A five-gene signature predicts overall survival of patients with papillary renal cell carcinoma

Ze Gao[1]�उ, Dong Zhang[2]�उ, Yi Duan[2], Lei Yan[1], Yidong Fan[1], Zhiqing Fang[1]‡*, Zhaoxu Liu[1,3,4]‡*

1 Department of Urology, Qilu Hospital of Shandong University, Jinan, Shandong, China, 2 Department of Breast Surgery, Qilu Hospital of Shandong University, Jinan, Shandong, China, 3 School of Nursing, Shandong University, Jinan, Shandong, China, 4 Department of Oncology and Pathology, Karolinska Institutet, Stockholm, Sweden

�उ These authors contributed equally to this work.
‡ These authors also contributed equally to this work.
* zhaoxvl@sdu.edu.cn (ZF); doctor_fangzq@sdu.edu.cn (ZL)

Abstract

Background

The present study aims to investigate the gene expression changes in papillary renal cell carcinoma(pRCC) and screen several genes and associated pathways of papillary renal cell carcinoma progression.

Methods

The papillary renal cell carcinoma RNA sequencing (RNA-seq) data set was downloaded from TCGA (The Cancer Genome Atlas). We identified the differentially expressed mRNAs between cancer and normal tissues and performed annotation of differentially expressed mRNAs to figure out the functions and pathways they were enriched in. Then, we constructed a risk score that relied on the 5-mRNA. The optimal value for the patients'classification risk level was identified by ROC analysis. The relationship between mRNA expression and prognosis of papillary renal cell carcinoma was evaluated by univariate Cox regression model. The 5-mRNA based risk score was validated in both complete set and testing set.

Result

In general, the 5-mRNA (CCNB2, IGF2BP3, KIF18A, PTTG1, and BUB1) were identified and validated, which can predict papillary renal cell carcinoma patient survival. This study revealed the 5-mRNA expression profile and the potential function of a single mRNA as a prognostic target for papillary renal cell carcinoma.

Conclusion

In addition, these findings may have significant implications for potential treatments options and prognosis for patients with papillary renal cell carcinoma.

1.Introduction

The incidence of kidney cancer approximately accounted for 2.4% of adult malignancies, estimated 338,000 new cases in 2012[1]. Renal cell carcinoma (RCC), the seventh most common histological type of cancer, affects men more frequently than women and comprises nearly 90% of all kidney tumors[2, 3], with clear cell (70%), papillary (10–15%), and chromophobe (5%) carcinoma the main histologic types[4]. Papillary renal cell carcinoma (pRCC) is a renal parenchyma malignant tumor with papillary or tubulopapillary architecture, including type 1 and type 2 pRCC[5]. According to the original description, type 1 is often multifocal, characterized by single layered small cell and scanty cytoplasm whereas type 2 is heterogeneous, characterized by large pseudostratified cells and eosinophilic cytoplasm[6]. The division of pRCCs into type 1 and type 2 tumors has been proved to have prognostic significance. The mutation of MET oncogene is a crucial step into the pathogenesis of hereditary pRCC forms, resulting in constitutive activation of the tyrosine kinase domain, which leads to increasing unregulated proliferation, invasion and metastases. However, they can be found only in a small proportion of sporadic cases[7] and a greater number show somatic copy number gains involving chromosome 7q[8, 9]. Although papillary renal cell carcinoma is indolent, bilateral and multifocal in some patients, other patients have a separate lesion with an aggressive clinical course[10]. The pRCC type 1 has been associated with overexpression or activation mutation of MET proto-oncogene which encodes for a hepatocyte growth factor receptor (HGFR)[11]. The pRCC type 2 is concerned with activation of the NRF2/antioxidant response element (ARE) pathway[12]. It has been reported that the 5-year overall survival of localized pRCC rates of 78–79% and cancer-specific 5-year survival rates of 86–94%[13, 14]. Unfortunately, there are currently no effective forms of therapy for patients with advanced disease. Although many patients have greatly different treatment responses and prognoses, their tumor types are identical in histology.

With the rise of gene sequencing, a large number of messenger RNAs (mRNA) have been discovered in the role of cancer. The mRNA is a large family of RNA molecules that convey genetic information from DNA to the ribosome, where they specify the amino acid sequence of the protein products of gene expression. A lot of evidences show that mRNAs are wildly involved in fundamental cellular processes, such as cell differentiation, proliferation, growth, mobility, and apoptosis, as well as carcinogenesis or cancer progression.

Recently, large-scale gene data sets regarding pRCC were downloaded from TCGA database (https://cancergenome.nih.gov/). The R Programming Language (R) was utilized for pre-processing and analysis of these data to obtain the differentially expressed genes (DEGs). In the present study, we screened the differentially expressed mRNAs between pRCC tissues and matched normal tissues, and found the association of different mRNAs expression with clinical data. We developed and validated a risk score for a prognosis based mRNA signature to demonstrate the relationship between mRNA and the prognosis of pRCC by using sample segmentation and Cox regression analysis.

2.Materials and methods

2.1 Data source and data pre-processing

The RNA sequencing data with papillary renal cell carcinoma (pRCC) was downloaded from TCGA dataset (June 13.2018). There were 321 pRCC samples in total, including 289 primary pRCC and 32 normal ones. In total, we screened 285 available samples from 321 samples, 1 sample information was repeated, and 3 samples lost survival data. The inclusion criteria were as follows:(1) a histological diagnosis of papillary renal cell carcinoma, which is a subtype of

renal cell carcinoma, (2) patients underwent radical nephrectomy or partial nephrectomy, (3) patients that didn't receive any neoadjuvant treatment, (4) patients with corresponding RNA-seq data, (5) patients with explicit clinical prognostic data. Besides, the corresponding clinical characteristics, including sex, age, pathological stage, clinical stage, definite TNM stage and histological type of our cohort were integrated as a clinical statistical chart in Table 1. Signal value for all genes were transformed to the log base 2. Standardization of the level of mRNA expression for each sample was normalized with median quantification. Since the data comes from the TCGA database, no further approval is required from the Ethics Committee.

2.2 Screening of differentially expressed mRNA

The RNA sequencing data (level 3) of 321 pRCC samples was downloaded from the TCGA data portal. The differentially expressed mRNAs between pRCC tissues and matched normal

Table 1. Clinical parameters of patients in the papillary renal cell carcinoma cohort obtained from The Cancer Genome Atlas (n = 285).

Parameter	Value	Hazard ratio	95%CI	P value
Age(years)	61(28–88)	1.004	0.997–1.031	0.792
Sex				0.149
Female	76	1.000		
Male	209	0.617	0.320–1.190	
Pathologic stage[a]				2.0E-14
StageI	170	1.000		
StageII	21	0.958	0.214–4.295	
StageIII	50	4.779	2.260–10.108	
StageIV	15	16.112	6.698–38.758	
Clinical stage[b]				
StageI	138	1.000		<2.0E-16
StageII	21	0.570	0.074–4.423	
StageIII	29	6.561	2.891–14.892	
StageIV	10	25.639	9.679–67.920	
Tumor type[c]				
Type 1	76	1.000		0.100
Type 2	84	2.769	1.013–7.571	
T stage [d]				
T1	191	1.000		2.0E-08
T2	32	3.161	1.307–7.645	
T3&T4	60	6.815	3.408–13.66	
N stage [e]				1.0E-11
N0	144	1.000		
N1&N2	28	7.715	3.864–15.400	
M stage [f]				<2.0E-16
M0	206	1.000		
M1	11	45.660	15.560–134.0	

[a]Pathologic stage was not available in 29 case

[b]clinical stage was not available in 87 case

[c]Tumor type was not available in 124 case

[d]T stage was not available in 2 case

[e]N stage was not available in 113 case

[f] M stage was not available in 68 case.

Age is reported as the median(range).

https://doi.org/10.1371/journal.pone.0211491.t001

ones were analyzed using the limma package in R[15]. The edgeR package and DESeq2 package in R were subsequently used for the calculation of DEGs. The adjust P value <0.05 and | $\log_2 FC|>1$ were set as the cut-off criteria. The genes that presented in both edgeR and DESeq2 package analysis results were selected as the final DEGs.

In our study, we mainly used the program code written in R language to analyze and deal with RNA-seq data.

2.3 Identification and selection of prognostic related mRNA

The patients were further randomly divided into the training set (n = 143) and the testing set (n = 142) with caret package in R language. In the training set, the association between mRNAs expression and the overall survival time of patients were analyzed. Univariate Cox regression analysis was performed with Survival package in R language[16]. In the survival analysis, We selected the time that patient's initial pathological diagnosis to papillary renal cell carcinoma as the start point, and the time that death or censored events occurred as the end point. Messenger RNAs with expressing significance p values <0.05 were selected as a target mRNA. To further ensure the reliability and feasibility of clinical prognosis based on mRNA, we made a Robust likelihood-based survival analysis by using Rbsurv package in R[17, 18]. All 285 samples were again randomly assigned to a training set or a testing set with the caret package in R language. We fitted a gene to the training set of samples and obtained a corresponding parameter estimate. Then we evaluated log likelihood with the parameter estimate and the testing set of samples. This procedure was repeated 10 times independently, which resulted in 10 log-likelihood for each gene. The best gene, g (1), with the largest mean log likelihood was selected. We searched the next best gene by evaluating every two-gene model and selected an optimal one with the largest mean log likelihood. Akaike information criterions (AICs) for all the candidate models were computed and an optimal predictive model was selected as with the lowest AIC value. The above procedure was repeated 1000 times, thus obtained 1000 loglike*s for each gene. The mRNA with the highest frequency (freq$>$300) was selected as the final target mRNA.

2.4 Establishment and validation risk score formula

The association between DEGs and patients'overall survival time was analysed in training set. Univariate Cox regression analysis was performed in R language by Survival package with the threshold of p value set as 0.05. In the univariate Cox regression analysis of the training set, these prognostic-related mRNAs were included in their estimated regression coefficients, and a risk scoring formula was established. In the training set, the patients were divided into high-risk score or low-risk score group based on the optimal risk score cutoff value, which represented the point at which the Youden index (sensitivity + specificity– 1) reached a maximum value in training set. Then, we applied the optimal risk score cutoff value to the validation set and the whole pRCC cohort to evaluate the classification performance of the model. By using survival ROC package in R, the receiver operating characteristic (ROC) curve was obtained. Then, choosing the optimal cutoff point with maximum sensitivity and specificity. Based on the optimal cut-off value, the survival difference between the low-risk and high-risk group was evaluated. The accuracy of the risk score formula was then further verified in the testing set and the complete set.

2.5 Functional and pathway enrichment analysis of DEGs

The ConsensusPathDB (http://cpdb.molgen.mpg.de/) database is a molecular functional interaction database, integrating information on protein interactions, genetic interactions

signaling, metabolism, gene regulation, and drug-target interactions in humans. In order to better understand the biological functions and characteristic, the present study performed Gene ontology (GO) using clusterProfiler package in R and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses in ConsensusPathDB database. All genes from human beings were used as reference. In this study, we analyzed the DEGs that were significantly up- and down-regulated as determined from pRCC data. The p value of the univariate survival analysis was less than 0.05 and considered statistically significant. To correct errors following multiple comparison analysis, the Benjamini-Hochberg step-up method was used to control the FDR.

3.Result

3.1 Differentially expressed mRNAs in pRCC

With a cut-off value of $P < 0.05$ and $|log2FC| > 1.0$, 4776 and 4831 differentially expressed mRNAs were identified by using the DESeq2 and edgeR package in R language, respectively (S1 File). The genes that presented in both edgeR and DESeq2 package analysis results were selected as the final DEGs. According to this standard, 2632 up-regulated mRNAs and 1921 down-regulaed ones were identified (Fig 1A and 1B). We displayed the distribution of all the differentially expressed mRNAs in both the–log(FDR) and logFC dimensions through a volcano plot in Fig 2.

3.2 Identification and selection of prognostic related mRNA

The total of 4553 mRNAs was further randomized into a training set and a testing set. A total of 749 differential expressed mRNAs were identified by univariable cox regression with the $P < 0.05$ in training set (S2 File). The top of 20 genes with the lowest p value were selected and presented in Table 2. Random data analysis was performed by using Robust likelihood-based modelling for 1000 times. Statistic frequency analysis of the significantly altered mRNAs indicated that all the selected mRNAs had a high frequency (Fig 3). In further analysis, differentially expressed mRNA with a frequency above 300 was picked out. Finally, there were 5 mRNAs identified as prognostic feature (Table 3). Among these genes, four mRNAs (CCNB2,
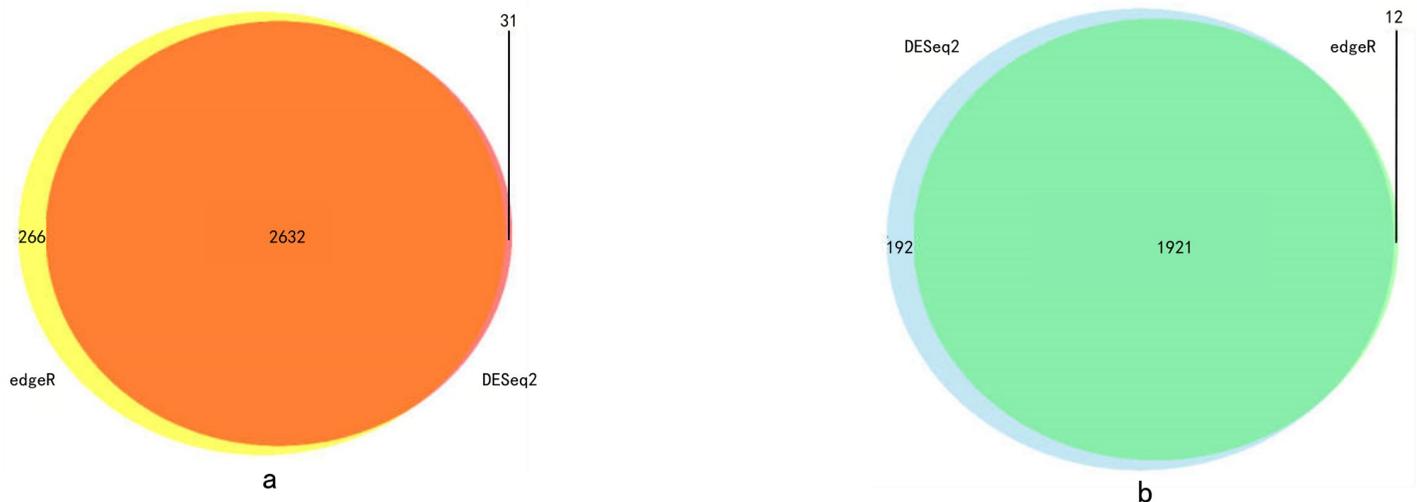
Fig 1. (a)The Venn polt of up-regulated DEG. (b)The Venn polt of down-regulated DEG.
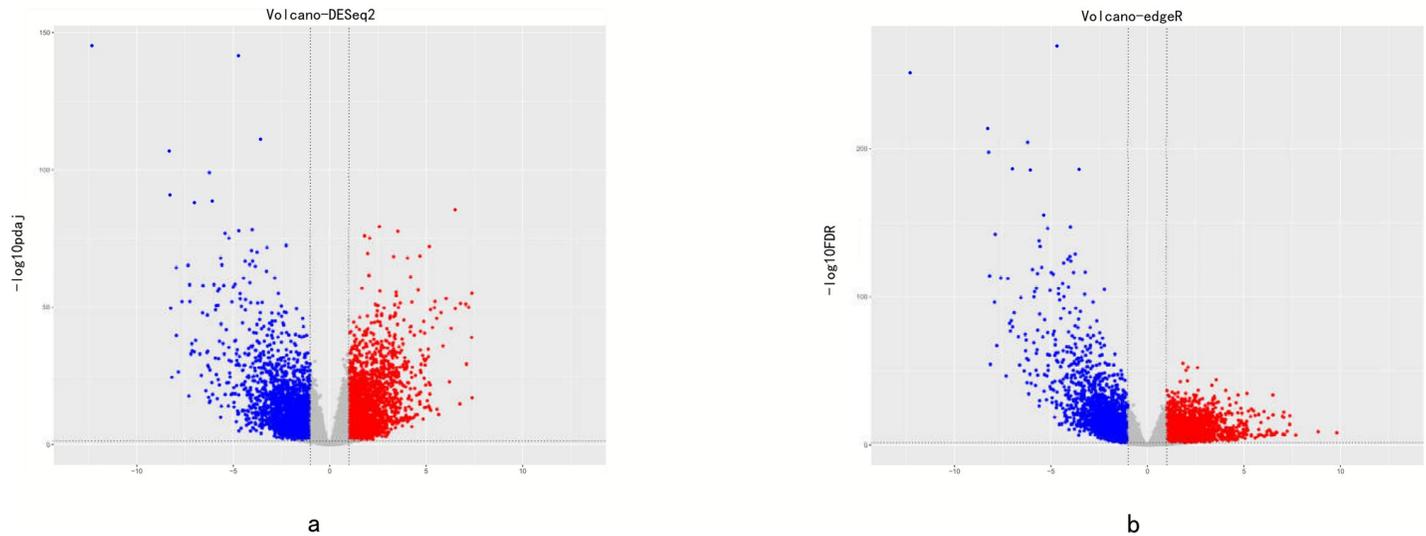
Fig 2. Volcano map with edgeR and DESeq2.

IGF2BP3, KIF18A, PTTG1) had positive coefficients which suggested that higher expression level was associated with worse survival time and one (BUB1) had negative coefficients suggested that higher levels of expression were related with better survival time.

3.3 The 5-mRNA signature predicts the survival of patients with pRCC

A risk score formula based on the expression level and coefficient of 5 mRNAs was created as follows: Risk score = (-0.560 * expression level of BUB1) + (0.337 * expression level of CCNB2) + (0.424 * expression level of IGF2BP3) + (0.516 * expression level of KIF18A) + (0.310 × expression level of PTTG1). Then, the mRNAs signature based risk score for each patient in training set were calculated, and patients in the cohort were assigned into high-risk group (n = 51) and low risk group (n = 92). The risk score was inversely related to the patients'survival time. The patients with lower risk score had longer survival time and death patients had higher risk score (Fig 4). The patients in high risk group had a tendency to have higher expression levels of IGF2BP3, KIF18A, PTTG1, BUB1 and CCNB2.

3.4 Available prognostic indicator in pRCC patients

We performed ROC analysis for the 5-mRNA risk score in the training set to further validate the sensitivity and specificity of survival prediction. The optimal cut-off value for false negative and false positive minimum was 0.951, and the area under the curve (AUC) was 0.82 (Fig 5A). The Kaplan-Meier curves showed that patients in the high-risk group had significantly shorter overall survival than those in the low-risk group (log-rank test $p < 0.0001$)(Fig 5B).

3.5 External validation of five-gene signature

The predictive efficiency of the five-mRNA signature in testing set with 142 patients was then evaluated. The patients in the testing set were classified into high-risk (n = 53) and low-risk groups (n = 89) by using the same model and criteria. Similar to the training set, overall survival was significantly lower in the high-risk group than in the low-risk group ($p < 0.0001$)(Fig 6A). Risk score-based classification of the external complete set also yielded similar results as shown in Fig 6B. Despite unbalanced samples in each group, the results of survival analysis

Table 2. Top 20 mRNAs significantly associated with the survival time of patients in the training set(n = 143).

mRNA	HR	Cox *p* value	CI 95
IGF2BP3	1.72	0	1.42–2.09
CENPA	2.18	1.00E-06	1.6–2.97
CCNB2	2.2	1.00E-06	1.6–3.02
KIF18A	2.69	2.00E-06	1.78–4.06
CLDN11	1.66	7.00E-06	1.33–2.08
LOC101928583	1.66	7.00E-06	1.33–2.08
TTK	1.97	7.00E-06	1.47–2.65
ZNF488	1.67	1.10E-05	1.33–2.09
KIF20A	1.92	1.20E-05	1.43–2.57
DEPDC1	1.65	1.20E-05	1.32–2.06
BUB1	1.93	1.20E-05	1.44–2.6
PTTG1	2.26	1.30E-05	1.57–3.27
ASPM	1.65	1.50E-05	1.32–2.07
NEK2	1.73	1.50E-05	1.35–2.21
CEP55	1.8	1.70E-05	1.38–2.35
TOP2A	1.7	2.00E-05	1.33–2.16
TPX2	1.94	2.20E-05	1.43–2.63
NUF2	1.8	2.50E-05	1.37–2.36
SOX11	1.56	2.60E-05	1.27–1.91
DLGAP5	1.76	2.80E-05	1.35–2.29

https://doi.org/10.1371/journal.pone.0211491.t002

were similar to those of low-risk patients with significantly longer survival times than those with high-risk patients ($p < 0.0001$). The clinical characteristics available from the TCGA database were integrated and univariate Cox regression analysis was performed to detect the candidate clinical prognostic parameters. To validate independent predictive power of five-gene signature from clinicopathological factors in pRCC cohort such as age, pathological stage, clinical stage, pathological type and so on, the stratified Cox proportional hazard analysis was constructed and visualized in Fig 7 and Fig 8, which suggested this signature could distinguish the high-risk subgroup from low-risk one in each clinical subtype and further convinced the independent prognostic efficacy of the final gene signature.

3.6 Functional and pathway enrichment analysis of DEGs

After performing GO analysis of DEGs with clusterProfiler package in R language, the DEGs were classified into three groups: molecular function group, biological process group and cellular component group. The biological results revealed that DEGs were primarily enriched in chromosome segregation, mitotic nuclear division, nuclear chromosome segregation, urogenital system development, renal system development and kidney development. The cellular component results indicated that DEGs were mainly riched in chromosomal region, proteinaceous extracellular matrix, condensed chromosome, spindle, chromosome and centromeric region. The molecular function results showed that DEGs were mainly enriched in ion channel activity and substrate-specific channel activity (Fig 9A–9C). To investigate pathway enrichment, KEGG signaling pathway analysis was used to identify the pathways, including cAMP signaling pathway, phospholipase D signaling pathway, Hippo signaling pathway, cell cycle, complement and coagulation cascades, TGF-beta signaling pathway, inflammatory mediator regulation of TRP channels and melanoma (Fig 9D).
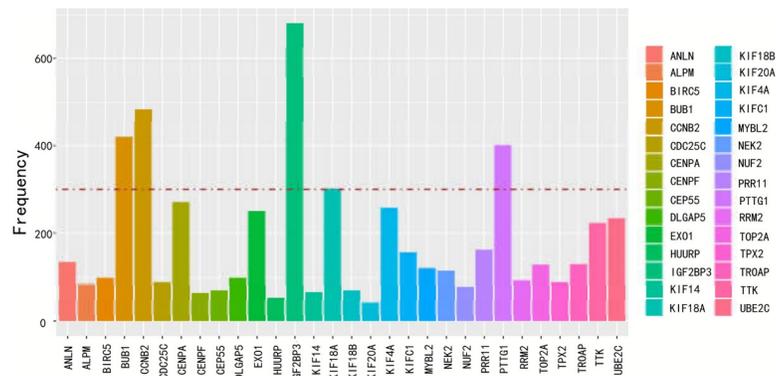
Fig 3. Statistic frequency analysis of the significantly altered mRNAs.

Discussion

In the present study, the R language was used to analysis the gene data downloaded from TCGA databases. The total of 4553 DEGs in pRCC compared with control samples were identified, which included 2632 up-regulated and 1921 down-regulated genes. The DEGs were mainly enriched in 60 GO terms, including ion channel activity, chromosome segregation and chromosomal region. The KEGG pathway enrichment analysis result showed that the DEGs were associated with calcium,cAMP,phosphollipase D, and hippo signaling pathway. The Ca^{2+}-mediated signaling pathways have been implicated either directly or indirectly related to tumorigenesis and tumor progression[19–21]. In this pathway, Orai1 exaggerates cell proliferation, migration, invasion and evasion of apoptosis[19]. The cAMP plays a complex and context-dependent role in regulating cell migration[22]. Usually the focus has been mainly on PKA-mediated migratory effects. Shaikh et al. suggests that hypoxia enhances cAMP-dependent protein kinase activity by up-regulating PKA gene expression in a HIF dependent mechanism and that PKA plays a key role in hypoxia-mediated EMT, migration, and invasion in lung cancer cells[23]. There is increasing evidence that the Hippo pathway is dysregulated in many human cancers, and dysregulation of the Hippo pathway exerts a significant impact on cancer development, including liver, breast, lung, colon, ovary, and others[24]. Aberrant phospholipase D (PLD) expression has been identified in multiple facets of complex pathological states, including cancer and inflammatory diseases. PLD contributes to various mitogenic or oncogenic signaling pathways[25]. Therefore, monitoring of these signaling pathway maybe beneficial to understanding the mechanism of carcinogenesis and researching treatment of prostate cancer.

We identified 5 mRNAs that are associated with the survival of patients with papillary renal cell carcinoma,namely CCNB2, IGF2BP3, KIF18A, PTTG1, and BUB1 in the training set. The prognostic related mRNAs were further selected to construct a risk score formula by Cox

Table 3. Significantly altered mRNAs with frequency above 300.

mRNA	Count
BUB1	421
CCNB2	484
IGF2BP3	681
KIF18A	302
PTTG1	401
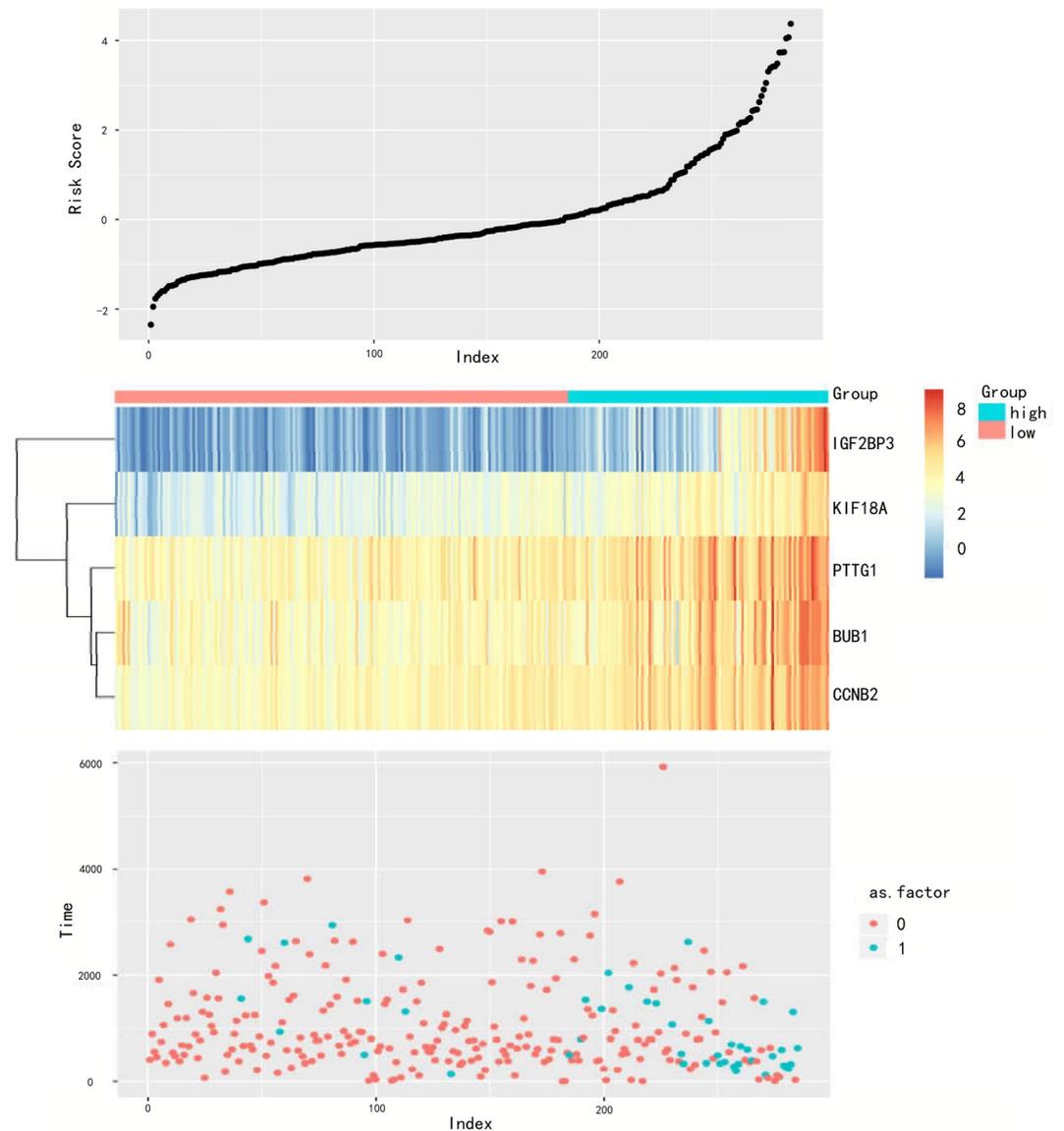
Fig 4. mRNA risk score analysis of the training set. The mRNA signature risk score distribution heat-map of the mRNA expression profiles.Rows represent mRNAs, and columns represent patients.

regression model. Besides, we used the ROC analysis to identify the optimal cutoff point, and divided patient into high-risk and low-risk groups. In the univariate Cox regression model, the survival time of high-risk patients was significantly shortened. Subsequently, the GO and KEGG pathway analysis suggest that mRNA plays a crucial role in molecular pathogenesis and progression of pRCC.

Among the 5 mRNAs which are associated with the prognosis of pRCC, some have been reported to express in cancer or other diseases, but have not been examined in pRCC. For example, elevated cytoplasmic CCNB2 protein levels are strongly associated with short-term disease-specific survival of breast cancer patients[26]. Cyclin B2 (CCNB2), the member of cyclin family proteins, regulates the activities of cyclin dependent kinases (CDKs) and different cyclins function spatially and temporally in specific phases of the cell cycle[27]. CCNB2
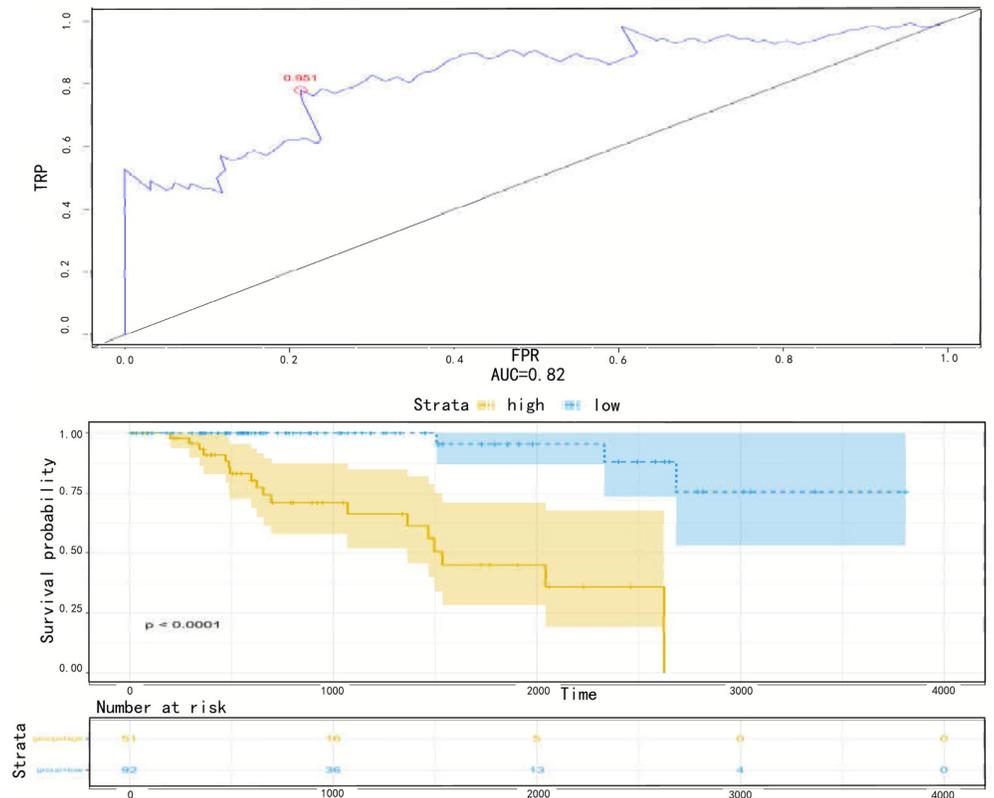
Fig 5. (a) ROC for 5 mRNA,Method = KM. Receiver operating characteristic (ROC) analyzes the sensitivity and specificity of the survival time by risk score based on the 5-mRNA signature. The red dot indicates the optimal cut-off point. (b) Method Kaplan Meier. Kaplan-Meier estimated the survival time of the training set patients using risk score based on the 5 mRNA signature. The plot was used to visualize the survival probability for the low-risk versus high-risk group of patients based on the optimal cut-off point.

https://doi.org/10.1371/journal.pone.0211491.g005
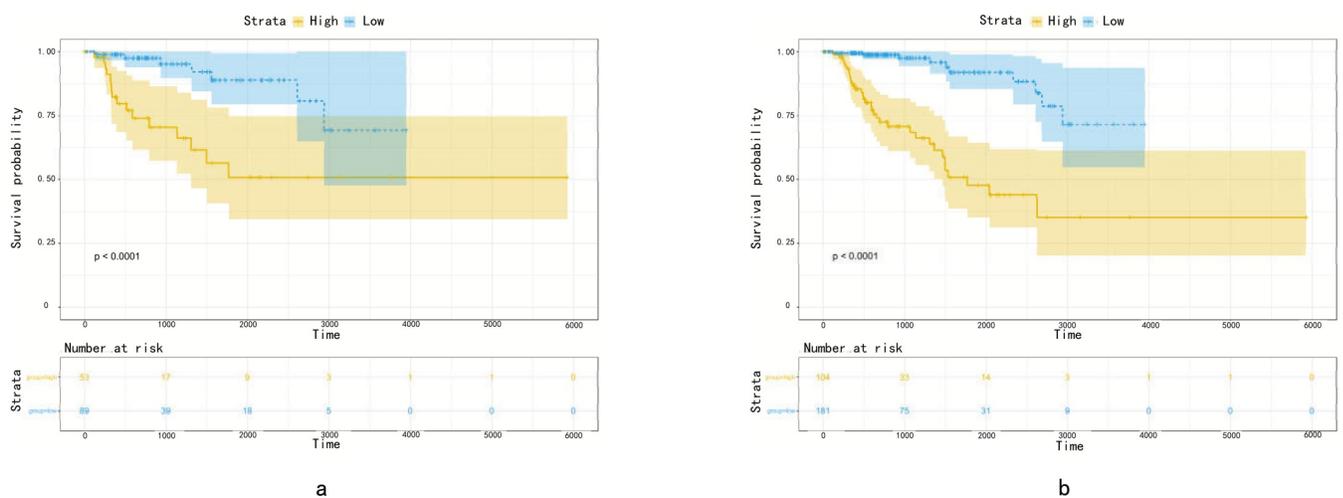
Fig 6. (a)Method Kaplan Meier. Kaplan-Meier estimated the survival time of the testing set patients using a risk score based on 5 mRNA signature. (b) Method Kaplan Meier. Kaplan-Meier estimated the survival time of the complete set patients from using a risk score based on 5 mRNA signature. The plot was used to visualize the survival probability for the low-risk versus high-risk group of patients based on the optimal cut-off point.

https://doi.org/10.1371/journal.pone.0211491.g006
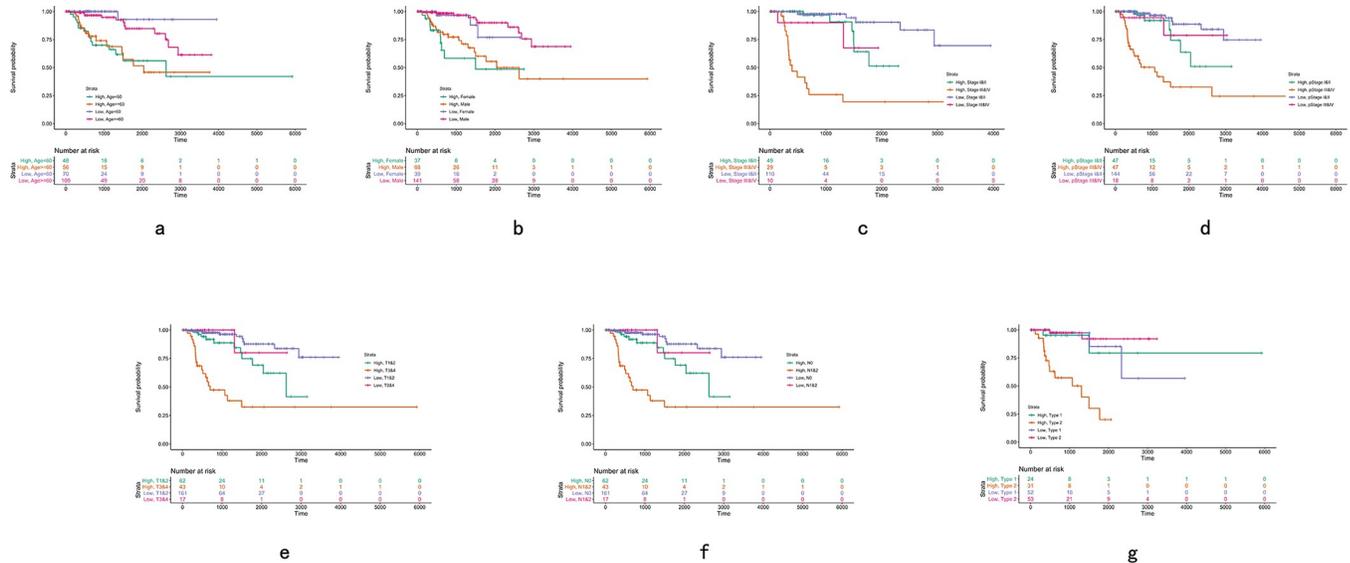
Fig 7. Stratified Cox hazard analysis of five gene signature on clinicopathological characteristics in pRCC cohort.

https://doi.org/10.1371/journal.pone.0211491.g007

commonly triggers the G2/M transfer by activating CDK1 kinase[28, 29]. CCNB2 is generally located in the Golgi apparatus during both interphase and mitosis[30]. Many evidences
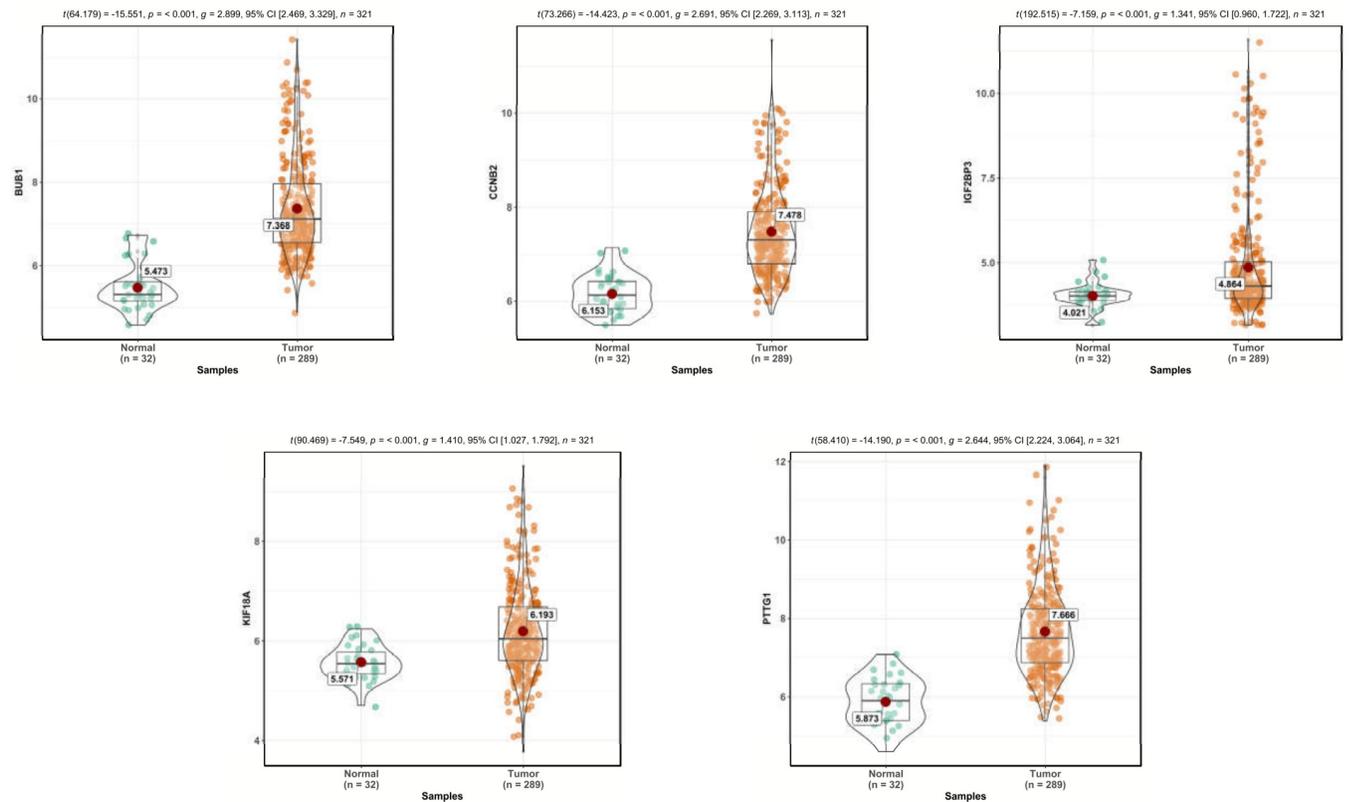
Fig 8. The comparative gene expression level of CCNB2, IGF2BP3, KIF18A, PTTG1, and BUB1 in normal tissue and pRCC tissue from TCGA database, respectively.

https://doi.org/10.1371/journal.pone.0211491.g008
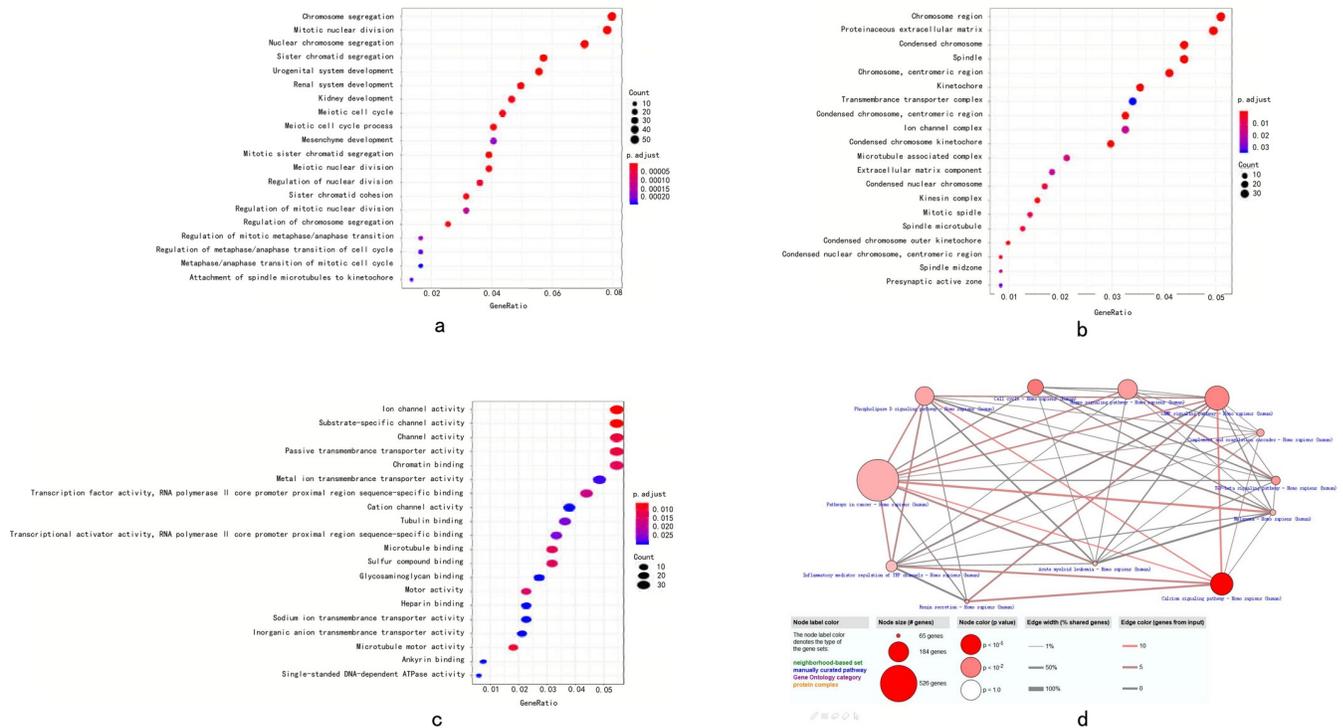
Fig 9. (a) biological process group of GO analysis. (b) cellular component group of GO analysis. (c) molecular function group of GO analysis.(d) The KEGG analysis. Biological function and KEGG pathway analysis of target genes. The overlapping target genes were predicted using the R language. (a-c) The enriched GO biological processes of target genes. (d) The enriched KEGG pathways of target genes.

suggested that CCNB2 expression was increased in a variety of human cancers, such as non-small cell lung cancer[31], breast carcinoma[26], gastric cancer[32], colorectal adenocarcinoma[33], pituitary adenoma[34] and adrenocortical carcinoma[35]. Therefore, CCNB2 is important in pRCC and may be used as a prognostic indicator. It has been reported that insulin-like growth factor 2 mRNA binding protein 3 (IGF2BP3) expression correlates with malignancy[36]. Schaeffer et al. suggested that IGF2BP3 was found to be selectively overexpressed in pancreatic ductal adenocarcinoma tissues but not in benign pancreatic tissues[37]. In ovarian cancer, high expression of IGF2BP3 was associated with poor survival, and women diagnosed at advanced stages with elevated IGF2BP3 was at higher risk of developing chemoresistance [38]. Lochhead et al. revealed that normal colorectal epithelium was negative for IGF2BP3 in patients of normal mucosa adjacent to carcinoma, and IGF2BP3 was associated with poor differentiation, stage III–IV disease, BRAF mutation, and LINE-1 hypomethylation[39]. Lin et al. demonstrated that IGF2BP3 enhances cell invasion ability and tumorigenicity in human OSCC in vitro and in vivo[40]. In the present study, it was demonstrated that the pRCC patients with IGF2BP3 alterations exhibited a poorer survival rate compared with those without the genetic alterations. This result suggested that the mutation in IGF2BP3 reduces the survival rate of patients with pRCC. Kif18A, a member of the kinesin super family of molecular motor proteins, regulates chromosome congregation and suppresses kinetochore movements to control mitotic chromosome alignment in the pre-anaphase state of the mammalian cell cycle[41]. Notably, Nagahara et al. demonstrated that colorectal cancer cells transfected with Kif18A cDNA demonstrated significant enhanced migration and invasion compared to mock-transfected cells[42]. KIF18A might be a biomarker for HCC diagnosis and an independent predictor of DFS and OS after surgical resection[43]. The results of the present study were

consistent with the results of previous studies. KIF18A may have an important role in progression of pRCC, however this requires further study, in order to verify the specific molecular marker role of KIF18A in the diagnosis of patients with pRCC. Pituitary tumour transforming gene 1 (PTTG1) is over-expressed in a vast array of malignancies including pituitary [44, 45], thyroid[46], colorectal[47] and lung[48] cancer. The high level of PTTG1 is commonly associated with an enhanced proliferative capacity, increased tumour grade and high invasive potential. Current evidences regarding the biological role of BUB1 in cancer is contradictory. Mutations in BUB1, some of which are functional, occur in some cancers, including those that originate in the lung, colon, and are reported to be associated with chromosomal instability and lymph node metastasis, suggesting that silencing of this kinase may mediate aggressive clinical behavior[49]. Moreover, Aurora b hyperactivation caused by overexpression of BUB1 leads to misaggregation of chromosomes, thereby inducing aneuploidy[50].

Conclusion

Taken all together, by performing a comprehensive analysis for differentially expressed mRNA profiles and corresponding clinical information, our study demonstrated that five-mRNA signature was a potential diagnostic marker in pRCC, and was an independent prognostic factor in pRCC patients. This signature has a lot of potential prognostic and therapeutic implications for the pRCC patient management. However, further research is needed to validate our findings and establish molecular mechanisms for mRNAs interactions and papillary renal cell carcinoma progression.

Supporting information

S1 File. Details of the result of univariable cox regression analysis.
(XLSX)

S2 File. The mRNA with p-value less than 0.05 in univariable cox regression analysis.
(XLS)

Acknowledgments

Author Contributions

Conceptualization: Zhaoxu Liu.

Data curation: Zhiqing Fang.

Investigation: Yi Duan.

Methodology: Yidong Fan.

Resources: Lei Yan.

Software: Dong Zhang.

Writing – original draft: Ze Gao.

References

1. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. International journal of cancer. 2015; 136(5):E359–86. https://doi.org/10.1002/ijc.29210 PMID: 25220842.

2. Badowska-Kozakiewicz AM, Budzik MP, Koczkodaj P, Przybylski J. Selected tumor markers in the routine diagnosis of chromophobe renal cell carcinoma. Archives of medical science: AMS. 2016; 12 (4):856–63. https://doi.org/10.5114/aoms.2015.51188 PMID: 27478468; PubMed Central PMCID: PMC4947610.

3. Jemal A, Siegel R, Ward E, Hao Y, Xu J, Murray T, et al. Cancer statistics, 2008. CA: a cancer journal for clinicians. 2008; 58(2):71–96. https://doi.org/10.3322/CA.2007.0010 PMID: 18287387.

4. Znaor A, Lortet-Tieulent J, Laversanne M, Jemal A, Bray F. International variations and trends in renal cell carcinoma incidence and mortality. European urology. 2015; 67(3):519–30. https://doi.org/10.1016/j.eururo.2014.10.002 PMID: 25449206.

5. Montironi R, Cheng L, Scarpelli M, Lopez-Beltran A. Pathology and Genetics: Tumours of the Urinary System and Male Genital System: Clinical Implications of the 4th Edition of the WHO Classification and Beyond. European urology. 2016; 70(1):120–3. https://doi.org/10.1016/j.eururo.2016.03.011 PMID: 26996660.

6. Delahunt B, Eble JN. Papillary renal cell carcinoma: a clinicopathologic and immunohistochemical study of 105 tumors. Modern pathology: an official journal of the United States and Canadian Academy of Pathology, Inc. 1997; 10(6):537–44. PMID: 9195569.

7. Courthod G, Tucci M, Di Maio M, Scagliotti GV. Papillary renal cell carcinoma: A review of the current therapeutic landscape. Critical reviews in oncology/hematology. 2015; 96(1):100–12. https://doi.org/10.1016/j.critrevonc.2015.05.008 PMID: 26052049.

8. Ridge CA, Pua BB, Madoff DC. Epidemiology and staging of renal cell carcinoma. Seminars in interventional radiology. 2014; 31(1):3–8. https://doi.org/10.1055/s-0033-1363837 PMID: 24596434; PubMed Central PMCID: PMC3930658.

9. Schmidt L, Duh FM, Chen F, Kishida T, Glenn G, Choyke P, et al. Germline and somatic mutations in the tyrosine kinase domain of the MET proto-oncogene in papillary renal carcinomas. Nature genetics. 1997; 16(1):68–73. https://doi.org/10.1038/ng0597-68 PMID: 9140397.

10. Lee BH. Commentary on: "Comprehensive molecular characterization of papillary renal-cell carcinoma." Cancer Genome Atlas Research Network.: N Engl J Med. 2016 Jan 14; 374(2):135–45. Urologic oncology. 2017;35(9):578–9. https://doi.org/10.1056/NEJMoa1505917 PMID: 26536169.

11. Albiges L, Guegan J, Le Formal A, Verkarre V, Rioux-Leclercq N, Sibony M, et al. MET is a potential target across all papillary renal cell carcinomas: result from a large molecular study of pRCC with CGH array and matching gene expression array. Clinical cancer research: an official journal of the American Association for Cancer Research. 2014; 20(13):3411–21. https://doi.org/10.1158/1078-0432.CCR-13-2173 PMID: 24658158.

12. Ooi A, Wong JC, Petillo D, Roossien D, Perrier-Trudova V, Whitten D, et al. An antioxidant response phenotype shared between hereditary and sporadic type 2 papillary renal cell carcinoma. Cancer cell. 2011; 20(4):511–23. https://doi.org/10.1016/j.ccr.2011.08.024 PMID: 22014576.

13. Steffens S, Janssen M, Roos FC, Becker F, Schumacher S, Seidel C, et al. Incidence and long-term prognosis of papillary compared to clear cell renal cell carcinoma—a multicentre study. European journal of cancer. 2012; 48(15):2347–52. https://doi.org/10.1016/j.ejca.2012.05.002 PMID: 22698386.

14. Patard JJ, Leray E, Rioux-Leclercq N, Cindolo L, Ficarra V, Zisman A, et al. Prognostic value of histologic subtypes in renal cell carcinoma: a multicenter experience. Journal of clinical oncology: official journal of the American Society of Clinical Oncology. 2005; 23(12):2763–71. https://doi.org/10.1200/JCO.2005.07.055 PMID: 15837991.

15. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic acids research. 2015; 43(7):e47. https://doi.org/10.1093/nar/gkv007 PMID: 25605792; PubMed Central PMCID: PMC4402510.

16. O'Quigley J, Moreau T. Cox's regression model: computing a goodness of fit statistic. Computer methods and programs in biomedicine. 1986; 22(3):253–6. PMID: 3524984.

17. Wang JY, Tai JJ. Robust quantitative trait association tests in the parent-offspring triad design: conditional likelihood-based approaches. Annals of human genetics. 2009; 73(2):231–44. https://doi.org/10.1111/j.1469-1809.2008.00502.x PMID: 19183344.

18. Renaud G, Stenzel U, Maricic T, Wiebe V, Kelso J. deML: robust demultiplexing of Illumina sequences using a likelihood-based approach. Bioinformatics. 2015; 31(5):770–2. https://doi.org/10.1093/bioinformatics/btu719 PMID: 25359895; PubMed Central PMCID: PMC4341068.

19. Monteith GR, McAndrew D, Faddy HM, Roberts-Thomson SJ. Calcium and cancer: targeting Ca2+ transport. Nature reviews Cancer. 2007; 7(7):519–30. https://doi.org/10.1038/nrc2171 PMID: 17585332.

20. Prevarskaya N, Skryma R, Shuba Y. Calcium in tumour metastasis: new roles for known actors. Nature reviews Cancer. 2011; 11(8):609–18. https://doi.org/10.1038/nrc3105 PMID: 21779011.

21. Motiani RK, Hyzinski-Garcia MC, Zhang X, Henkel MM, Abdullaev IF, Kuo YH, et al. STIM1 and Orai1 mediate CRAC channel activity and are essential for human glioblastoma invasion. Pflugers Archiv: European journal of physiology. 2013; 465(9):1249–60. https://doi.org/10.1007/s00424-013-1254-8 PMID: 23515871; PubMed Central PMCID: PMC3748246.

22. Howe AK. Cross-talk between calcium and protein kinase A in the regulation of cell migration. Current opinion in cell biology. 2011; 23(5):554–61. https://doi.org/10.1016/j.ceb.2011.05.006 PMID: 21665456; PubMed Central PMCID: PMC3175278.

23. Shaikh D, Zhou Q, Chen T, Ibe JC, Raj JU, Zhou G. cAMP-dependent protein kinase is essential for hypoxia-mediated epithelial-mesenchymal transition, migration, and invasion in lung cancer cells. Cellular signalling. 2012; 24(12):2396–406. https://doi.org/10.1016/j.cellsig.2012.08.007 PMID: 22954688.

24. Yu FX, Zhao B, Guan KL. Hippo Pathway in Organ Size Control, Tissue Homeostasis, and Cancer. Cell. 2015; 163(4):811–28. https://doi.org/10.1016/j.cell.2015.10.044 PMID: 26544935; PubMed Central PMCID: PMC4638384.

25. Kang DW, Choi KY, Min do S. Functional regulation of phospholipase D expression in cancer and inflammation. The Journal of biological chemistry. 2014; 289(33):22575–82. https://doi.org/10.1074/jbc.R114.569822 PMID: 24990948; PubMed Central PMCID: PMC4132765.

26. Shubbar E, Kovacs A, Hajizadeh S, Parris TZ, Nemes S, Gunnarsdottir K, et al. Elevated cyclin B2 expression in invasive breast carcinoma is associated with unfavorable clinical outcome. BMC cancer. 2013; 13:1. https://doi.org/10.1186/1471-2407-13-1 PMID: 23282137; PubMed Central PMCID: PMC3545739.

27. Vermeulen K, Van Bockstaele DR, Berneman ZN. The cell cycle: a review of regulation, deregulation and therapeutic targets in cancer. Cell proliferation. 2003; 36(3):131–49. PMID: 12814430.

28. Petri ET, Errico A, Escobedo L, Hunt T, Basavappa R. The crystal structure of human cyclin B. Cell cycle. 2007; 6(11):1342–9. https://doi.org/10.4161/cc.6.11.4297 PMID: 17495533.

29. Wu T, Zhang X, Huang X, Yang Y, Hua X. Regulation of cyclin B2 expression and cell cycle G2/m transition by menin. The Journal of biological chemistry. 2010; 285(24):18291–300. https://doi.org/10.1074/jbc.M110.106575 PMID: 20404349; PubMed Central PMCID: PMC2881754.

30. Jackman M, Firth M, Pines J. Human cyclins B1 and B2 are localized to strikingly different structures: B1 to microtubules, B2 primarily to the Golgi apparatus. The EMBO journal. 1995; 14(8):1646–54. PMID: 7737117; PubMed Central PMCID: PMC398257.

31. Qian X, Song X, He Y, Yang Z, Sun T, Wang J, et al. CCNB2 overexpression is a poor prognostic biomarker in Chinese NSCLC patients. Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie. 2015; 74:222–7. https://doi.org/10.1016/j.biopha.2015.08.004 PMID: 26349989.

32. Wang DG, Chen G, Wen XY, Wang D, Cheng ZH, Sun SQ. Identification of biomarkers for diagnosis of gastric cancer by bioinformatics. Asian Pacific journal of cancer prevention: APJCP. 2015; 16(4):1361–5. PMID: 25743799.

33. Park SH, Yu GR, Kim WH, Moon WS, Kim JH, Kim DG. NF-Y-dependent cyclin B2 expression in colorectal adenocarcinoma. Clinical cancer research: an official journal of the American Association for Cancer Research. 2007; 13(3):858–67. https://doi.org/10.1158/1078-0432.CCR-06-1461 PMID: 17289878.

34. De Martino I, Visone R, Wierinckx A, Palmieri D, Ferraro A, Cappabianca P, et al. HMGA proteins up-regulate CCNB2 gene in mouse and human pituitary adenomas. Cancer research. 2009; 69(5):1844–50. https://doi.org/10.1158/0008-5472.CAN-08-4133 PMID: 19223528.

35. Fernandez-Ranvier GG, Weng J, Yeh RF, Khanafshar E, Suh I, Barker C, et al. Identification of biomarkers of adrenocortical carcinoma using genomewide gene expression profiling. Archives of surgery. 2008; 143(9):841–6; discussion 6. https://doi.org/10.1001/archsurg.143.9.841 PMID: 18794420.

36. Ennajdaoui H, Howard JM, Sterne-Weiler T, Jahanbani F, Coyne DJ, Uren PJ, et al. IGF2BP3 Modulates the Interaction of Invasion-Associated Transcripts with RISC. Cell reports. 2016; 15(9):1876–83. https://doi.org/10.1016/j.celrep.2016.04.083 PMID: 27210763; PubMed Central PMCID: PMC4889463.

37. Schaeffer DF, Owen DR, Lim HJ, Buczkowski AK, Chung SW, Scudamore CH, et al. Insulin-like growth factor 2 mRNA binding protein 3 (IGF2BP3) overexpression in pancreatic ductal adenocarcinoma correlates with poor survival. BMC cancer. 2010; 10:59. https://doi.org/10.1186/1471-2407-10-59 PMID: 20178612; PubMed Central PMCID: PMC2837867.

38. Hsu KF, Shen MR, Huang YF, Cheng YM, Lin SH, Chow NH, et al. Overexpression of the RNA-binding proteins Lin28B and IGF2BP3 (IMP3) is associated with chemoresistance and poor disease outcome in

ovarian cancer. British journal of cancer. 2015; 113(3):414–24. https://doi.org/10.1038/bjc.2015.254 PMID: 26158423; PubMed Central PMCID: PMC4522643.

39. Lochhead P, Imamura Y, Morikawa T, Kuchiba A, Yamauchi M, Liao X, et al. Insulin-like growth factor 2 messenger RNA binding protein 3 (IGF2BP3) is a marker of unfavourable prognosis in colorectal cancer. European journal of cancer. 2012; 48(18):3405–13. https://doi.org/10.1016/j.ejca.2012.06.021 PMID: 22840368; PubMed Central PMCID: PMC3613860.

40. Lin CY, Chen ST, Jeng YM, Yeh CC, Chou HY, Deng YT, et al. Insulin-like growth factor II mRNA-binding protein 3 expression promotes tumor formation and invasion and predicts poor prognosis in oral squamous cell carcinoma. Journal of oral pathology & medicine: official publication of the International Association of Oral Pathologists and the American Academy of Oral Pathology. 2011; 40(9):699–705. https://doi.org/10.1111/j.1600-0714.2011.01019.x PMID: 23647076.

41. Malumbres M, Barbacid M. Cell cycle, CDKs and cancer: a changing paradigm. Nature reviews Cancer. 2009; 9(3):153–66. https://doi.org/10.1038/nrc2602 PMID: 19238148.

42. Nagahara M, Nishida N, Iwatsuki M, Ishimaru S, Mimori K, Tanaka F, et al. Kinesin 18A expression: clinical relevance to colorectal cancer progression. International journal of cancer. 2011; 129(11):2543–52. https://doi.org/10.1002/ijc.25916 PMID: 21213216.

43. Liao W, Huang G, Liao Y, Yang J, Chen Q, Xiao S, et al. High KIF18A expression correlates with unfavorable prognosis in primary hepatocellular carcinoma. Oncotarget. 2014; 5(21):10271–9. https://doi.org/10.18632/oncotarget.2082 PMID: 25431949; PubMed Central PMCID: PMC4279371.

44. Saez C, Japon MA, Ramos-Morales F, Romero F, Segura DI, Tortolero M, et al. hpttg is over-expressed in pituitary adenomas and other primary epithelial neoplasias. Oncogene. 1999; 18(39):5473–6. https://doi.org/10.1038/sj.onc.1202914 PMID: 10498901.

45. Zhang X, Horwitz GA, Heaney AP, Nakashima M, Prezant TR, Bronstein MD, et al. Pituitary tumor transforming gene (PTTG) expression in pituitary adenomas. The Journal of clinical endocrinology and metabolism. 1999; 84(2):761–7. https://doi.org/10.1210/jcem.84.2.5432 PMID: 10022450.

46. Boelaert K, McCabe CJ, Tannahill LA, Gittoes NJ, Holder RL, Watkinson JC, et al. Pituitary tumor transforming gene and fibroblast growth factor-2 expression: potential prognostic indicators in differentiated thyroid cancer. The Journal of clinical endocrinology and metabolism. 2003; 88(5):2341–7. https://doi.org/10.1210/jc.2002-021113 PMID: 12727994.

47. Heaney AP, Singson R, McCabe CJ, Nelson V, Nakashima M, Melmed S. Expression of pituitary-tumour transforming gene in colorectal tumours. Lancet. 2000; 355(9205):716–9. https://doi.org/10.1016/S0140-6736(99)10238-1 PMID: 10703804.

48. Rehfeld N, Geddert H, Atamna A, Rohrbeck A, Garcia G, Kliszewski S, et al. The influence of the pituitary tumor transforming gene-1 (PTTG-1) on survival of patients with small cell lung cancer and non-small cell lung cancer. Journal of carcinogenesis. 2006; 5:4. https://doi.org/10.1186/1477-3163-5-4 PMID: 16426442; PubMed Central PMCID: PMC1360069.

49. Davidson B, Nymoen DA, Elgaaen BV, Staff AC, Trope CG, Kaern J, et al. BUB1 mRNA is significantly co-expressed with AURKA and AURKB mRNA in advanced-stage ovarian serous carcinoma. Virchows Archiv: an international journal of pathology. 2014; 464(6):701–7. https://doi.org/10.1007/s00428-014-1577-7 PMID: 24756216.

50. Ricke RM, van Deursen JM. Aurora B hyperactivation by Bub1 overexpression promotes chromosome missegregation. Cell cycle. 2011; 10(21):3645–51. https://doi.org/10.4161/cc.10.21.18156 PMID: 22033440; PubMed Central PMCID: PMC3266005.