RBM47) + (0.06504 * RPL22L1) + (0.04107 * IGF2BP2) + (0.07886 * RNASE2) + (0.09797 * CSDA) + (0.32254 * CNP) + (-0.13138 * DDX25)$. To evaluate the prognostic value of the risk score model, patients were divided into high-risk and low-risk groups based on the median risk score. PCA analysis showed different distribution patterns between high-risk and low-risk patients (Fig. 6A), of which high-risk patients gathered together and low-risk patients gathered together. Kaplan–Meier curves and log-rank test were used and showed that high-risk patients had poor prognosis compared with low-risk patients (Fig. 6B). The prognostic performance of risk score model was evaluated using ROC curves (Fig. 6C). The AUCs of the curves were 0.756 at 1 year, 0.717 at 2 years, 0.738 at 3 year and 0.761 at 5 year, showing the risk score model had moderate sensitivity and specificity. The risk score model was further validated in E-MTAB-3267 set. The risk score for each patient was calculate using the same formula and patients were divided into high- and low-risk based on median risk score. PCA analysis also showed different distribution patterns (Fig. 6D). Kaplan–Meier curve showed high-risk patients had poor prognosis despite the P value larger than 0.05 (Fig. 6E). The ROC curves also showed the good performance of the risk score model (Fig. 6F). These results showed the reliability and stability of the risk score model.

8. Risk score model showed strong associations with clinicopathological factors in ccRCC

The heat map showed the expression levels of ten genes for each patient in TCGA dataset (Fig. 7A). The clinicopathological factors between high-risk and low-risk patients were compared. We found that high-risk patients had high proportion of advanced stage ($P < 0.0001$), high grade ($P < 0.0001$), high pathological T staging ($P < 0.0001$), male ($P < 0.05$) and death ($P < 0.0001$). We performed univariate Cox regression analysis for these clinical factors and risk score model, and we found the risk score model was risk factor for overall survival of ccRCC patients (Fig. 7B). We also performed multiple Cox regression analysis and found the risk score model was still an independent risk factor for overall survival after integrating with age, gender, stage, grade, pathological T staging, pathological N staging and pathological M staging (Fig. 7C). These results showed the risk score model was closely correlated with clinicopathological factors in ccRCC.

Discussion

Despite advances in diagnosis and treatment of ccRCC over the past decades, reducing mortality and improving prognosis of ccRCC patients remain a formidable task which faces many challenges such as lacking of clinically available biomarkers, and insufficient understanding of the underlying molecular

mechanism of ccRCC[18]. Therefore, it is evident that a comprehensive understanding of potential functions and identification of sensitive biomarkers of ccRCC will facilitate the improvement of survival in patients with ccRCC. Previous studies shown that RBPs dysregulation leads to tumorigenesis and progression of various human tumors[19–21]. However, most of the RBPs that participate in ccRCC tumorigenesis and progression remained unexplored. In present study, we firstly mapped these RBP genes on TCGA dataset and performed consensus clustering. A total of 115 differently expressed RBPs were obtained, comprising 71 up-regulated and 44 down-regulated RBPs. Then, we explored the biological pathways and constructed a PPI network for these RBPs. The expression levels, genomic variation and diagnostic value of hub RBPs were comprehensively evaluated. Moreover, we constructed and validated a risk score model to evaluate the prognosis of ccRCC based on survival-related RBPs.

RBPs, mainly function as post-transcriptional regulatory factors, play key roles in maintaining the homeostasis of cellular physiology[22], and the disturbance of these essential cellular functions regulated by RBPs can lead to various disease, including cancer[23]. In ccRCC, many RBPs have been revealed playing important roles in tumorigenesis[7, 24]. RBPs have different functions, where some RBPs can enhance tumorigenesis[10, 11], and some RBPs can repress tumorigenesis[8, 9, 12]. Taking into account the important roles of RBPs in ccRCC, we comprehensively explored the roles of all these RBPs in ccRCC using consensus clustering method. Two ccRCC subgroups, cluster 1 and 2, were identified based on the expression similarity of RBPs. PCA analysis confirmed the reliability of the consensus clustering and showed different distribution between two clusters. We showed different prognosis between the two clusters and cluster 2 were closely correlated with the malignancy of ccRCC. Moreover, we found several oncogenic pathways, including epithelial mesenchymal transition, G2M checkpoint, KRAS signaling and IL6 JAK STAT3 signaling were correlated with the malignancy of ccRCC.

RBPs have been shown to critically regulate multiple events of RNA lifecycle, including transcriptional, splicing, RNA trafficking and sequestration[25]. RBPs also contribute to translational and post-translational regulation through binding to 3' untranslated regions of mRNAs[26]. An interaction between the SRP receptor and the translocon is critical in the process of protein cotranslational translocation[27]. Many studies showed deregulated expression of RBPs in different diseases, including cancer[22]. In present study, we obtained several differently expressed RBPs of ccRCC, which might play vital roles in tumorigenesis and progression of ccRCC. Functional enrichment analysis showed that these differentially expressed RBPs were principally enriched in SRP-dependent cotranslational protein targeting to membrane, mRNA processing, RNA binding, nucleotide binding and ribosome. To further explore the functions of these differently expressed RBPs, we constructed a PPI network and screened key modules which might play more important roles in ccRCC. The biological functions of these key modules were involved in cytosolic large ribosomal subunit, cytosolic ribosome, ribosome assembly, ribosome biogenesis, 2'-5'-oligoadenylate synthetase activity, negative regulation of viral genome replication and type I interferon signaling pathway.

To further investigate the roles of these RBPs in tumorigenesis of ccRCC, we screened the top ten hub genes from the PPI network, containing RPLP0, RPS14, OASL, RPS20, RPL35, RPS2, RPL10, RPL30,

RPS15 and RPL18. We found that all these hub genes were up-regulated in ccRCC. A part of them has been showed to be involved in tumorigenesis. Previous studies showed that RPLP0 deficiency could suppress tumor growth in gynecologic tumor[28] and gastric cancer[29] through regulating cell cycle and apoptosis. RPS20 was identified as a new oncogene in glioblastoma[30] and colon cancer[31]. RPL35 was identified as an key factors for tumorigenesis in neuroblastoma[32]. RPS2 was overexpressed in malignant prostate cancer and had the potential to be therapeutic target for prostate cancer[33]. RPL10 played a vital role in tumorigenesis and progression of pancreatic cancer by regulating productions of reactive oxygen species[34]. The overexpression of RPL30 was associated with adverse prognosis in medulloblastoma[35]. RPS15 was identified as a new driver gene in aggressive and chemorefractory cases of chronic lymphocytic leukemia[36]. However, the roles of these genes in ccRCC were needed to be further investigated. In present study, we validated the high expressions of these hub genes in HPA database. Moreover, we estimated the diagnostic values for ccRCC of these hub genes. The diagnostic accuracy rates of all these hub genes were larger than 0.85, showing excellent performance. All these genes might be the potential diagnostic biomarkers for ccRCC. The genomic alterations of ten hub genes were low, showing genomic alterations were not the main cause of change of gene expressions.

To obtained RBPs with excellent prognostic value, we used univariate and Lasso Cox regression analysis and Kaplan–Meier curves to filter out these differently expressed RBPs. A total of ten RBPs with prognostic value were obtained, including CNP, CSDA, DDX25, IGF2BP2, IGF2BP3, PABPC1L, PAIP2B, RBM47, RNASE2 and RPL22L1. Previous studies found that CSDA expression was associated with Fuhrman grade of ccRCC[37]. High plasma and tissue levels of IGF2BP3 are independently associated with poor prognosis in ccRCC[38]. CNP was correlated with survival and could be used as a novel and potentially useful clinical prognosis biomarker in glioblastoma multiforme[39]. The expression of IGF2BP2 was up-regulated and led to a poor prognosis in pancreatic ductal adenocarcinoma[40] and acute myelocytic leukemia[41]. PABPC1L was emerging as a promising prognostic factor for colon cancer[42]. RBM47 was identified as a protective factor in lung adenocarcinoma[43] and breast cancer[44], which could suppressed tumor growth. RPL22L1 was a prognostic marker for colorectal cancer[45] and prostate cancer[46]. While, the potential roles and prognostic values of other RBPs remains to be further investigated. Taking into account the significantly prognostic values of these ten genes, an risk score model was established to help guiding the stratification of ccRCC patients.

With the advent of precision cancer medicine, more performant risk stratification models are urgently needed to improve management strategies for ccRCC patients. Considering the important roles of RBPs in life processes and tumor, a risk score model based on RBPs with excellent prognostic values may provide new insights for clinical practice of ccRCC patients. The risk score model based on RBPs have been established in lung squamous cell carcinoma[47], lung adenocarcinoma[48] and breast cancer[49], and showed good performance. In present study, we were the first to establish a risk score model based on RBPs with prognostic values in ccRCC patients. The model could be used to stratify patients with different prognosis. The model showed good performance with moderate sensitivity and specificity for predicting patients' prognosis. The model was further validated in an external dataset, showing the reliability and stability of risk score model. Moreover, we further investigated the association between the

model and clinical factors. We found high-risk patients tended to be advanced stage, high grade, high pathological T staging and male, and had higher proportions of death. These results showed close correlations between the model and traditional clinical factors. We also found the risk score model could be an independent risk factor for overall survival of ccRCC patients. Therefore, the model could be used as a supplement for clinical factors to improve clinical evaluation of ccRCC patients. However, there are several limitations in this study. First, we used data from multiple public database. These potential biomarkers with diagnostic and prognostic values and the risk score model identified in this study were needed to further validated in our own tissue samples. Second, our study was a retrospective study and prospective studies were needed to be performed to confirm our outcomes.

Conclusions

In conclusion, our study systematically investigated the roles of RBPs in ccRCC using a series of bioinformatics methods. We identified and validated several potential biomarkers for ccRCC patients with good diagnostic value. These potential biomarkers could be used for accurate and early diagnosis of ccRCC. We also identified several prognostic biomarkers for ccRCC, which could be used to predict patients' prognosis. Moreover, it was the first time to establish a prognostic model based on RBPs with prognostic value. The model showed good performance in stratification of ccRCC patients and prediction of prognosis. In addition, the model was an independent risk factor for overall survival of ccRCC patients. Therefore, it could be used as a complementation for clinical factors to improve prognostic evaluation of ccRCC patients.

Abbreviations

RBPs: RNA binding proteins; RCC: renal cell carcinoma; ccRCC: clear cell renal cell carcinoma; PPI: protein-protein interaction; HPA: The Human Protein Atlas; ROC: receiver operating characteristic; TCGA: The Cancer Genome Atlas; PCA: principal component analysis; GO: gene ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; GSEA: gene set enrichment analysis; AUC: area under the curve; CNA: copy number alteration; NES: normalized enrichment score.

Declarations

Acknowledgements

We thank the TCGA and ArrayExpress database (E-MTAB-3267) for providing valuable datasets.

Authors' contributions

C.Z.L. and L.Z. conceived and designed the experiments. X.L.H. and J.C. acquired and analyzed the data. X.L.H. and S.D.G. wrote this manuscript. H.B.X. checked the manuscript. All authors read and approved the final manuscript.

Funding

This study was funded by the National Natural Science Foundation of China (No. 81630019, 81870519), Scientific Research Foundation of the Institute for Translational Medicine of Anhui Province (2017ZHXY02).

Availability of data and materials

All data generated during this study are included in this published article.

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing interests.

References

1. Ljungberg B, Bensalah K, Canfield S, Dabestani S, Hofmann F, Hora M, Kuczyk MA, Lam T, Marconi L, Merseburger AS, Mulders P, Powles T, Staehler M, Volpe A, Bex A. EAU guidelines on renal cell carcinoma: 2014 update. *Eur Urol*. 2015;67(5):913–24.
2. Siegel RL, Miller KD, Jemal A. Cancer statistics. 2015. *CA Cancer J Clin*. 2015;65(1):5–29.
3. Znaor A, Lortet-Tieulent J, Laversanne M, Jemal A, Bray F. International variations and trends in renal cell carcinoma incidence and mortality. *Eur Urol*. 2015;67(3):519–30.
4. Ferlay J, Steliarova-Foucher E, Lortet-Tieulent J, Rosso S, Coebergh JW, Comber H, Forman D, Bray F. Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012. *Eur J Cancer*. 2013;49(6):1374–403.
5. Hsieh JJ, Purdue MP, Signoretti S, Swanton C, Albiges L, Schmidinger M, Heng DY, Larkin J, Ficarra V. Renal cell carcinoma. *Nat Rev Dis Primers*. 2017;3:17009.
6. Gerstberger S, Hafner M, Tuschl T. A census of human RNA-binding proteins. *Nat Rev Genet*. 2014;15(12):829–45.
7. Perron G, Jandaghi P, Solanki S, Safisamghabadi M, Storoz C, Karimzadeh M, Papadakis AI, Arseneault M, Scelo G, Banks RE, Tost J, Lathrop M, Tanguay S, Brazma A, Huang S, Brimo F, Najafabadi HS, Riazalhosseini Y. A General Framework for Interrogation of mRNA Stability Programs Identifies RNA-Binding Proteins that Govern Cancer Transcriptomes. *Cell Rep*. 2018;23(6):1639–50.

8. Zhang RL, Yang JP, Peng LX, Zheng LS, Xie P, Wang MY, Cao Y, Zhang ZL, Zhou FJ, Qian CN, Bao YX. RNA-binding protein QKI-5 inhibits the proliferation of clear cell renal cell carcinoma via post-transcriptional stabilization of RASA1 mRNA. *Cell cycle*. 2016;15(22):3094–104.
9. Shi F, Wei D, Zhu Z, Yan F, Wang F, Zhang K, Li X, Zheng Y, Yuan J, Lu Z, Yuan J. The RNA-binding protein QKI suppresses tumorigenesis of clear cell renal cell carcinoma by regulating the expression of HIF-1alpha. *J Cancer*. 2020;11(6):1359–70.
10. Jiang J, Chen X, Liu H, Shao J, Xie R, Gu P, Duan C. Polypyrimidine Tract-Binding Protein 1 promotes proliferation, migration and invasion in clear-cell renal cell carcinoma by regulating alternative splicing of PKM. *Am J Cancer Res*. 2017;7(2):245–59.
11. Zhou KW, Jiang K, Zhu W, Weng G. Expression of cold-inducible RNA-binding protein (CIRP) in renal cell carcinoma and the effect of CIRP downregulation cell proliferation and chemosensitivity to gemcitabine. *Oncol Lett*. 2018;15(5):7611–6.
12. Huang W, Wei XL, Ni W, Cao M, Meng L, Yang H. The expression of RNA-binding protein RBM38 decreased in renal cell carcinoma and represses renal cancer cell proliferation, migration, and invasion. *Tumour Biol*. 2017;39(5):1010428317701635.
13. Wu J, Cui Y, Sun X, Cao G, Li B, Ikeda DM, Kurian AW, Li R. Unsupervised Clustering of Quantitative Image Phenotypes Reveals Breast Cancer Subtypes with Distinct Prognoses and Molecular Pathways. *Clin Cancer Res*. 2017;23(13):3334–42.
14. Uhlen M, Zhang C, Lee S, Sjostedt E, Fagerberg L, Bidkhori G, Benfeitas R, Arif M, Liu Z, Edfors F, Sanli K, von Feilitzen K, Oksvold P, Lundberg E, Hober S, Nilsson P, Mattsson J, Schwenk JM, Brunnstrom H, Glimelius B, Sjoblom T, Edqvist PH, Djureinovic D, Micke P, Lindskog C, Mardinoglu A, Ponten F. A pathology atlas of the human cancer transcriptome. *Science*. 2017;357(6352):pii: eaan2507.
15. Sing T, Sander O, Beerenwinkel N, Lengauer T. ROCr: visualizing classifier performance in R. *Bioinformatics*. 2005;21(20):3940–1.
16. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson E, Cerami E, Sander C, Schultz N. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal*. 2013;6(269):pl1.
17. Tibshirani R. The lasso method for variable selection in the Cox model. *Stat Med*. 1997;16(4):385–95.
18. Bhalla S, Chaudhary K, Kumar R, Sehgal M, Kaur H, Sharma S, Raghava GP. Gene expression-based biomarkers for discriminating early and late stage of clear cell renal cancer. *Sci Rep*. 2017;7:44997.
19. Pereira B, Billaud M, Almeida R. RNA-Binding Proteins in Cancer: Old Players and New Actors. *Trends Cancer*. 2017;3(7):506–28.
20. Dong W, Dai ZH, Liu FC, Guo XG, Ge CM, Ding J, Liu H, Yang F. The RNA-binding protein RBM3 promotes cell proliferation in hepatocellular carcinoma by regulating circular RNA SCD-circRNA 2 production. *Ebiomedicine*. 2019;45:155–67.
21. Velasco MX, Kosti A, Penalva LOF, Hernandez G. The Diverse Roles of RNA-Binding Proteins in Glioma Development. *Adv Exp Med Biol*. 2019;1157:29–39.

22. Masuda K, Kuwano Y. Diverse roles of RNA-binding proteins in cancer traits and their implications in gastrointestinal cancers. *Wiley Interdiscip Rev RNA*. 2019;10(3):e1520.
23. Wang ZL, Li B, Luo YX, Lin Q, Liu SR, Zhang XQ, Zhou H, Yang JH, Qu LH. Comprehensive Genomic Characterization of RNA-Binding Proteins across Human Cancers. *Cell Rep*. 2018;22(1):286–98.
24. Danilin S, Sourbier C, Thomas L, Lindner V, Rothhut S, Dormoy V, Helwig JJ, Jacqmin D, Lang H, Massfelder T. Role of the RNA-binding protein HuR in human renal cell carcinoma. *Carcinogenesis*. 2010;31(6):1018–26.
25. Picchiarelli G, Dupuis L. Role of RNA Binding Proteins with prion-like domains in muscle and neuromuscular diseases. *Cell stress*. 2020;4(4):76–91.
26. Mayr C, Bartel DP. Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell*. 2009;138(4):673–84.
27. Jiang Y, Cheng Z, Mandon EC, Gilmore R. An interaction between the SRP receptor and the translocon is critical during cotranslational protein translocation. *J Cell Biol*. 2008;180(6):1149–61.
28. Wang CH, Wang LK, Wu CC, Chen ML, Lee MC, Lin YY, Tsai FM. The Ribosomal Protein RPLP0 Mediates PLAAT4-induced Cell Cycle Arrest and Cell Apoptosis. *Cell Biochem Biophys*. 2019;77(3):253–60.
29. Teller A, Jechorek D, Hartig R, Adolf D, Reissig K, Roessner A, Franke S. Dysregulation of apoptotic signaling pathways by interaction of RPLP0 and cathepsin X/Z in gastric cancer. *Pathol Res Pract*. 2015;211(1):62–70.
30. Yong WH, Shabihkhani M, Telesca D, Yang S, Tso JL, Menjivar JC, Wei B, Lucey GM, Mareninov S, Chen Z, Liau LM, Lai A, Nelson SF, Cloughesy TF, Tso CL. Ribosomal Proteins RPS11 and RPS20, Two Stress-Response Markers of Glioblastoma Stem Cells, Are Novel Predictors of Poor Prognosis in Glioblastoma Patients. *Plos One*. 2015;10(10):e0141334.
31. Nieminen TT, O'Donohue MF, Wu Y, Lohi H, Scherer SW, Paterson AD, Ellonen P, Abdel-Rahman WM, Valo S, Mecklin JP, Jarvinen HJ, Gleizes PE, Peltomaki P. Germline mutation of RPS20, encoding a ribosomal protein, causes predisposition to hereditary nonpolyposis colorectal carcinoma without DNA mismatch repair deficiency. *Gastroenterology*. 2014;147(3):595–8.e5.
32. Liu PY, Tee AE, Milazzo G, Hannan KM, Maag J, Mondal S, Atmadibrata B, Bartonicek N, Peng H, Ho N, Mayoh C, Ciaccio R, Sun Y, Henderson MJ, Gao J, Everaert C, Hulme AJ, Wong M, Lan Q, Cheung BB, Shi L, Wang JY, Simon T, Fischer M, Zhang XD, Marshall GM, Norris MD, Haber M, Vandesompele J, Li J, Mestdagh P, Hannan RD, Dinger ME, Perini G, Liu T. The long noncoding RNA IncNB1 promotes tumorigenesis by interacting with ribosomal protein RPL35. *Nat Commun*. 2019;10(1):5026.
33. Wang M, Hu Y, Stearns ME. RPS2: a novel therapeutic target in prostate cancer. *J Exp Clin Cancer Res*. 2009;28:6.
34. Yang J, Chen Z, Liu N, Chen Y. Ribosomal protein L10 in mitochondria serves as a regulator for ROS level in pancreatic cancer cells. *Redox Biol*. 2018;19:158–65.

35. De Bortoli M, Castellino RC, Lu XY, Deyo J, Sturla LM, Adesina AM, Perlaky L, Pomeroy SL, Lau CC, Man TK, Rao PH, Kim JY. Medulloblastoma outcome is adversely associated with overexpression of EEF1D, RPL30, and RPS20 on the long arm of chromosome 8. *BMC Cancer*. 2006;6:223.
36. Ljungstrom V, Cortese D, Young E, Pandzic T, Mansouri L, Plevova K, Ntoufa S, Baliakas P, Clifford R, Sutton LA, Blakemore SJ, Stavroyianni N, Agathangelidis A, Rossi D, Hoglund M, Kotaskova J, Juliusson G, Belessi C, Chiorazzi N, Panagiotidis P, Langerak AW, Smedby KE, Oscier D, Gaidano G, Schuh A, Davi F, Pott C, Strefford JC, Trentin L, Pospisilova S, Ghia P, Stamatopoulos K, Sjoblom T, Rosenquist R. Whole-exome sequencing in relapsing chronic lymphocytic leukemia: clinical impact of recurrent RPS15 mutations. *Blood*. 2016;127(8):1007–16.
37. Dupasquier S, Delmarcelle AS, Marbaix E, Cosyns JP, Courtoy PJ, Pierreux CE. Validation of housekeeping gene and impact on normalized gene expression in clear cell renal cell carcinoma: critical reassessment of YBX3/ZONAB/CSDA expression. *BMC Mol Biol*. 2014;15:9.
38. Jiang Z, Chu PG, Woda BA, Liu Q, Balaji KC, Rock KL, Wu CL. Combination of quantitative IMP3 and tumor stage: a new system to predict metastasis for patients with localized renal cell carcinomas. *Clin Cancer Res*. 2008;14(17):5579–84.
39. Zorniak M, Clark PA, Leeper HE, Tipping MD, Francis DM, Kozak KR, Salamat MS, Kuo JS. Differential expression of 2',3'-cyclic-nucleotide 3'-phosphodiesterase and neural lineage markers correlate with glioblastoma xenograft infiltration and patient survival. *Clin Cancer Res*. 2012;18(13):3628–36.
40. Huang S, Wu Z, Cheng Y, Wei W, Hao L. Insulin-like growth factor 2 mRNA binding protein 2 promotes aerobic glycolysis and cell proliferation in pancreatic ductal adenocarcinoma via stabilizing GLUT1 mRNA. *Acta Biochim Biophys Sin*. 2019;51(7):743–52.
41. He X, Li W, Liang X, Zhu X, Zhang L, Huang Y, Yu T, Li S, Chen Z. IGF2BP2 Overexpression Indicates Poor Survival in Patients with Acute Myelocytic Leukemia. *Cell Physiol Biochem*. 2018;51(4):1945–56.
42. Zhang B, Babu KR, Lim CY, Kwok ZH, Li J, Zhou S, Yang H, Tay Y. A comprehensive expression landscape of RNA-binding proteins (RBPs) across 16 human cancer types. *RNA Biol*. 2020;17(2):211–26.
43. Sakurai T, Isogaya K, Sakai S, Morikawa M, Morishita Y, Ehata S, Miyazono K, Koinuma D. RNA-binding motif protein 47 inhibits Nrf2 activity to suppress tumor growth in lung adenocarcinoma. *Oncogene*. 2016;35(38):5000–9.
44. Vanharanta S, Marney CB, Shu W, Valiente M, Zou Y, Mele A, Darnell RB, Massague J. Loss of the multifunctional RNA-binding protein RBM47 as a source of selectable metastatic traits in breast cancer. *Elife*. 2014;3.
45. Rao S, Peri S, Hoffmann J, Cai KQ, Harris B, Rhodes M, Connolly DC, Testa JR, Wiest DL. RPL22L1 induction in colorectal cancer is associated with poor prognosis and 5-FU resistance. *Plos One*. 2019;14(10):e0222392.
46. Liang Z, Mou Q, Pan Z, Zhang Q, Gao G, Cao Y, Gao Z, Pan Z, Feng W. Identification of candidate diagnostic and prognostic biomarkers for human prostate cancer: RPL22L1 and RPS21. *Med Oncol*.

2019;36(6):56.

47. Li W, Li X, Gao LN, You CG. Integrated Analysis of the Functions and Prognostic Values of RNA Binding Proteins in Lung Squamous Cell Carcinoma. *Front Genet.* 2020;11:185.
48. Li W, Gao LN, Song PP, You CG. Development and validation of a RNA binding protein-associated prognostic model for lung adenocarcinoma. *Aging.* 2020;12(4):3558–73.
49. Wang K, Li L, Fu L, Yuan Y, Dai H, Zhu T, Zhou Y, Yuan F. Integrated Bioinformatics Analysis the Function of RNA Binding Proteins (RBPs) and Their Prognostic Value in Breast Cancer. *Front Pharmacol.* 2019;10:140.

Table

Table 1

Variables	TCGA set (n = 522), n(%)	Risk score		P value
		Cluster1	Cluster2	
Age (mean ± SD, years)	60.4 ± 12.0	61.25 ± 11.95	59.87 ± 12.08	0.196
Gender				< 0.001
Male	342 (65.5)	53 (15.50)	289 (84.50)	
Female	180 (34.5)	157 (87.22)	23 (12.78)	
Stage				< 0.001
I	260 (49.8)	126 (48.46)	134 (51.54)	
II	55 (10.5)	17 (30.91)	38 (69.09)	
III	121 (23.2)	45 (37.19)	76 (62.81)	
IV	83 (15.9)	21 (25.30)	62 (74.70)	
Unkonwn	3 (0.6)	1 (33.33)	2 (66.67)	
Grade				< 0.001
G1	12 (2.3)	8 (66.67)	4 (33.33)	
G2	225 (43.1)	113 (50.22)	112 (49.78)	
G3	203 (38.9)	67 (33.00)	136 (67.00)	
G4	74 (14.2)	18 (24.32)	56 (75.68)	
Unkonwn	8 (1.5)	4 (50.00)	4 (50.00)	
T stage				< 0.001
T1	266 (51.0)	129 (48.50)	137 (51.50)	
T2	67 (12.8)	19 (28.36)	48 (71.64)	
T3	178 (34.1)	61 (34.27)	117 (65.73)	
T4	11 (2.1)	1 (9.09)	10 (90.91)	
N stage				0.607
N0/Nx	506 (96.9)	205 (40.51)	301 (59.49)	
N1	16 (3.1)	5 (31.25)	11 (68.75)	
M stage				0.008
M0/Mx	443 (84.9)	189 (42.66)	254 (57.34)	

Variables	TCGA set (n = 522), n(%)	Risk score		P value
		Cluster1	Cluster2	
M1	79 (15.1)	21 (26.58)	58 (73.42)	

Supplementary Figure Legends

Figure S1. ROC curves for 10 hub genes from PPI network in TCGA dataset. (A) OASL; (B) RPL10; (C) RPL18; (D) RPL30; (E) RPL35; (F) RPLP0; (G) RPS2; (H) RPS14; (I) RPS15; (J) RPS20. (K) Mutation frequency of these 10 hub genes.

Figure S2. Survival analysis for the 12 survival-related genes. (A, B) The coefficients of LASSO cox regression analysis to select the most useful prognostic genes were showed. Kaplan–Meier curves for CNP (C), CSDA (D), DDX25 (E), IGF2BP2 (F), IGF2BP3 (G), PABPC1L (H), PAIP2B (I), RBM47 (J), RNASE2 (K), RPL22L1 (L), NYNRIN (M) and RALYL (N).

Figures

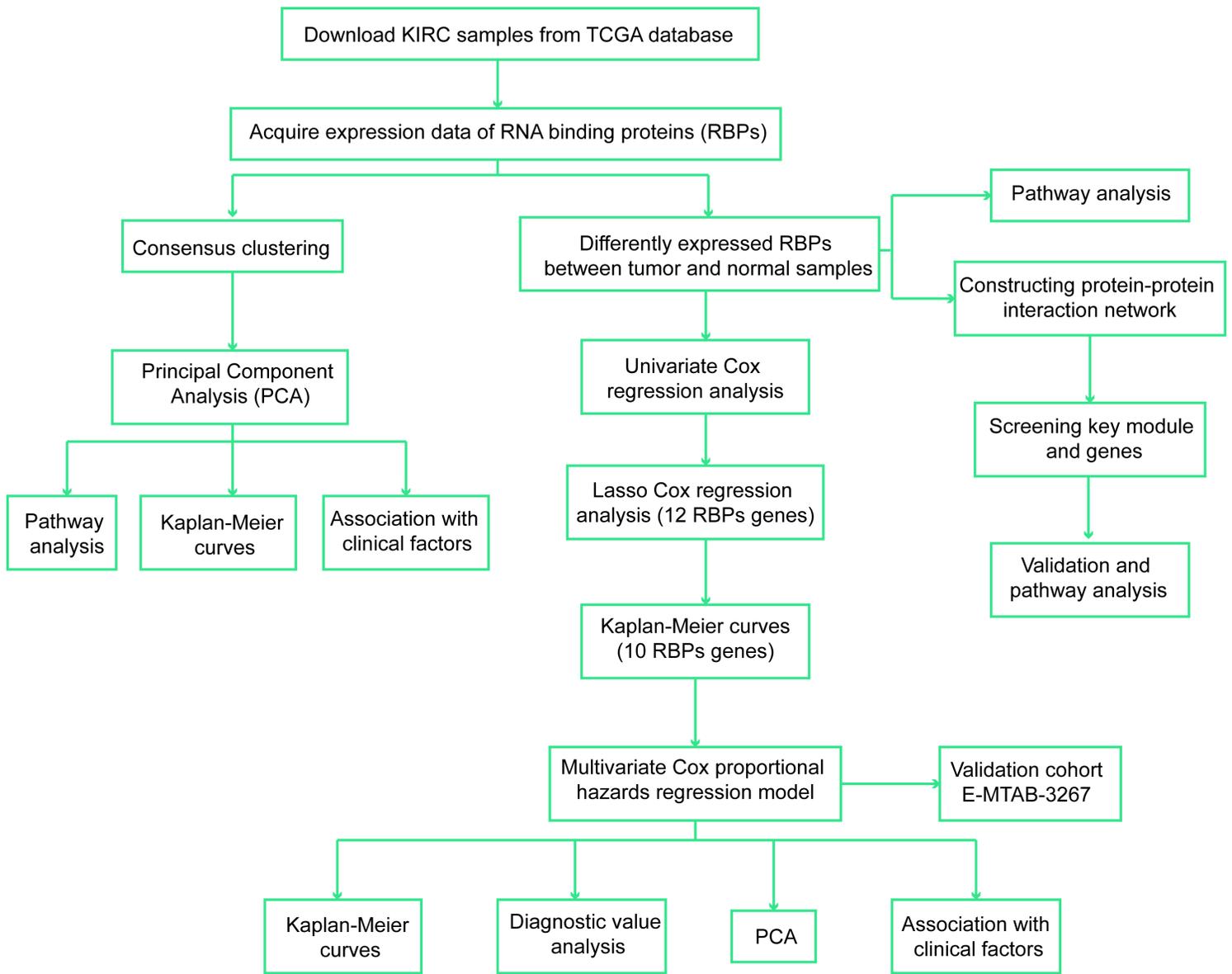


Figure 1

The flow chart of this study design.

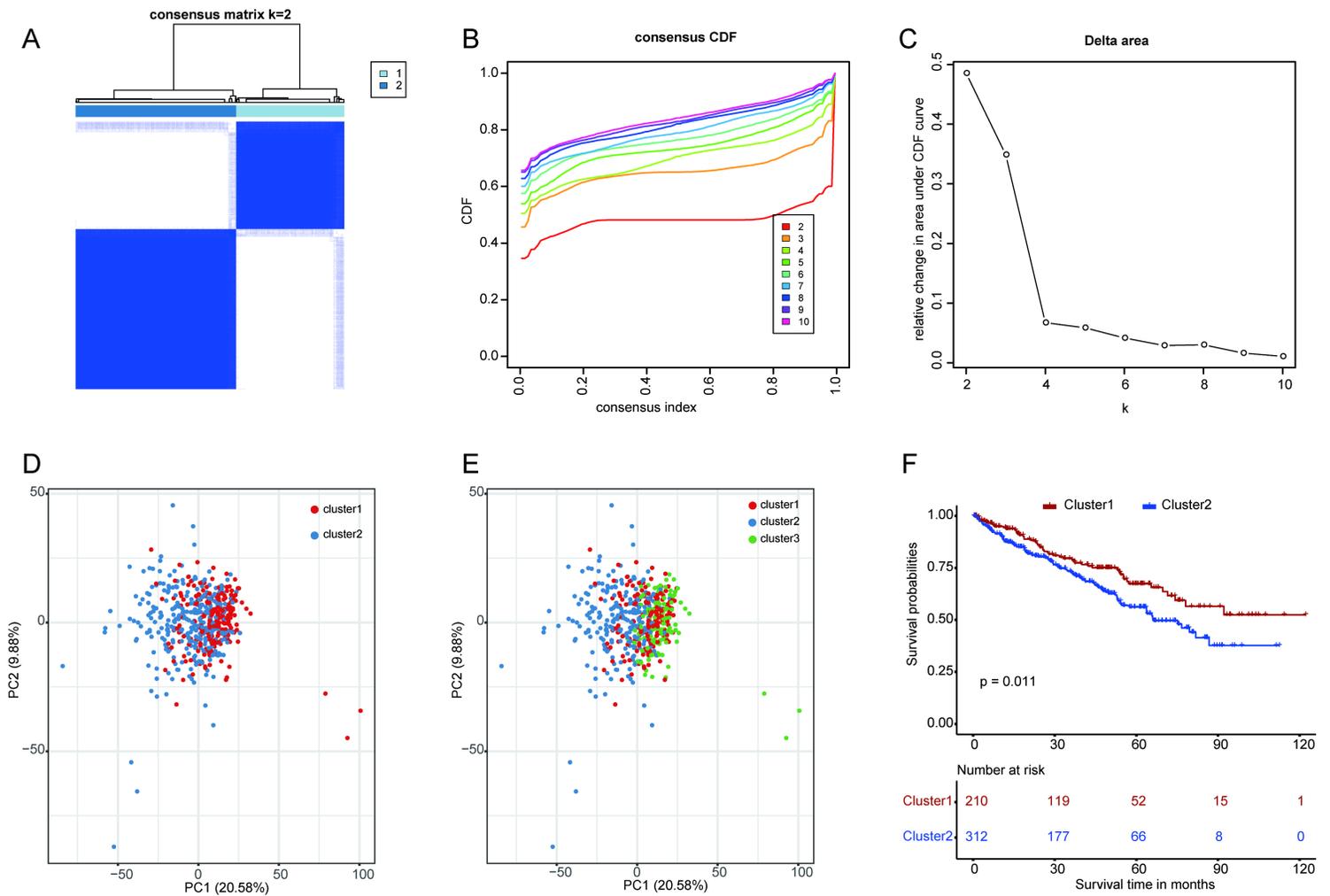


Figure 2

Identification of cluster numbers using consensus clustering. (A) Consensus clustering matrix for $k = 2$. (B) Consensus clustering cumulative distribution function (CDF) for $k = 2$ to 10. (C) Relative change in area under CDF curve for $k = 2$ to 10. Principal component analysis for evaluating the distribution of two clusters (D) and three clusters (E) identified by consensus clustering. (F) Kaplan–Meier curve showed different prognosis between the two clusters.

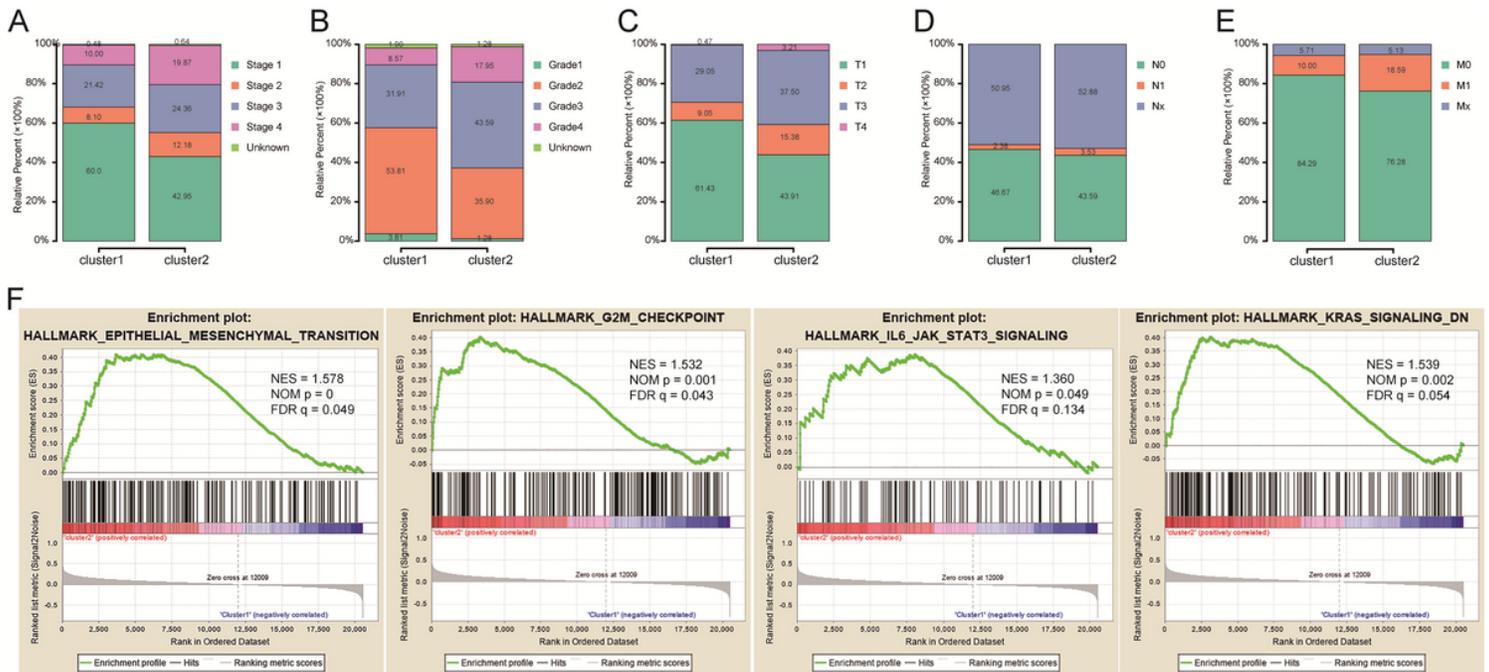


Figure 3

Distributions of clinical factors between two clusters and GSEA analysis. Different distributions between two clusters for stage (A), grade (B), pathological T staging (C), lymph node metastasis (D) and distant metastasis (E). (F) GSEA revealed several activated oncogenic pathways in cluster 2.

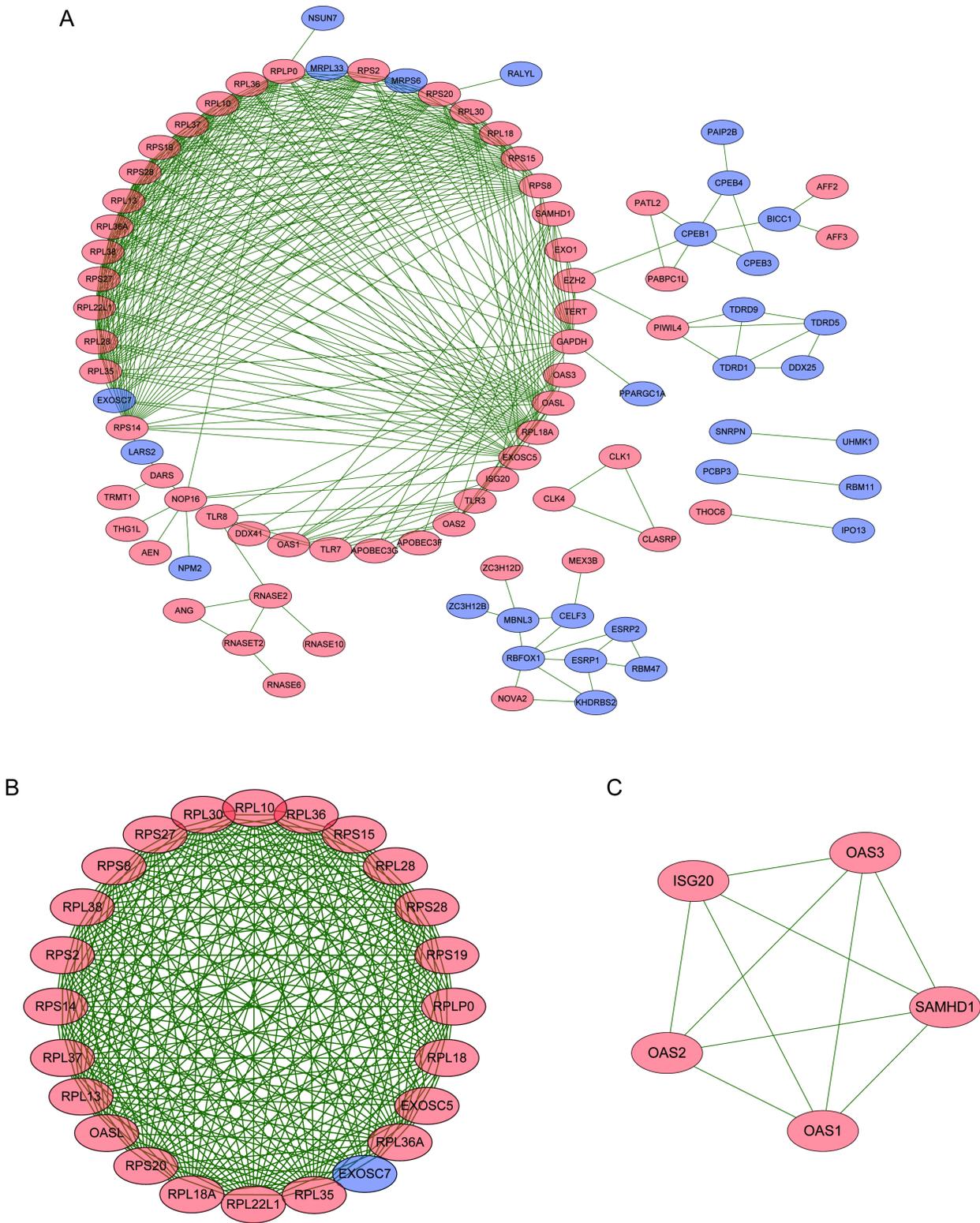


Figure 4

Protein-protein interaction (PPI) network construction. (A) PPI network for differentially expressed RNA binding proteins. (B) Critical module 1 in PPI network. (C) Critical module 2 in PPI network.

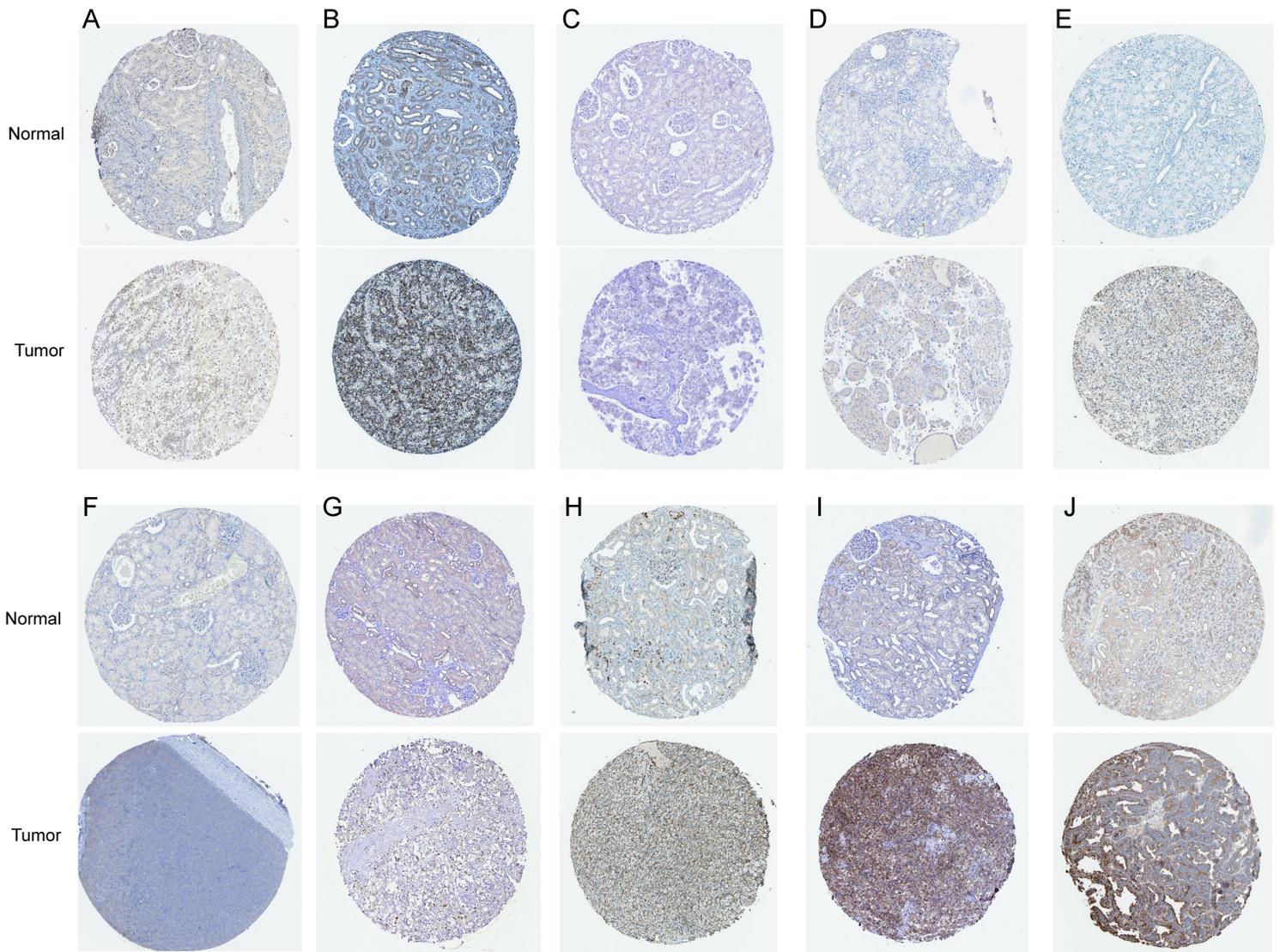


Figure 5

Validation of hub genes for normal and ccRCC samples in The Human Protein Atlas database. (A) OASL; (B) RPL10; (C) RPL18; (D) RPL30; (E) RPL35; (F) RPLP0; (G) RPS2; (H) RPS14; (I) RPS15; (J) RPS20.

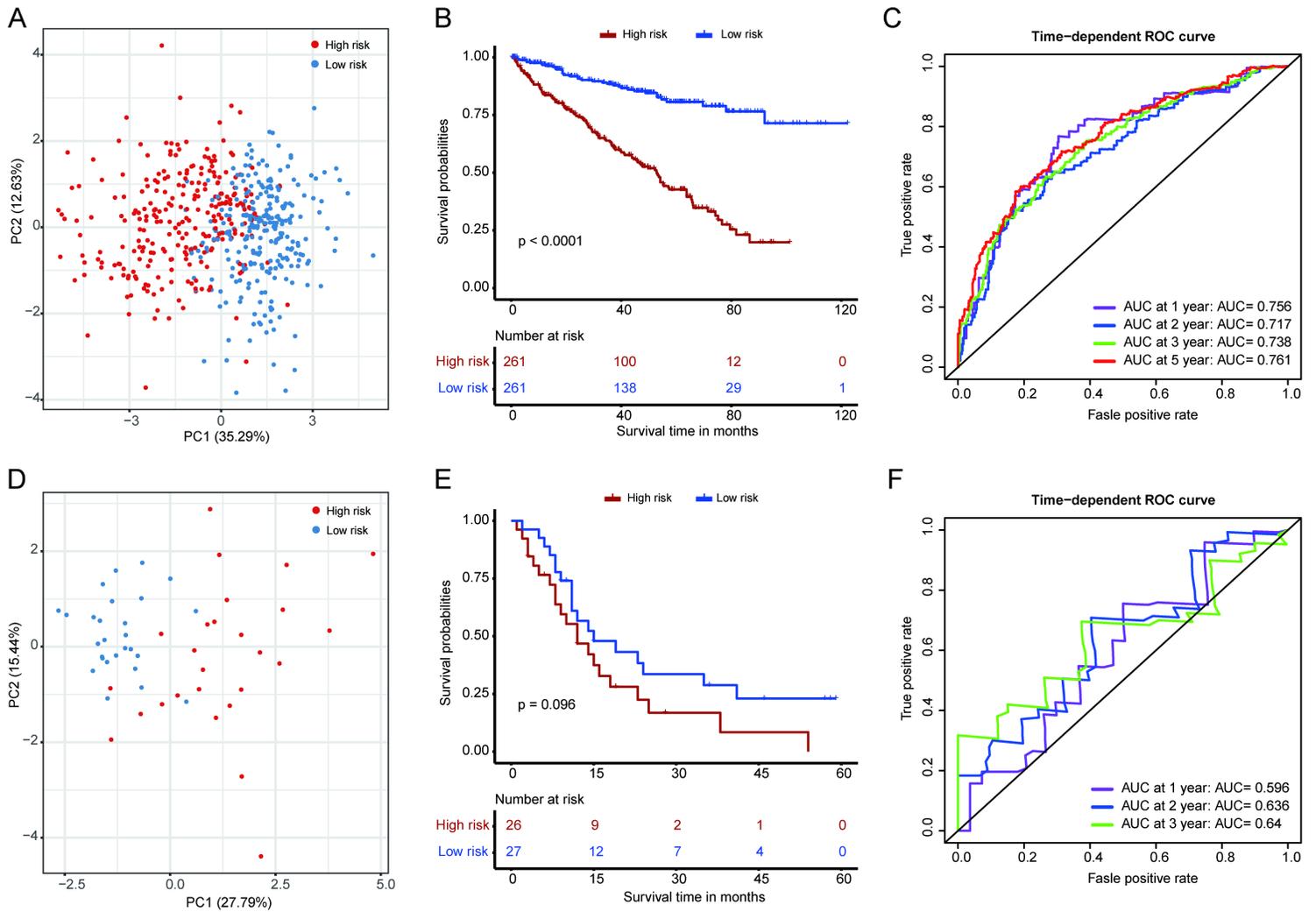


Figure 6

Construction and validation of a risk score model. Principal component analysis (PCA) was performed to evaluate the distribution characteristics for high-risk and low-risk patients in TCGA (A) and E-MTAB-3267 sets (D). Kaplan-Meier curves showed poor prognosis for high-risk patients in TCGA (B) and E-MTAB-3267 sets (E). ROC curves showed good performance of the risk score model in TCGA (C) and E-MTAB-3267 sets (F).

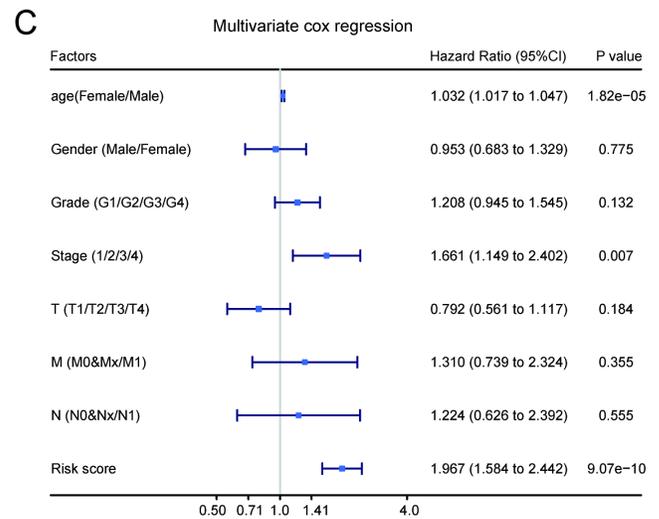
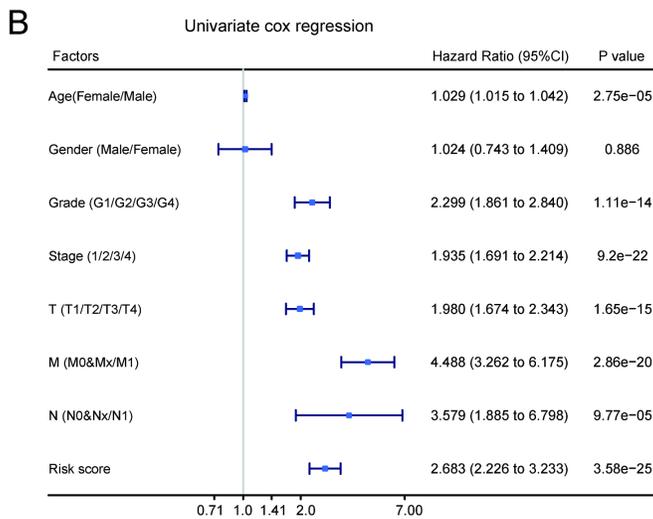
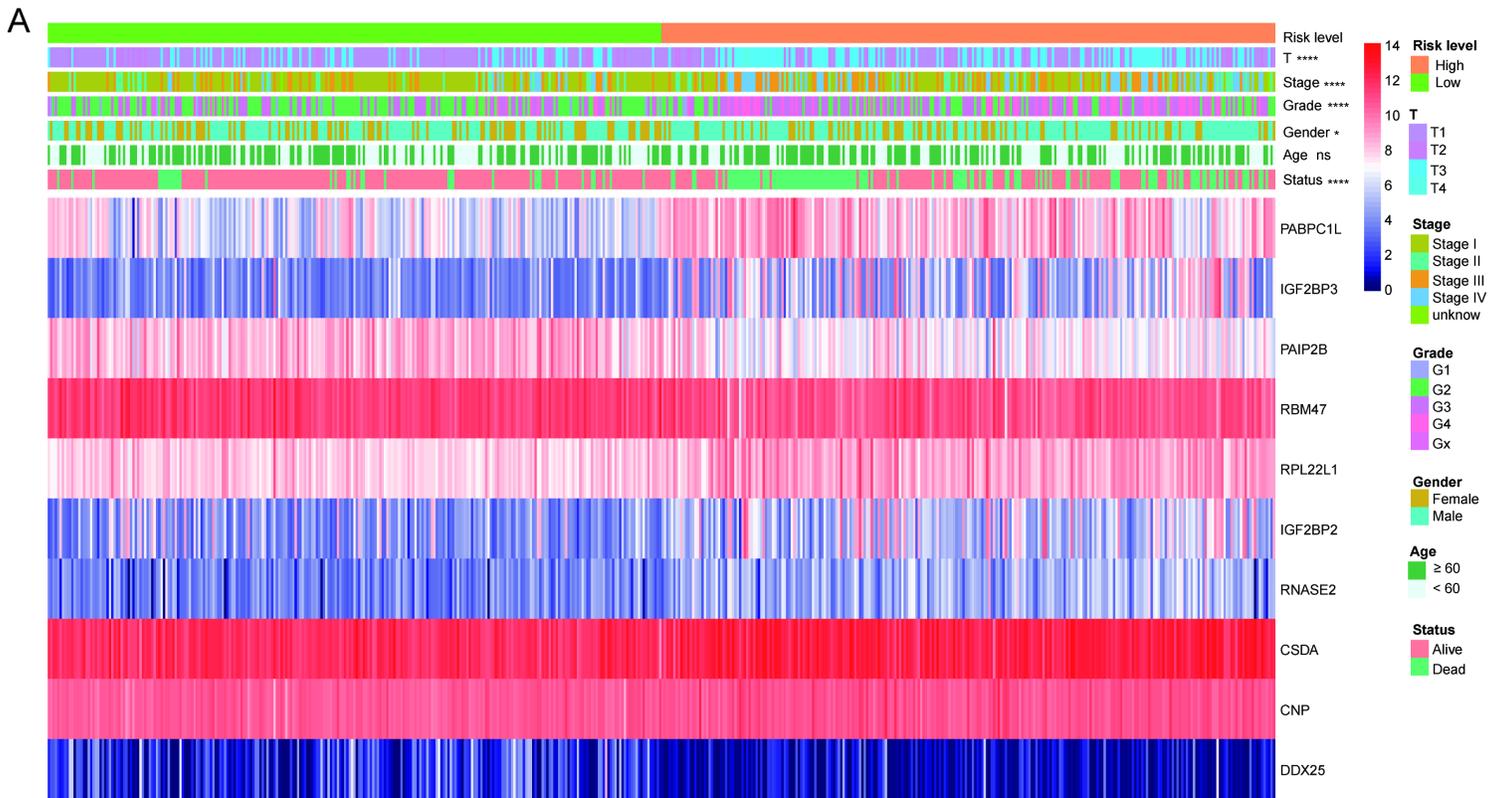


Figure 7

Relationship between clinical factors and risk score model. (A) Heat map showed the expression levels of selected genes in high-risk and low-risk patients. The clinical factors were compared between the high-risk and low-risk patients. Univariate (B) and multivariate (C) Cox regression analyses for evaluating the relationship between the risk score model and clinical factors. ns $P > 0.05$, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [FigureS1.tif](#)
- [TableS3.xlsx](#)
- [TableS1.xlsx](#)
- [TableS2.docx](#)
- [FigureS2.tif](#)