

# Genome-wide profiling identification of prognostic novel signatures in papillary renal cell carcinoma based on large-scale sequencing data

**Xi Tian**

Fudan University Shanghai Cancer Center <https://orcid.org/0000-0003-1965-0647>

**Wen-Hao Xu**

Fudan University Shanghai Cancer Center

**Aihetaimujiang anwaier**

Fudan University Shanghai Cancer Center

**Hong-Kai Wang**

Fudan University Shanghai Cancer Center

**Fang-Ning Wan**

Fudan University Shanghai Cancer Center

**Yu Zhu**

Fudan University Shanghai Cancer Center

**Da-Long Cao**

Fudan University Shanghai Cancer Center

**Yi-Ping Zhu**

Fudan University Shanghai Cancer Center

**Guo-Hai Shi**

Fudan University Shanghai Cancer Center

**Yuan-Yuan Qu** (✉ [quyy1987@163.com](mailto:quyy1987@163.com))

<https://orcid.org/0000-0003-1179-2800>

**Hai-Liang Zhang**

Fudan University Shanghai Cancer Center

**Ding-Wei Ye**

Fudan University Shanghai Cancer Center

---

## Primary research

**Keywords:** papillary renal cell carcinoma, PTTG1, biomarker, prognosis, protein-protein interaction

**Posted Date:** February 12th, 2020

**DOI:** <https://doi.org/10.21203/rs.2.23309/v1>

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

[Read Full License](#)

---

## Abstract

Background : This study aims to identify potential biomarkers with prognostic value in papillary renal cell carcinoma (PRCC) by combining protein interaction networks with gene expression profiles from multiply cohorts. Methods : Two microarray datasets were downloaded from the Gene Expression Omnibus (GEO) database and differentially expressed genes (DEGs) were identified based on standardized labeling information. The protein-protein interaction network (PPI) and functional annotations of DEGs was established and the modules were analyzed by using STRING and Cytoscape. Survival analysis of significant DEGs was analyzed by Kaplan-Meier comprehensive expression score in the Cancer Genome Atlas (TCGA) cohort. Receiver operating characteristic (ROC) curves were constructed to describe binary classifier value of genes using area under the curve (AUC) score. Additionally, immunohistochemical staining of PTTG1 protein was performed, and survival analysis was validated in Fudan University Shanghai Cancer Center (FUSCC) cohort. Results : A total of 473 DEGs and 38 functionally related hub genes were identified as candidate prognostic biomarkers. Eight genes, including BUB1B , CCNB1 , CCNB2 , MAD2L1 , TTK , CDC20 , PTTG1 and MCM, enrichment with cell cycle process were identified for further analysis. Statistical analysis of TCGA cohort indicates that expression level of the eight genes was higher in PRCC tumor tissues and was negatively correlated with patient's outcome. Significantly elevated PTTG1 expression and its negative correlativity with patient's outcome was validated in FUSCC cohorts. Conclusions : Expression levels of eight hub genes have strong prognostic value and may help better understand the potential carcinogenesis and develop targeted therapy strategies for PRCC.

## Background

Papillary renal cell carcinoma (PRCC), the second most common histologic subtype of renal cell carcinoma, originates from tubular epithelial cells, accounts for about 10%-15% of all renal tumors[1, 2]. The histological features of PRCC are fibrovascular cores with papillary arrangement of tumor cells[2]. PRCC can be divided into two subtypes: type I and type II. Type 1 PRCC was characterized by monolayer, light staining and basophilic small cells, often with rich foam macrophage infiltration, and type II showed high Fuhrman grade, eosinophilic cytoplasm and pseudostratified arrangement in the center of the nipple[3].

Recent studies have shown that type II PRCC is a heterogeneous tumor, which can be further subdivided into other subtypes according to the genetic and molecular composition of the tumor, reflecting different clinical course and prognosis[4]. It is reported that PRCC type II patients have worse outcome than clear cell renal cell carcinoma (ccRCC) [5, 6] and it is difficult to predict the prognosis of patients. Some genetic features of PRCC have been recognized such as mutations of MET, NF2, SETD2, and Nrf2 pathway genes[7–9]. However, these mutations were found in only 10–15% of PRCC tumors in these studies[10, 11]. In recent years, with better understanding of RCC molecular biology, targeted drugs and therapeutic effects have been improved[12].

With the rapid development of microarray technology and bioinformatics analysis, our understanding of differentially expressed genes (DEGs) and functional pathways related to the occurrence and development of PRCC has become more comprehensive. However, the rarity of this tumor has become an obstacle in identifying potential markers to distinguish PRCC and provide potential therapeutic targets. In this study, two mRNA microarray data sets were downloaded from Gene Expression Omnibus (GEO)[13] for analysis to obtain DEGs between cancer tissues and normal adjacent tissues. Subsequently, in order to further understand the molecular mechanism of tumorigenesis, gene ontology (GO)[14] and Kyoto genome encyclopedia (KEGG)[15] pathway enrichment analysis were carried out. The protein-protein interaction (PPI) network reveals the specific functions of all proteins and describes the importance of these interactions in biological processes, molecular functions, and signal transduction.

To determine the candidate biomarkers and their possible role in PRCC, this work focused on analyzing gene expression profiles, reveal potential biointeraction networks, and assessing prognostic value. We speculated that the oncogenic activity of important hub genes was associated with poor prognosis, which may be a potential therapeutic target for PRCC.

## Materials And Methods

### Original biological microarray data

Gene Expression Omnibus (GEO) is a public functional genomic database that stores high throughout gene expression data, chips and microarrays. The Original DNA microarray data were obtained from GEO, for patients with PRCC. Two chip datasets GSE48352, GSE26574 were downloaded from GEO (Affymetrix GPL16311 platform, Affymetrix GPL11433 platform, respectively). The corresponding genes transformed into a probe were converted into a symbol according to the annotation information on the platform.

### Screening and identification of DEGs

The DEGs between PRCC and non-cancerous samples were screened and identified by GEO2R. Delineating parameters such as P-values, Benjamini and Hochberg false discovery rate (FDR) and fold change were used to filter DEGs and applied to provide a balance between the discovery of statistically important genes and the limitations of false positives. A probe set that does not have a corresponding gene symbol or a gene that has multiple probe sets is knocked out or averaged.  $\log_{10}FC$  (fold change)  $>= 1$  or  $<=-1$  and  $P-value < 0.01$  were considered statistically significant.

### PPI network construction and module analysis

In this study, we used Search Tool for the Retrieval of Interacting Genes (STRING; <http://string-db.org>) (version 10.0) online database to predict PPI networks of DEGs and analyze functional interactions between proteins[16]. This may help to further understand the underlying mechanisms of the development and progression of PRCC. Cytoscape (version 3.5), an open source bioinformatics software platform, was designed to visualize molecular interaction networks[17]. Cytoscape's plug-in Molecular Complex Detection (MCODE) (version 1.4.2) can cluster a given network based on topology to find dense

connection areas[18]. The most important modules in the PPI network are selected as follows: MCODE Score > 24. Subsequently, KEGG and GO analysis of the genes in the module was performed using Database for Annotation, Visualization and Integrated Discovery (DAVID; <http://david.ncifcrf.gov>; Version 6.8)[19].

## Hub genes selection and analysis

After applying the MCODE plug-in, a network of the 38 genes and their co-expression genes was analyzed using cBioPortal (<http://www.cbioportal.org>) online platform[20]. ClueGO, a Cytoscape plug-in that could visualize the non-redundant biological terms for large clusters of genes in a functionally grouped network[21]. The biological processes from GO and KEGG pathway analysis of hub genes were visualized using ClueGO (version 2.5.3) and CluePedia (version 1.5.3), a functional extension of ClueGO[22, 23] (Bindea, Galon et al. 2013, G, J et al. 2013)[22, 23]. Heat map, based on hierarchical clustering algorithm, was constructed using the phenotype and gene expression profile of 323 samples in TCGA.

## Functional enrichment of DEGs

Biological properties such as biological processes (BP), molecular functions (MF), and cellular components (CC) were extracted from gene ontology (GO) enrichment analysis to determine the role of DEGs in PRCC. Kyoto Encyclopedia of Genes and Genomes (KEGG) is a database resource for understanding high-level functions and biological systems from large-scale molecular datasets generated by high-throughput experimental technologies. DAVID was applied to discern the role of development-related signaling pathways in PRCC. P-values < 0.05 are considered statistically significant. GO and KEGG enrichment was analyzed and displayed using bubble chart.

## Statistical analysis of TCGA cohort

Phenotype and expression profiles of hub genes in 323 PRCC patients from TCGA were analyzed and displayed to predict prognostic value. The expression levels of each gene were divided into three groups: high, medium and low. Subsequently, to compare the outcome of high group and low group, the univariate survival analysis of the 8 hub genes was performed using Kaplan-Meier curve. Univariate and multivariate analyses were performed with Cox logistic regression models to find independent variables, including age (ref.  $\geq 60$  years), gender (ref. Male), laterality (ref. Right), pT stage (ref. T1-T2), pN stage (ref. N0), pM stage (ref. M0), AJCC stage (ref. I-II) and gene expression (ref. Low). The results of COX regression analyses are displayed by forest plot. Receiver operating characteristic (ROC) curves were constructed to describe binary classifier value of genes using area under the curve (AUC) score. Cutoff values were taken using X-tile software[24].

## Immunohistochemical (IHC) staining and evaluation

Rabbit anti-PTTG1 monoclonal antibody was used (ab128040, Abcam, USA). The positive or negative staining of a certain protein in a FFPE slide was independently evaluated by two experienced pathologists and determined as follows. The overall IHC score from 0 to 12 was evaluated according to the multiply of the staining intensity and extent score, as previously described[25].

# Statistical analysis of FUSCC cohort and potential networks of the hub genes

According to the IHC score, the patients were divided into two groups: high expression group and low expression group of PTTG1. Correlation analyses between the expression of PTTG1 and clinicopathological features were carried out by chi-square test. In order to explore the factors related to prognosis, age at surgery (ref.  $\geq 60$  years), BMI, tumor size, laterality (ref. Right), pT stage (ref. T1-T2), pN stage (ref. N0), pM stage (ref. M0), AJCC stage (ref. I-II), Furhman grade (Ref. 1-2) and gene expression (ref. Low) were analyzed by Cox regression. R software (Version 3.4.2) was implemented to describe potential networks associated with functional annotations of the hub genes.

## Results

This study is divided into three phases. In the first phase, we used the information in GEO platform to evaluate DEGs; in the second phase, we constructed a PPI network and evaluated their interactions specificity based on co-expression and functional annotation. In the third phase, phenotype and expression profiles of hub genes in 323 PRCC patients from TCGA were analyzed and displayed to predict prognostic value.

## Identification of DEGs in PRCC

After normalizing the microarray results, we identified DEGs, including 1,270 probes in GSE48352, 826 probes in GSE26574. As shown in the Venn diagram, the overlap in both two datasets contained 473 differentially expressed genes between tumor tissues and adjacent normal tissues (Fig. 1A).

## PPI network construction and module analysis

PPI network of DEGs was built (Fig. 1B) and the most important modules were identified using Cytoscape plugin (Fig. 1C). Functional analysis of the 38 genes involved in the module was performed using DAVID. The functional analyses of 38 genes involved in this module were analyzed using DAVID. Enrichment profiles showed that hub genes in this module were primarily enriched in sister chromatid segregation (49.37%), mitotic nuclear division (13.92%), anaphase-promoting complex-dependent catabolic process (11.39%; Supplementary Fig. 1). Through functional analysis, we subjectively selected eight genes related to the cell cycle and highlighted them in the PPI network of 38 hub genes.

## Hub gene selection and analysis

After statistical selection, 38 functionally related hub genes of network includes PTTG1, CDC20, TPX2, AURKA, MAD2L1, TRIP13, CCNB2, KIF20A, CKS2, CCNB1, UBE2C, BUB1B, ZWINT, RACGAP1, TTK, HMMR, TOP2A, TYMS, TK1, KIF4A, PRC1, PBK, NUSAP1, EZH2, RFC4, DTL, CENPK, MCM5, MELK, UBE2T, RRM2, CEP55, KIAA0101, RAD51AP1, CDCA7, CKAP2, CDKN3 and STIL. A network of the 38 genes and their co-expression genes was visualized and displayed (Fig. 2A). The biologic process and KEGG enrichment

analysis of the hub genes were shown in Fig. 2B. The selected 8 genes are as follow: BUB1B, CCNB1, CCNB2, MAD2L1, TTK, CDC20, PTTG1, MCM5. Using hierarchical clustering algorithm, the heat map showed that the highlighted 8 genes relating to cell cycle could essentially distinguish tumor samples from adjacent normal samples (Fig. 2C). To make it clearer, we used DAVID in bubble chart to show the results of Functional and pathway enrichment analysis. GO analysis showed that the biological process changes of 473 DEGs were significantly enriched in oxidation-reduction process and positive regulation of cell migration; Changes in cellular components of 473 DEGs were mostly enriched in the extracellular exosome and plasma membrane; Changes in KEGG pathway of 473 DEGs were mainly enriched in metabolic pathway; Changes in molecular functions of 473 DEGs were mainly enriched in protein binding (Fig. 2D). After selection of the 38 hub genes, DAVID was used again in bubble chart. GO analysis results showed that changes in biological processes of 38 DEGs were significantly enriched in mitotic cytokinesis; Changes in cellular components of 38 DEGs were mostly enriched in the nucleus and cytoplasm; Changes in KEGG pathway of 38 DEGs were mainly enriched in cell cycle and oocyte meiosis; Changes in molecular functions of 38 DEGs were mainly enriched in ATP binding and Chromatin binding (Fig. 2E). The functional enrichment results of 38 DEGs are listed in Table 1, modules with both P value < 0.05 and FDR < 0.1 were considered significant.

Table 1  
GO and KEGG pathways enrichment analysis of DEGs in the most significant module

Term	Description	Count in gene set	P value
cfa04110	Cell cycle	8	9.40E-09
cfa04114	Oocyte meiosis	5	1.26E-04
GO:0000281	Mitotic cytokinesis	4	1.86E-05
GO:0005524	ATP binding	13	1.23E-06

GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; DEGs: differentially expressed genes

## Gene expression profiles and survival analyses in ccRCC patients from TCGA cohort

The 8 hub gene expression profiles were relatively high in tumor tissues compared to normal tissues (Fig. 3). The OS of PRCC patients with elevated 8 gene expression was significantly worse. Apart from MCM5, PRCC patients with other 7 hub gene expression showed markedly significant worse DFS (Fig. 4A). Univariate Cox regression analyses was carried out, expression levels of the eight hub genes and traditional prognostic factors such as AJCC stage and pTNM stage were considered significantly related to OS ( $p < 0.05$ ; Fig. 4B). In multivariate Cox regression analysis models, traditional prognostic factors were not always relevant to OS. Expression levels of the eight hub genes were still significantly correlated with OS ( $p < 0.05$ ; Fig. 5A, C, E, G, I, K, M, O) and the AUC indices for the TCGA-OS were from 0.686 to 0.819 (Fig. 5B, D, F, H, J, L, N, P). After integrating all the significant clinicopathological parameters and

gene expression profiles in the Cox regression models, we generated eight formulas for OS (Supplementary table 1).

## **PTTG1 expression and clinicopathological characteristics in XXX PRCC patients from the FUSCC cohort**

Although the AUC of MCM5 is the highest in the eight genes, we did not choose to verify the prognostic value of MCM5 due to the little correlation with PFS. As the AUC of PTTG1 ranked in second place and it correlated significantly with both OS and PFS, we chose it to make further verification. In order to evaluate the expression level of PTTG1 protein in FUSCC tumor samples, we performed IHC staining and found that the density and intensity of PTTG1 expression in PRCC tissues in the FUSCC cohort were significantly higher than those in adjacent normal renal tissues (Fig. 6A-C). High expression group of PTTG1 showed poorer OS ( $HR = 2.673, p = 0.001$ ) and PFS ( $HR = 2.683, p < 0.001$ ) than low expression group of PTTG1 in FUSCC cohort (Fig. 6D-E). Correlation analyses between the expression of PTTG1 and clinicopathological features were carried out. Increased PTTG1 expression in PRCC patients significantly correlated with advanced pT ( $p = 0.049$ ), pN ( $p = 0.014$ ), and pM stage ( $p = 0.017$ ), AJCC stage ( $p = 0.007$ ) and Furhman grade ( $p = 0.005$ ) in the FUSCC cohort (Table 2).

Table 2

Clinicopathological characteristics in relation to PTTG1 expression status in PRCC patients from FUSCC cohort.

Variable	Entire group (n = 126)	PTTG1 expression		$\chi^2$	P value
		Low expression (n = 63)	High expression (n = 63)		
Age at surgery (y, median $\pm$ SD)		57.0 $\pm$ 11.7	57.0 $\pm$ 13.2		
BMI (kg/m <sup>2</sup> , median $\pm$ SD)		22.9 $\pm$ 3.4	22.5 $\pm$ 3.2		
Tumor size (cm, median $\pm$ SD)		4.0 $\pm$ 2.5	5.0 $\pm$ 3.0		
OS time (month, median $\pm$ SD)		54.0 $\pm$ 35.6	42.0 $\pm$ 35.2		
PFS time (month, median $\pm$ SD)		54.0 $\pm$ 38.7	36.0 $\pm$ 39.7		
Sex (n, %)					
Male	85, 67.5%	45, 71.4%	40, 63.5%		
Female	41, 32.5%	18, 28.6%	23, 36.5%	0.904	0.342
Laterality (n, %)					
Left	71, 56.3%	39, 61.9%	32, 50.8%		
Right	55, 43.7%	24, 38.1%	31, 49.2%	1.581	0.209
T stage at presentation (n, %)					
T1-T2	90, 71.4%	50, 79.4%	40, 63.5%		
T3-T4	36, 28.6%	13, 20.6%	23, 36.5%	3.889	0.049
N stage at presentation (n, %)					
N0	101, 80.2%	56, 88.9%	45, 71.4%		
N1	25, 19.8%	7, 11.1%	18, 28.6%	6.038	0.014

Abbreviations: PRCC, papillary renal cell carcinoma; FUSCC, Fudan University Shanghai Cancer Center; BMI, body mass index; AJCC, American Joint Committee on Cancer;

\*P value less than 0.05 was considered as statistically significant and marked in bold.

Variable	Entire group (n = 126)	PTTG1 expression		$\chi^2$	P value
		Low expression (n = 63)	High expression (n = 63)		
M stage at presentation (n, %)					
M0	99, 78.6%	55, 87.3%	44, 69.8%		
M1	27, 21.4%	8, 12.7%	19, 30.2%	5.704	0.017
AJCC stage					
I-II	73, 57.9%	44, 69.8%	29, 46.0%		
III-IV	53, 42.1%	19, 30.2%	34, 54.0%	7.327	0.007
Furhman grade (n, %)					
1-2	27, 21.4%	20, 31.7%	7, 11.1%		
3-4	99, 78.6%	43, 68.3%	56, 88.9%	7.966	0.005

Abbreviations: PRCC, papillary renal cell carcinoma; FUSCC, Fudan University Shanghai Cancer Center; BMI, body mass index; AJCC, American Joint Committee on Cancer;

\*P value less than 0.05 was considered as statistically significant and marked in bold.

## Cox regression analyses of FUSCC cohorts and potential networks of the hub genes

In univariate Cox regression analysis models, pTNM stage, AJCC stage, and Furhman grade, tumor size, BMI, were significantly relevant to PFS ( $p < 0.05$ ; Table 3) and OS ( $p < 0.05$ ; Table 3). Importantly, PTTG1 expression markedly correlated with poor PFS (HR = 2.46,  $p < 0.01$ ) and poor OS (HR = 2.90,  $p < 0.01$ ). The potential networks associated with functional annotations of the hub genes include inhibition, protein interaction, ubiquitination, phosphorylation, activating expression, activation, indirect relation, dephosphorylation (Fig. 7).

Table 3

Univariate Cox regression analyses of OS and PFS in 126 enrolled PRCC patients from FUSCC cohort.

	OS		PFS	
Covariates	HR (95%CI)	P value	HR (95%CI)	P value
Age at surgery	1.02 (1.00, 1.05)	0.08	1.02 (1.00, 1.04)	0.17
BMI (kg/m <sup>2</sup> )	0.83 (0.75, 0.92)	< 0.01	0.87 (0.79, 0.95)	< 0.01
Tumor size (cm)	1.17 (1.07, 1.29)	< 0.01	1.17 (1.08, 1.28)	< 0.01
Laterality (Ref. left)	0.75 (0.41, 1.35)	0.33	0.97 (0.55, 1.73)	0.92
T stage (T1-T2)	1.88 (1.02, 3.45)	0.04	1.90 (1.05, 3.43)	0.03
N stage (Ref. N0)	5.37 (2.91, 9.91)	< 0.01	5.75 (3.18, 10.40)	< 0.01
M stage (Ref. M0)	14.21 (7.24, 27.86)	< 0.01	14.36 (7.50, 27.47)	< 0.01
AJCC stage	7.52 (3.67, 15.43)	< 0.01	6.30 (3.25, 12.23)	< 0.01
Furhman grade (Ref. 1–2)	1.43 (0.64, 3.20)	0.34	1.23 (0.59, 2.54)	0.58
PTTG1 expression (Ref. Low)	2.90 (1.52, 5.55)	< 0.01	2.46 (1.35, 4.51)	< 0.01

Abbreviations: PFS: progression free survival; PRCC: papillary renal cell carcinoma; FUSCC, Fudan University Shanghai Cancer Center; HR: hazard ratio; CI: confidence interval; BMI: body mass index;

\*P value less than 0.05 was considered as statistically significant and marked in bold.

## Discussion

The concept of PRCC was first proposed by Mancilla-Jimenez in 1976, thirty-four cases of RCC showed papillary structures. Of these, 85.3% PRCC patient have a better prognosis than other types of RCC[26]. Since PRCC is relatively rare in clinical practice and has been rarely studied, major molecular mechanisms in the pathogenesis are poorly understood. Therefore, potential biomarkers for efficient diagnosis and treatment are urgently needed. In this study A total of 473 DEGs and 38 hub genes were identified by microarray data analysis. Among the 38 hub genes, 8 genes relating to cell cycle including BUB1B, CCNB1, CCNB2, MAD2L1, TTK, CDC20, PTTG1, MCM5 were subjectively selected. After statistical analysis, the 8 genes showed clear prognostic value.

BUB1B (spindle detection point protein, also known as BUBR1) is an important functional protein at the detection point of mitosis and the change of BUB1B expression plays an important role in tumorigenesis and progression[27]. Studies have found that BUB1B is overexpressed in kidney cancer and breast cancer, its mutation and overexpression is strongly correlated with Chromosomal instability[27, 28]. Yet the prognostic value of BUB1B in PRCC has been rarely reported.

As a member of the cell cycle family, CCNB1 is one of the key factors related to cell detection points[29, 30]. Currently, Cyclin B1 overexpression has been found in a variety of human tumors, such as esophageal cancer, non-small cell lung cancer, tongue cancer, and is related to tumor grade, differentiation, invasion and metastasis and prognosis[31]. Thus, there is enough evidence to doubt the role of CCNB1 in human PRCC as an oncogene.

It has been reported that CCNB2 is highly expressed in tumor tissues, such as breast cancer[32], adrenal cortical carcinoma[33], colorectal adenocarcinoma[34]and pituitary adenoma[35]. It has also been reported that serum circulating CCNB2 mRNA level in cancer patients is significantly higher than that in normal population and benign diseases[36]. Thus, CCNB2 may have play an important role in the generation and development of PRCC.

It has been shown that interrupting the function of MAD2L1 in mammalian cells can affect the process of spindle examination and lead to the development of aneuploid cells or tumors. The deletion of MAD2L1 gene can cause chromosome instability and drive the development of tumor. In mouse models, deletion of MAD2L1 gene can cause liver and lung cancer[37].

TTK is the basic component of spindle assembly checkpoint(SAC), it plays an important role in the replication of mitotic centrosomes and the correct separation of chromosomes[38]. In order to maintain the division and proliferation of tumor cells, TTK was highly expressed in tumor cells to maintain the normal function of SAC. After inhibiting the function of TTK, SAC is damaged, errors in mitotic metaphase cannot be detected, chromosomes cannot be separated into daughter cells on average, and heteroploidy is further increased, exceeding a certain threshold will cause tumor cell apoptosis, so TTK can serve as an effective anti-tumor target[39, 40].

Multiple studies have shown that CDC20 could degrade several important substrate factors to regulate cell cycle progression including Securin[41], Cyclin A[42, 43], p21[44] and Mcl-1[45]. The protein p21 is considered to be an effector of various tumor inhibition signaling pathways, partly deactivate Cyclin dependent kinases to promote anti - tumor proliferation. Downregulation of p21 expression was detected in a variety of human malignancies, so CDC20 may play its carcinogenic role in part by degrading tumor suppressor protein p21. Thus, the generation of PRCC may be caused part by CDC20.

MCMs protein is closely related to the cell cycle. It is a promoter of DNA replication and plays a key role in regulating cells going from G0 phase to S phase[46]. As one of the MCM protein families, MCM5 protein is closely related to cell proliferation. The expression of MCMs protein can be indirectly understood through the detection of MCM5. Williams et al. reported the application of MCM5 protein antibody to diagnose abnormal prodromal malignant cells in pap cervical smear, and found that the MCM had high sensitivity and specificity in detecting predromal malignant cells in cervical smear[47]. Going has proposed this conclusion in esophageal tissues[48]. These results indicate that MCM5 protein has some predictive value.

PTTG1 is a tumor transforming gene, which can cause cell transformation without the participation of any auxiliary gene, and is closely related to the occurrence of many tumors[49]. PTTG1 has been identified as an oncogene. The expression level of PTTG1 is closely related to tumor formation, angiogenesis and metastasis[50]. In our study, we detected the expression level of PTTG1 protein in PRCC from FUSCC cohort by using immunohistochemical staining and we found that PTTG1 protein increased in tumor tissues, and higher expression of PTTG1 is significantly relevant to both OS and PFS.

There are still several limitations in this study. First, the data utilized in the study consisted of unbalanced PRCC and control normal samples, which were restricted in quantity and downloaded from the Gene Expression Omnibus database, not generated by us. The chip data contains relatively small PRCC samples in public database and only 323 patients were enrolled from TCGA cohort with corresponding RNA sequence. Second, we only verified the expression and prognostic values of PTTG1 in FUSCC cohort and the potential mechanism of signaling pathway in PRCC is not clear, while a series of functional annotations and enrichment analyses have been carried out. The detailed mechanism between the eight genes and PRCC need to be further studied.

## Conclusion

In conclusion, DEGs and hub genes may be involved in poor prognosis and recurrence of PRCC in this study. The transcription profiles of BUB1B, CCNB1, CCNB2, MAD2L1, TTK, CDC20, PTTG1 and MCM5 are predictive and may contribute to a better understanding of the potential carcinogenesis and progression of PRCC. Further research needs to elucidate the molecular mechanism and signaling pathway changes of these genes in PRCC.

## Abbreviations

RCC

renal cell carcinoma

PRCC

papillary renal cell carcinoma

GEO

Gene Expression Omnibus

DEGs

differentially expressed genes

PPI

protein-protein interaction network

TCGA

The Cancer Genome Atlas

ROC

Receiver operating characteristic

AUC

area under the curve

FUSCC

Fudan University Shanghai Cancer Center

ccRCC

clear cell renal cell carcinoma

GO

gene ontology

KEGG

Kyoto Encyclopedia of Genes and Genomes

FDR

false discovery rate

FC

fold change

MCODE

Molecular Complex Detection

DAVID

Database for Annotation, Visualization and Integrated Discovery

BP

biological processes

MF

molecular functions

CC

cellular components

## Declarations

**Ethics approval and consent to participate:** The Ethics approval and consent to participate of the current study was approved and consented by the ethics committee of Fudan University Shanghai Cancer center.

**Availability of data and material:** The datasets during and/or analyzed during the current study available from the corresponding author on reasonable request.

**Conflict of interests:** The authors declare no competing interests.

**Funding:** This work is supported by Grants from the National Natural Science Foundation of China (No.81772706 and No.81802525) and National Key Research and Development Project

**Authors' contributions:** The work presented here was carried out in collaboration among all authors. YDW, ZHL and QYY defined the theme of the study and discussed analysis, interpretation and presentation. TX and XWH drafted the manuscript, analyzed the data, developed the algorithm, and explained the results. Aihetaimujiang, WHK and WFN, participated in the collection of relevant data and helped draft the manuscript. ZY and CDL helped to perform the statistical analysis. ZYP and SGH helped revise the manuscript and provided guiding suggestions. All the authors read and approved the final manuscript.

**Acknowledgements:** We thank the TCGA and GEO (ID: XXX) databases for providing PRCC gene expression profiles.

## References

1. Jemal, A., et al., *Cancer Statistics, 2008*. 2010. **51**(1): p. 15-36.
2. Kuroda, N., ., et al., *Review of papillary renal cell carcinoma with focus on clinical and pathobiological aspects*. 2003. **18**(2): p. 487-94.
3. B, D., E.J.J.M.p.a.o.j.o.t.U. States, and I. Canadian Academy of Pathology, *Papillary renal cell carcinoma: a clinicopathologic and immunohistochemical study of 105 tumors*. 1997. **10**(6): p. 537-44.
4. M, A., A.-B. IA, and A.H.T.J.A.i.a. pathology, *Papillary Renal Cell Carcinoma (PRCC): An Update*. 2019. **26**(2): p. 124-132.
5. LJ, M., G. AB, and W.L.J. Cancer, *Renal cell carcinoma. Prognostic significance of morphologic parameters in 121 cases*. 1988. **61**(8): p. 1639-51.
6. B, D., et al., *Morphologic typing of papillary renal cell carcinoma: comparison of growth kinetics and patient survival in 66 cases*. 2001. **32**(6): p. 590-5.
7. Schmidt, L., et al., *EARLY ONSET HEREDITARY PAPILLARY RENAL CARCINOMA: GERMLINE MISSENSE MUTATIONS IN THE TYROSINE KINASE DOMAIN OF THE MET PROTO-ONCOGENE*. **172**(4): p. 1256-1261.
8. Carole, S., et al., *Targeting loss of the Hippo signaling pathway in NF2-deficient papillary kidney cancers*. **9**(12).
9. M, K., et al., *Recurrent chromosomal gains and heterogeneous driver mutations characterise papillary renal cancer evolution*. 2015. **6**: p. 6336.
10. Fay, A.P, S. Sabina, and T.K. Choueiri, %J Clinical Cancer Research An Official Journal of the American Association for Cancer Research, *MET as a target in papillary renal cell carcinoma*. 2014. **20**(13): p. 3361-3.

11. S. D., et al., *Spectrum of diverse genomic alterations define non-clear cell renal carcinoma subtypes*. 2015. **47**(1): p. 13-21.
12. A. S.-G., et al., *Systemic treatment of renal cell cancer: A comprehensive review*. 2017. **60**(undefined): p. 77-89.
13. R. E., D. M, and L.A.J.N.a. research, *Gene Expression Omnibus: NCBI gene expression and hybridization array data repository*. 2002. **30**(1): p. 207-10.
14. Ashburner, M., ., et al., *Gene ontology: tool for the unification of biology*. *The Gene Ontology Consortium*. 2000. **25**(1): p. 25-9.
15. symposium, K.M.J.N.F., *The KEGG database*. 2002. **247**(undefined): p. 91-101; discussion 101-3, 119-28, 244-52.
16. A. F., et al., *STRING v9.1: protein-protein interaction networks, with increased coverage and integration*. 2013. **41**(null): p. D808-15.
17. ME, S., et al., *Cytoscape 2.8: new features for data integration and network visualization*. 2011. **27**(3): p. 431-2.
18. WP, B., et al., *MultiContrast Delayed Enhancement (MCODE) improves detection of subendocardial myocardial infarction by late gadolinium enhancement cardiovascular magnetic resonance: a clinical validation study*. 2012. **14**(undefined): p. 83.
19. DW, H., et al., *The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists*. 2007. **8**(9): p. R183.
20. E, C., et al., *The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data*. 2012. **2**(5): p. 401-4.
21. G, B., et al., *ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks*. 2009. **25**(8): p. 1091-3.
22. Bindea, G., J. Galon, and B. Mlecnik, *CluePedia Cytoscape plugin: pathway insights using integrated experimental and in silico data*. Bioinformatics, 2013. **29**(5): p. 661-3.
23. G, B., G. J, and M.B.J. Bioinformatics, *CluePedia Cytoscape plugin: pathway insights using integrated experimental and in silico data*. 2013. **29**(5): p. 661-3.
24. RL, C., D.-F. M, and R.D.J.C.c.r.a.o.j.o.t.A.A.f.C. Research, *X-tile: a new bio-informatics tool for biomarker assessment and outcome-based cut-point optimization*. 2004. **10**(21): p. 7252-9.
25. J, W., et al., *Elevated MRE11 expression associated with progression and poor outcome in prostate cancer*. 2019. **10**(18): p. 4333-4340.
26. R, M.-J., S. RJ, and B.R.J. Cancer, *Papillary renal cell carcinoma: a clinical, radiologic, and pathologic study of 34 cases*. 1976. **38**(6): p. 2469-80.
27. Myrie, K.A., et al., *Mutation and expression analysis of human BUB1 and BUB1B in aneuploid breast cancer cell lines*. 2000. **152**(2): p. 193-199.
28. Scantu, M., et al., *Genomic instability and increased expression of BUB1B and MAD2L1 genes in ductal breast carcinoma*. 2007. **254**(2): p. 298-307.

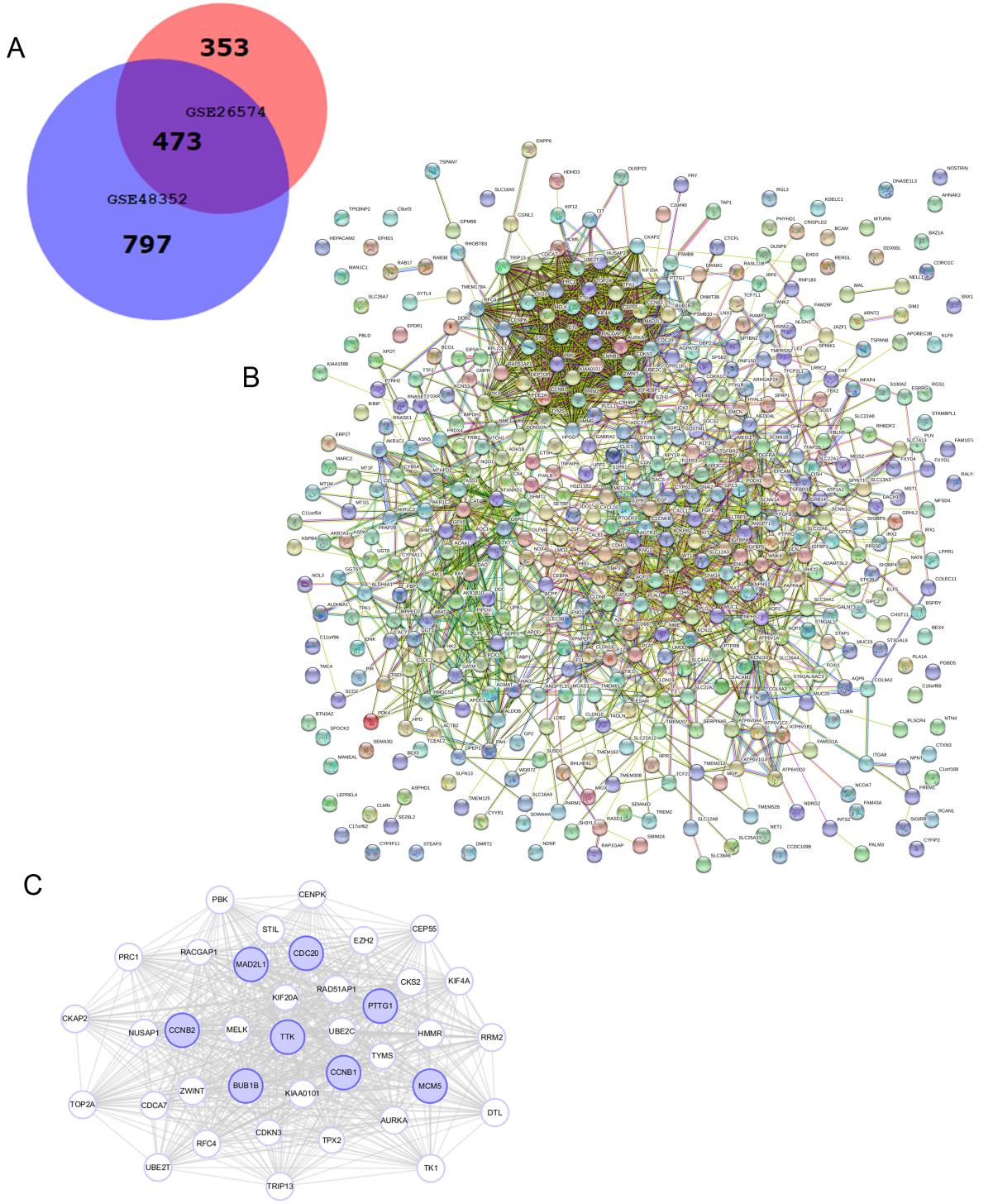
29. Gavet, O. and J.J.D.C. Pines, *Progressive activation of CyclinB1-Cdk1 coordinates entry to mitosis*. 2010. **18**(4): p. 533-543.
30. Yuan, J., et al., *Cyclin B1 depletion inhibits proliferation and induces apoptosis in human tumor cells*. 2004. **23**(34): p. 5843-5852.
31. Harada, H., et al., *Cyclin B1 is useful to predict occult cervical lymph node metastases in tongue carcinoma*. 2006. **25**(3): p. 351.
32. Shubbar, E., et al., *Elevated cyclin B2 expression in invasive breast carcinoma is associated with unfavorable clinical outcome*. 2013. **13**(1): p. 1-10.
33. Fernandez-Ranvier, G.G., et al., *Identification of biomarkers of adrenocortical carcinoma using genomewide gene expression profiling*. 2008. **143**(9): p. 841-846.
34. Seon-Hwa, P., et al., *NF-Y-dependent cyclin B2 expression in colorectal adenocarcinoma*. 2007. **13**(3): p. 858-867.
35. Ivana, D.M., et al., *HMGA proteins up-regulate CCNB2 gene in mouse and human pituitary adenomas*. 2009. **69**(5): p. 1844-50.
36. Mo, M.L., et al., *Use of serum circulating CCNB2 in cancer surveillance*. 2010. **25**(4): p. 236-242.
37. Loren, M., et al., *Complete loss of the tumor suppressor MAD2 causes premature cyclin B degradation and mitotic failure in human somatic cells*. 2004. **101**(13): p. 4459-4464.
38. Liu, X. and M.J.A.R.o.B. Winey, *The MPS1 family of protein kinases*. 2012. **81**(81): p. 561.
39. Jemaà, M., et al., *Whole-genome duplication increases tumor cell sensitivity to MPS1 inhibition*. 2016. **7**(1): p. 885-901.
40. Marieke, A., L. Spiros, and N.C. Turner, %J Current Opinion in Pharmacology, *Tumour selective targeting of cell cycle kinases for cancer treatment*. 2013. **13**(4): p. 529-535.
41. Chu, C.M.J.J.o.G. and Hepatology, *Chu, C.M. Natural history of chronic hepatitis B infection in adults with emphasis on the occurrence of cirrhosis and hepatocellular carcinoma*. *J. Gastroenterol. Hepatol.* 15 (Suppl) E25-E30. 2000. **15 Suppl**(Suppl): p. E25-30.
42. Geley, S., et al., *Anaphase-Promoting Complex/Cyclosome-Dependent Proteolysis of Human Cyclin a Starts at the Beginning of Mitosis and Is Not Subject to the Spindle Assembly Checkpoint*. 2001. **153**(1): p. 137-148.
43. Ohtoshi, A., ., et al., *Human p55(CDC)/Cdc20 associates with cyclin A and is phosphorylated by the cyclin A-Cdk2 complex*. 2000. **268**(2): p. 530-534.
44. Amador, V., et al., *APC/CCdc20 Controls the Ubiquitin-Mediated Degradation of p21 in Prometaphase*. 2007. **27**(3): p. 462-473.
45. Harley, M.E., H.S. Allan LASanderson, and P.R.J.E.J. Clarke, *Phosphorylation of Mcl-1 by CDK1-cyclin B1 initiates its Cdc20-dependent destruction during mitotic arrest*. 2014. **29**(14): p. 2407-2420.
46. Tye, B.K.J.A.R.o.B., *MCM Proteins in DNA Replication*. 1999. **68**(68): p. 649-686.
47. Williams, G.H., et al., *Diagnosis of oesophageal cancer by detection of minichromosome maintenance 5 protein in gastric aspirates*. 2004. **91**(4): p. 714-719.

48. Going, J.J., et al., *Aberrant expression of minichromosome maintenance proteins 2 and 5, and Ki-67 in dysplastic squamous oesophageal epithelium and Barrett's mucosa*. 2002. **50**(3): p. 373-7.
49. Zhang, X., ., et al., *Structure, expression, and function of human pituitary tumor-transforming gene (PTTG)*. 1999. **13**(1): p. 156-166.
50. McCabe, C.J., et al., *Vascular endothelial growth factor, its receptor KDR/Flk-1, and pituitary tumor transforming gene in pituitary tumors*. 2002. **87**(9): p. 4238-4244.

## Supplementary Figure 1

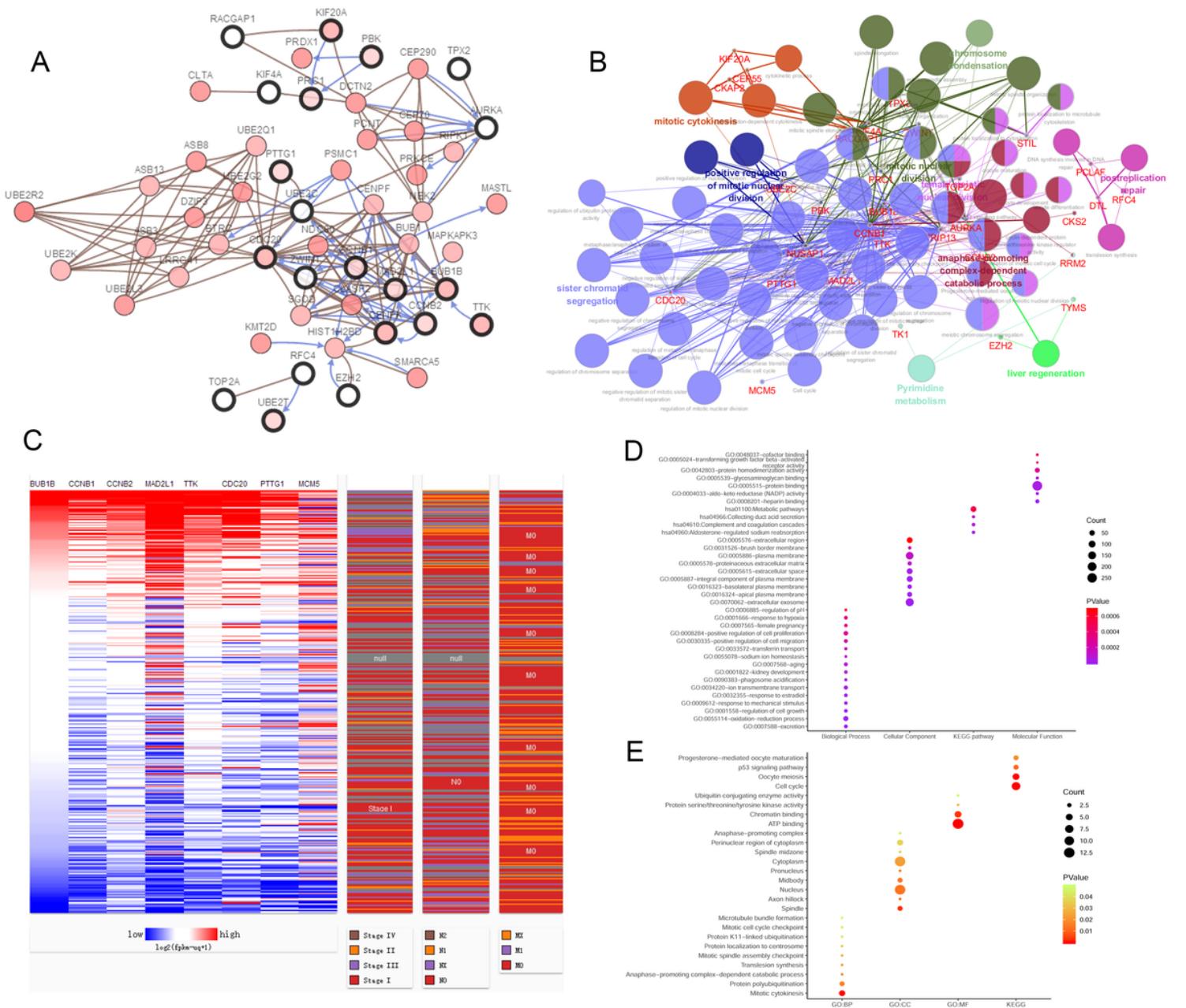
**Supplementary figure1.** Enrichment profiles showed that hub genes in this module were primarily enriched in sister chromatid segregation (49.37%), mitotic nuclear division (13.92%), anaphase-promoting complex-dependent catabolic process (11.39%)

## Figures



**Figure 1**

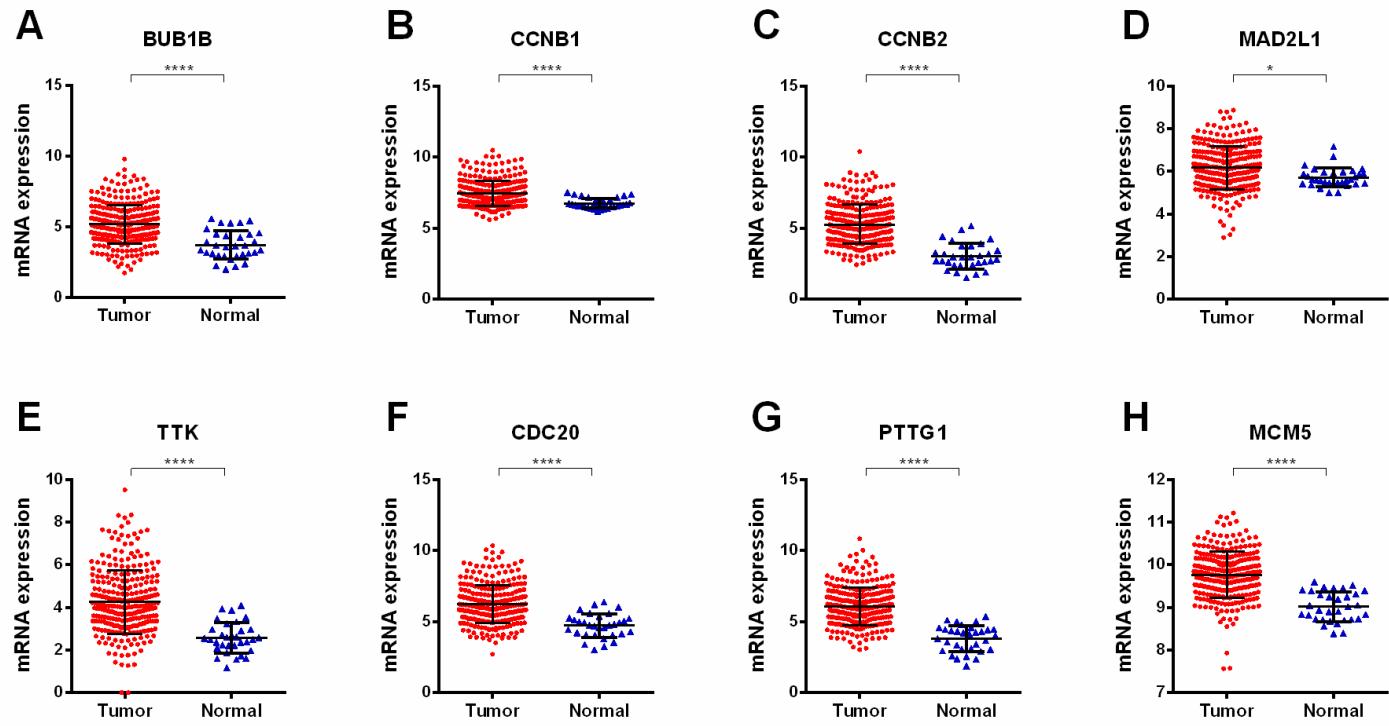
Venn diagram, PPI network and the most significant module of DEGs. (A) DEGs were selected with a fold change  $>2$  and P-value  $<0.01$  among the mRNA expression profiling chip datasets GSE48352 and GSE26574. The 2 datasets show an overlap of 473 genes in the Venn diagram. (B) The PPI network of DEGs was constructed using STRING. (C) The most significant module was obtained from PPI network with 38 nodes. Significant edges are related to cell cycle and they are marked in light blue.



**Figure 2**

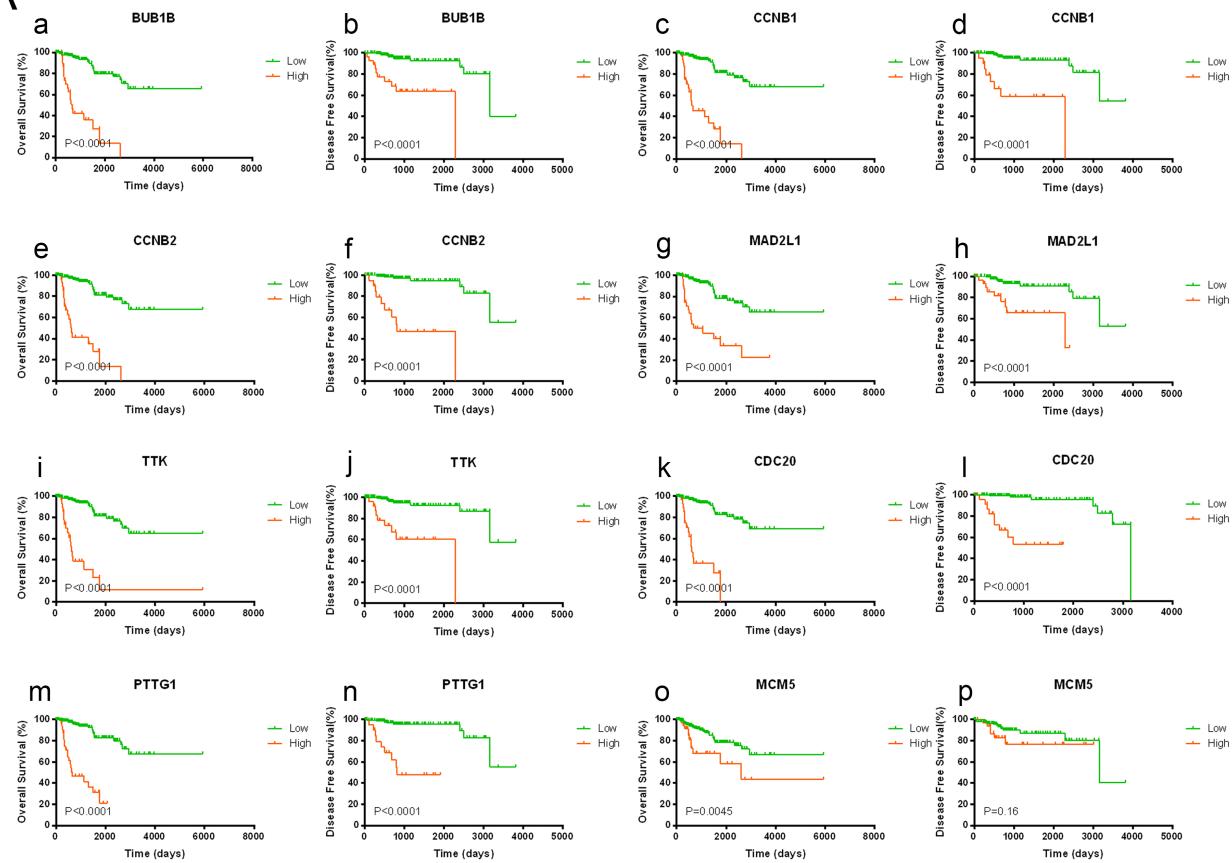
Interaction network and biological process analysis of the hub genes. (A) Hub genes and their co-expression network were analyzed using cBioPortal. Nodes with bold black outline represent hub genes. Nodes with thin black outline represent the co-expression genes. (B) The biological process analysis of hub genes was constructed using ClueGO. Different color of nodes refers to the functional annotation of ontologies. Corrected P-value  $< 0.05$  was considered statistically significant. (C) Hierarchical partitioning of 38 hub genes was obtained from DNA microarrays. It represent the level of expression of 38 genes across a number of comparable samples with high expression samples marked in red and low in blue. (D) GO analysis showed that the biological process changes of 473 DEGs were significantly enriched in oxidation-reduction process and positive regulation of cell migration; Changes in cellular components of 473 DEGs were mostly enriched in the extracellular exosom and plasma membrane; Changes in KEGG

pathway of 473 DEGs were mainly enriched in metabolic pathway; Changes in molecular funtions of 473 DEGs were mainly enriched in protein binding. (E) GO analysis results showed that changes in biological processes of 38 DEGs were significantly enriched in mitotic cytokinesis; Changes in cellular components of 38 DEGs were mostly enriched in the nucleus and cytoplasm; Changes in KEGG pathway of 38 DEGs were mainly enriched in cell cycle and oncyte meiosis; Changes in molecular funtions of 38 DEGs were mainly enriched in ATP binding and Chromatin binding.

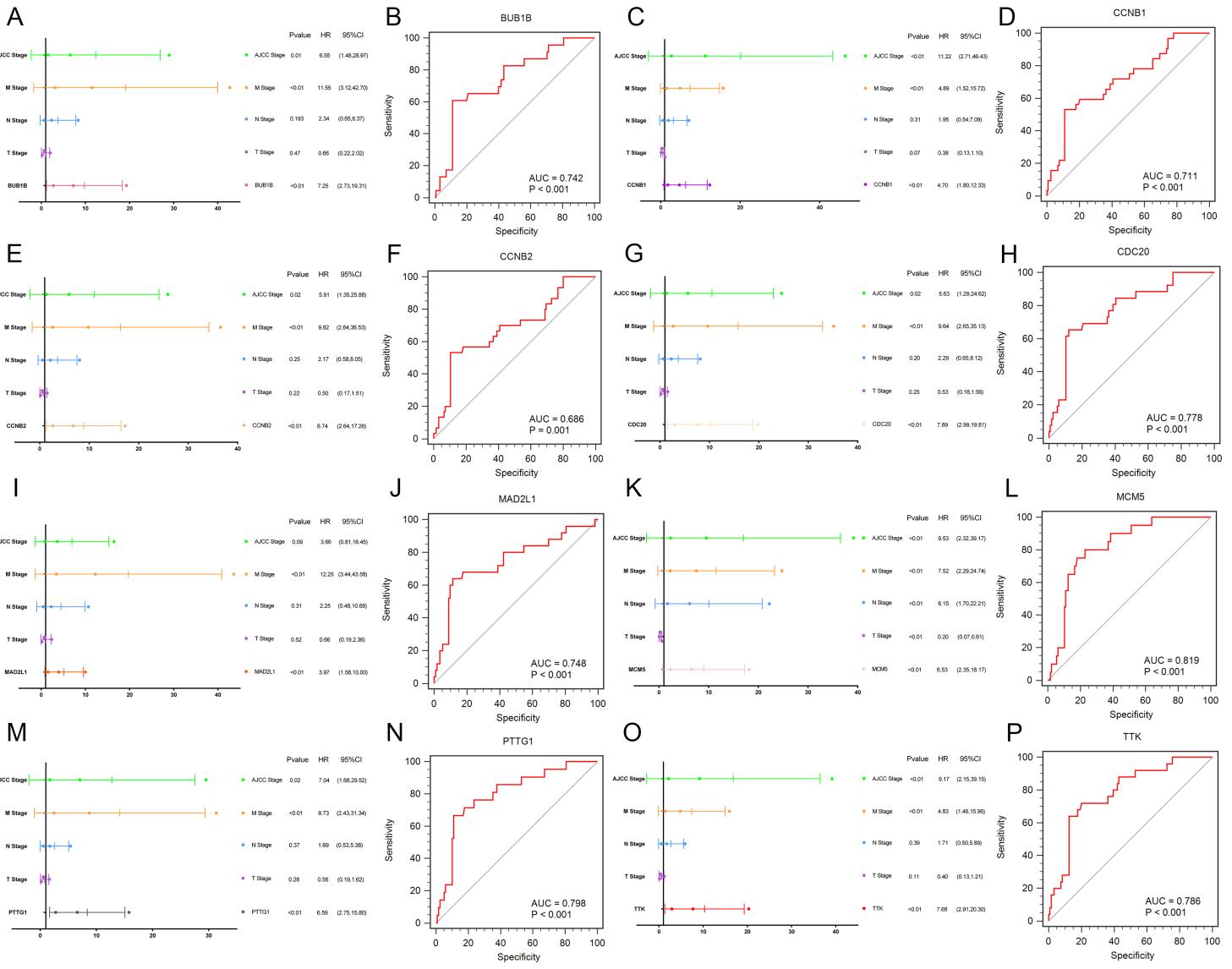


**Figure 3**

The eight hub gene expression profiles showed relatively lower expression in the matched normal tissues compared with tumor tissues.

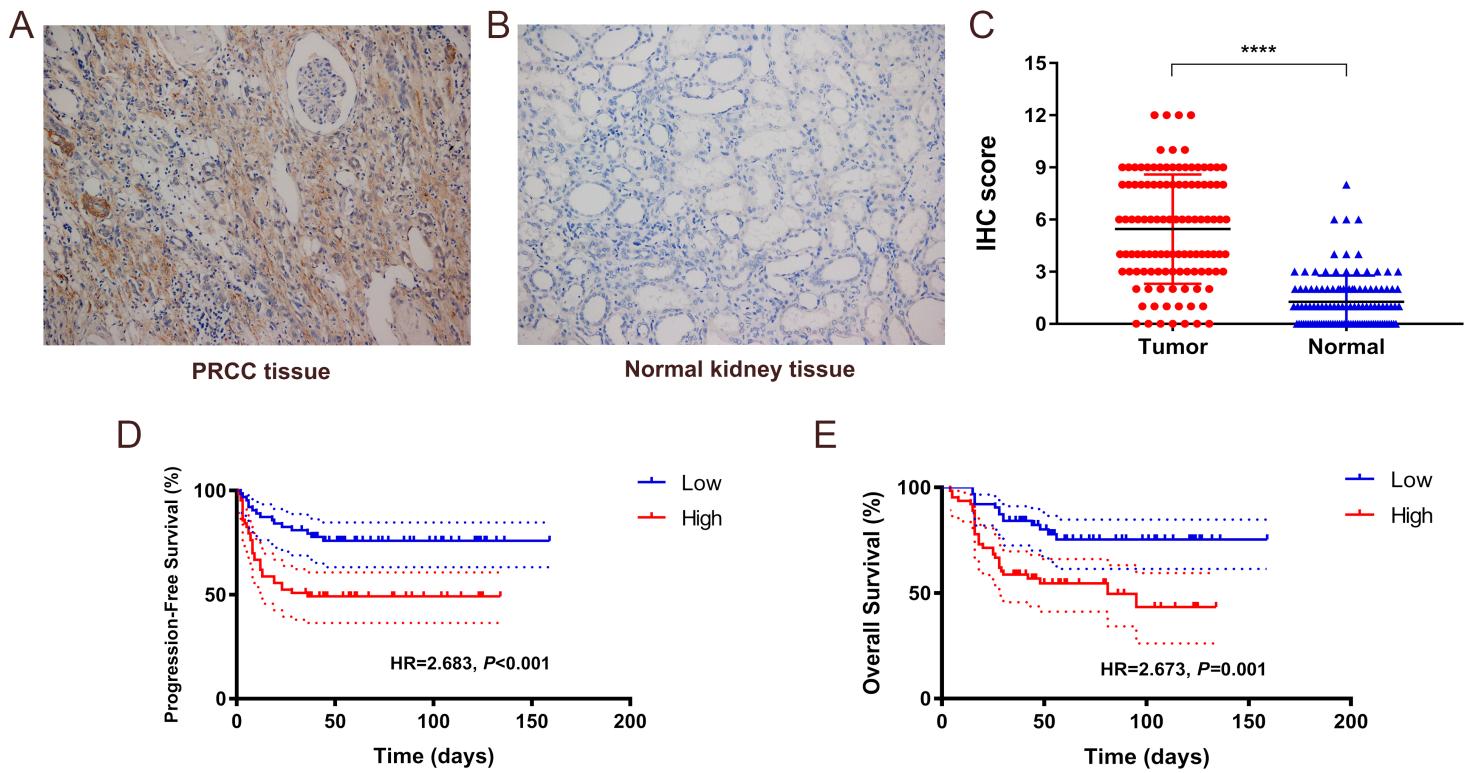
**A****Figure 4**

Kaplan-Meier curve and univariate Cox regression analyses of eight hub genes. (A) PRCC patients with elevated eight hub gene expression showed markedly significant worse OS. Apart from MCM5, PRCC patients with other seven hub gene expression showed markedly significant worse DFS. (B) Expression levels of the eight hub genes and traditional prognostic factors such as AJCC stage and pTNM stage were considered significantly related to OS ( $p<0.01$ ).



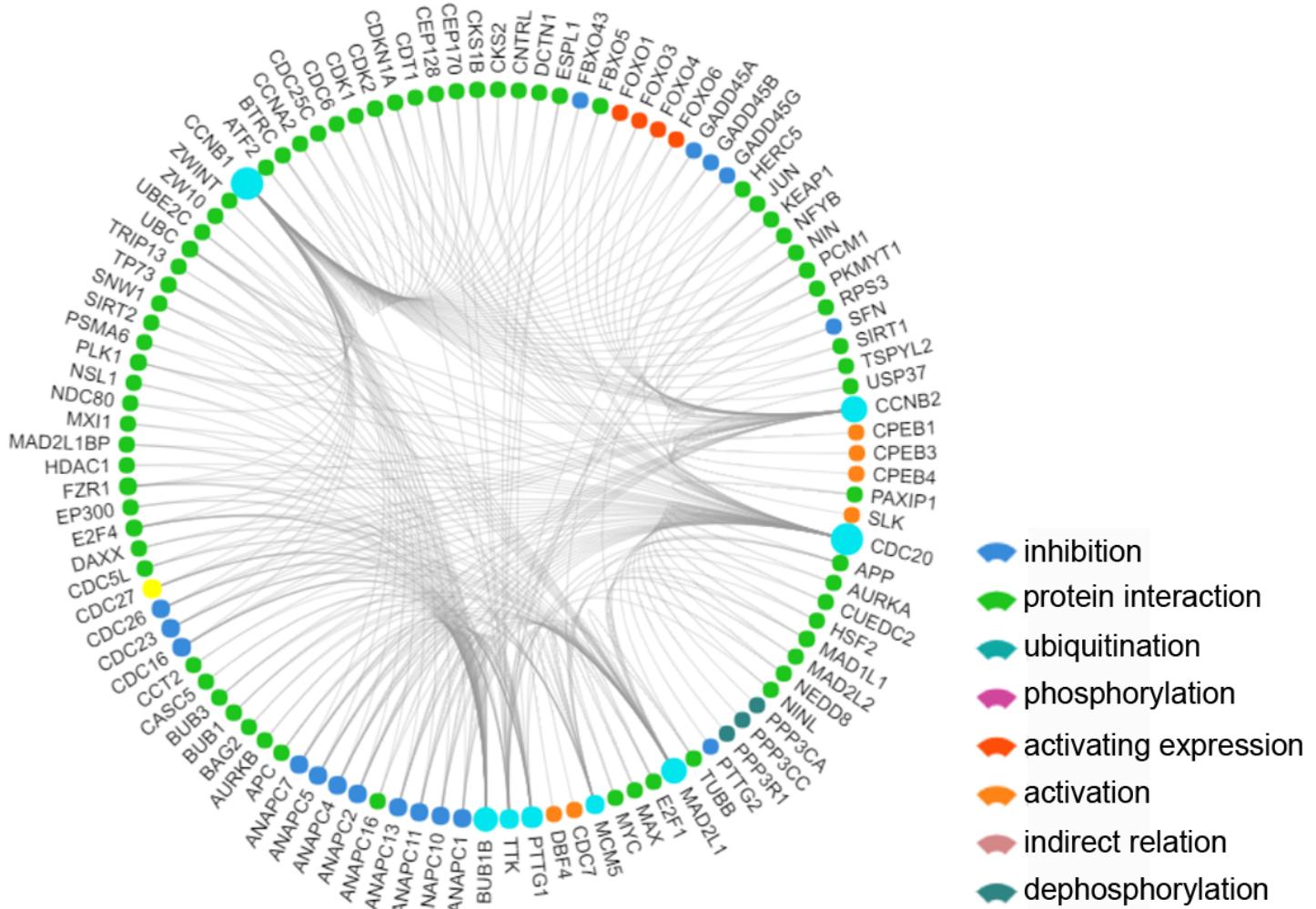
**Figure 5**

Multivariate Cox regression analyses and Receiver operating characteristic (ROC) curves of eight genes. (A, C, E, G, I, K, M, O) Traditional prognostic factors were not always relevant to OS. However, the expression levels of the eight hub genes were still significant to OS ( $p<0.01$ ). (B, D, F, H, J, L, N, P) The area under the curve (AUC) score ranked as follow: MCM5 (0.819), PTTG1 (0.798), TTK (0.786), CDC20 (0.778), MAD2L1 (0.748), BUB1B (0.742), CCNB1 (0.711), CCNB2 (0.686).



**Figure 6**

PTTG1 expression and Kaplan-Meier curve in FUSCC cohort. (A-C) IHC staining was performed and the density and intensity of PTTG1 expression in PRCC tissues in the FUSCC cohort were significantly higher than those in adjacent normal renal tissues. (D-E) High expression group of PPTG1 showed poorer OS (HR=2.673, P=0.001) and PFS (HR=2.683, P<0.001) than low expression group of PPTG1 in FUSCC cohort.



**Figure 7**

Potential networks associated with functional annotations of the hub genes include inhibition, protein interaction, ubiquitination, phosphorylation, activating expression, activation, indirect relation, dephosphorylation.

## Supplementary Files

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

- [supplementaryfigure1.tif](#)
- [SupTable1.docx](#)