ORIGINAL RESEARCH

An integrated analysis for long noncoding RNAs and microRNAs with the mediated competing endogenous RNA network in papillary renal cell carcinoma

Chuiguo Huang[1,*]
Naijun Yuan[2,*]
Liying Wu[3]
Xiaofu Wang[1]
Junqiang Dai[4]
Pan Song[5]
Fengxi Li[6]
Changbao Xu[1]
Xinghua Zhao[1]

[1]Department of Urology, The Second Affiliated Hospital of Zhengzhou University, Zhengzhou, [2]College of Traditional Chinese Medicine of Jinan University, Institute of Integrated Traditional Chinese and Western Medicine of Jinan University, Guangzhou, [3]Department of Nephrology, The Second Affiliated Hospital of Hainan Medical College, Hainan, [4]Department of Neurosurgery, The Second Affiliated Hospital of Lanzhou University, Lanzhou, [5]Department of Urology, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, [6]Department of General Surgery, The First Affiliated Hospital of Guangxi Medical University, Guangxi, People's Republic of China

*These authors contributed equally to this work

Correspondence: Xinghua Zhao;
Chuiguo Huang
Department of Urology, The Second Affiliated Hospital of Zhengzhou University, Jingba Road, Zhengzhou 450014, Henan Province, People's Republic of China
Email xhzhao2006@163.com;
huangcg0727@163.com

Abstract: Papillary renal cell carcinoma (PRCC) is the second most common subtype of renal cell carcinoma, and it lacks effective therapeutic targets and prognostic molecular biomarkers. Attention has been increasingly focused on long noncoding RNAs (lncRNAs), which can act as competing endogenous RNA (ceRNA) to compete for shared microRNAs (miRNAs) in the tumorigenesis of human tumors. Therefore, to clarify the functional roles of lncRNAs with respect to the mediated ceRNA network in PRCC, we comprehensively integrated expression profiles, including data on mRNAs, lncRNAs and miRNAs obtained from 289 PRCC tissues and 32 normal tissues in The Cancer Genome Atlas. As a result, we identified 2,197 differentially expressed mRNAs (DEmRNAs) and 84 differentially expressed miRNAs (DEmiRNAs) using a threshold of |log2 (fold change)| >2.0 and an adjusted P-value <0.05. To determine the hub DEmRNAs that could be key target genes, a weighted gene co-expression network analysis was performed. A total of 28 hub DEmRNAs were identified as potential target genes. Seven dysregulated DEmiRNAs were identified that were significantly associated with the 28 hub potential target genes. In addition, we found that 16 differentially expressed lncRNAs were able to interact with the DEmiRNAs. Finally, we used Cytoscape software to visualize the ceRNA network with these differently expressed molecules. From these results, we believe that the identified ceRNA network plays a crucial role in the process of PRCC deterioration, and some of the identified genes are strongly related to clinical prognosis.

Keywords: papillary renal cell carcinoma, The Cancer Genome Atlas, weighted gene co-expression network analysis, long noncoding RNA, competing endogenous RNA network, survival prognosis

Introduction

Renal cell carcinoma (RCC) is a heterogeneous disease with varying histological abnormalities, cytogenetic aberrations, responses to therapy, and prognoses and accounts for 90% of renal malignancies in adults.[1] The World Health Organization (WHO) stratified RCC into multiple distinct subtypes, including clear cell, papillary, and chromophobe neoplasms. Papillary renal cell carcinoma (PRCC), the second most common subtype, accounts for about 10%–15% of RCC.[2] Within standard therapeutic protocols, malignant PRCC patients have a different pathological appearance and clinical outcome. Current treatment strategies of PRCC include surgery, radiation therapy, and chemotherapy, but a specialized, effective therapeutic remedy is still lacking.[3] Moreover, relevant molecular targets for clinical therapy are deficient, and thus require

further validation by more comprehensive clinical trials with a larger number of samples. Furthermore, molecular targets of drugs are subject to drug resistance and can cause toxicity leading to many restrictions in clinical application. Therefore, it is urgent and necessary that newly discovered treatments are found, based on the specific mechanisms of PRCC carcinogenesis, to improve treatment efficiency and avoid the adverse effects of conventional methods.

Advances in RNA sequencing technologies have brought to light the complexity of our genome. Noncoding RNAs (ncRNAs) account for the majority (98%) of the transcriptome, and different classes of regulative RNA with important biological functions are being continually discovered.[4] Investigating the world of RNA is one of the most important challenges facing biologists today, and long noncoding RNAs (lncRNAs) represent potential new biomarkers and drug targets. Currently, a growing body of evidence has indicated that there are interactions between lncRNAs and microRNAs (miRNAs), the downstream target genes of which have been closely related to tumor pathogenesis. In 2011, Salmena et al presented a competing endogenous RNA (ceRNA) hypothesis, which described a complex posttranscriptional regulatory network, including lncRNAs, mRNAs, and other types of RNAs, which can act as natural miRNA sponges to suppress miRNA functions by sharing one or more miRNA response elements.[5] Several experimental studies have also confirmed this hypothesis. For example, Cesana et al identified a muscle-specific lncRNA, known as long intergenic non-protein-coding RNA-muscle differentiation 1 (linc-MD1), which can mediate the time of muscle differentiation by acting as a ceRNA in human myoblasts.[6] Furthermore, Poliseno et al showed that phosphatase and tensin homolog pseudogene 1 (PTENP1) can upregulate the gene expression of PTEN by acting as a molecular sponge, causing adsorption of miR-19 and miR-20a in prostate cancer, inhibiting tumor cell growth.[7] However, to date, there is still little research on ceRNA-related lncRNAs involved in relevant regulatory mechanisms in PRCC.

Weighted gene co-expression network analysis (WGCNA) has been widely used in determining changes in transcriptome expression patterns in various diseases, finding clusters that include highly correlated genes, and summarizing these clusters by the module eigengenes.[8,9] Correlation networks facilitate network-based gene screening methods that can be used to identify candidate biomarkers or therapeutic targets. These methods have been successfully applied in complex biological contexts and fields, including cancer, mouse genetics, yeast genetics, and analyzing brain imaging data.[10] For example, carcinoembryonic antigen-related cell adhesion molecule 1 (CEACAM1), mitochondrial calcium

uniporter, voltage-dependent anion channel 1 (VDAC1), and cytochrome C (CYC) have been identified as prognostic biomarkers in the development of pancreatic ductal adenocarcinoma eigengenes.[11] Giulietti et al found that miR-302d-3p, miR-513c-5p, miR-3918, and miR-6076 could be candidate miRNA biomarkers for pancreatic ductal adenocarcinoma by targeting CYC.[12] Tang et al showed that five novel lncRNAs (CYP4F26P, RP11-108M12.3, RP11-38M8.1, RP11-54H7.4, and ZNF503-AS1), based on the prognostic signature, could be therapeutic targets for lung squamous cell carcinoma patients.[13] Consequently, focusing on the coherence function of network modules is a powerful systematic analysis method.[14]

The Cancer Genome Atlas (TCGA) project has provided a comprehensive way to improve our ability to diagnose, treat, and prevent tumor progression, through a better understanding of the genetic basis of cancer.[15] By the end of 2015, TCGA contained 33 types of cancer and >11,000 patients with gene expression profiles and relevant clinical data. Such an enormous amount of data provides an ideal opportunity to identify the mechanism and prognostic molecular signatures of cancer in a comprehensive way. Therefore, our research was based on TCGA, which analyzed the integrated RNA expression profiles from 289 PRCC samples and 32 normal samples. We also used the WGCNA to analyze the hub differentially expressed mRNAs (DEmRNAs), which can be target genes. The databases, miRTarBase and miRcode, were used to screen the key differentially expressed lncRNAs (DElncRNAs) and differentially expressed microRNAs (DEmiRNAs), as well as DEmiRNAs–DEmRNAs interactions. Hereafter, we constructed a ceRNA network to elaborate the interactions and potential cross talk between the hub DElncRNAs, DEmiRNAs, and DEmRNAs. In addition, relevant survival analyses were performed to determine the prognostic genes that possess clinical traits.

Materials and methods
Study population
We manually searched TCGA database, and a total of 289 PRCC cases were enrolled for comprehensive integrated analysis. In addition, we used the Data Transfer Tool (provided by GDC Apps) to download the level 3 mRNASeq gene expression data, miRNASeq data of samples, and clinical information of these patients (https://tcga-data.nci.nih.gov/). The sequenced data were derived from Illumina HiSeq RNASeq and Illumina HiSeq miRNASeq platforms. Our research meets the publication guidelines provided by TCGA (http://cancergenome.nih.gov/publications/publicationguidelines).

Differentially expressed analysis

The PRCC mRNASeq and miRNASeq data were derived from 321 samples, including 289 renal carcinoma tissue samples (cohort Tumor) and 32 normal tissue samples that were extracted from adjacent tissues (cohort Normal), and they were downloaded from TCGA. In addition, we merged data from tumor samples and normal samples, and deleted expressed data close to zero. In order to compare the normal group with PRCC, we used the "DESeq" package in R software to identify the DEmRNAs and miRNAs with thresholds of |log2 (fold change [FC])| >2.0 and adjusted P-value <0.05.[16]

The GENECARDS database (http://www.genecards.org/) was used to define and annotate the DElncRNAs.[17] In our study, we identified the DElncRNAs from differentially expressed RNAs, with the cutoff criteria of |log2FC| >2.0 and adjusted P-value <0.05.

Weighted correlation network analysis

WGCNA, a systematic biology method, describes the co-expression patterns between genes across microarray samples. WGCNA can be used for identifying modules of highly correlated genes, for summarizing these clusters using the module eigengene or an intramodular hub gene, for relating modules to one another and to external sample traits, and for calculating module membership measures. Co-expression networks facilitate network-based gene screening methods that can be used to identify candidate biomarkers or therapeutic targets.[9] Therefore, to identify the interactions between the differentially expressed genes (DEGs), the WGCNA, which uses the topological overlapping measurement, was performed to identify co-expression modules, with a power cutoff threshold of 12 and a module size cutoff ≤50 in this study. In addition, for investigating the relation between modules and clinical trait information, we plotted the heatmaps of all clusters at two different conditions (survival time and survival status) and showed the relation cutoff and P-value of the modules.

Functional annotation and protein–protein interaction (PPI)

To identify Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways of DEGs, the "clusterProfiler" package of bioconductor was performed for annotating the function of different module genes, with a threshold of a false discovery rate <0.05.[18]

In order to identify protein interactions between DEGs of different modules, we used the String (Search Tool for the Retrieval of Interacting Genes) online tools to construct a PPI network for DEGs (minimum required interaction score >0.4).[19] In addition, the Cytoscape software was used to visualize the PPI networks.

Construct the ceRNA network

To clarify the roles of lncRNA and miRNA with the mediated ceRNA network, we constructed a co-expression network of DEGs, lncRNAs, and miRNAs and visualized this network using Cytoscape software. We selected the hub DEmRNA from WGCNA modules as the target genes and discovered miRNA–mRNA pairs by miRTarBase (http://mirtarbase.mbc.nctu.edu.tw/).[20] The miRNA-targeted mRNAs were verified by various methods including the reporter assay, quantitative real-time PCR, Western blotting, microarrays, and next-generation sequencing experiments in miRTarBase. In addition, the lncRNA–miRNA interactions were determined based on the miRcode database, which is a comprehensive searchable online tool of putative miRNA target sites across the complete transcriptome annotation of Encyclopedia of DNA Elements (ENCODE), including >10,000 lncRNA genes in the current version (http://www.mircode.org/).[21]

Survival analysis

In order to identify the prognostic lncRNA and miRNA signature, by combining the clinical data of PRCC patients in TCGA, we established the life curves of these samples with hub miRNAs and some DElncRNAs. For the survival analysis for lncRNAs, we took advantage of a new and powerful web-based tool, Gene Expression Profiling Interactive Analysis (GEPIA), which can deliver fast and customizable functionalities based on TCGA and Genotype-Tissue Expression data.[22] As our research was based on the complete samples of PRCC in TCGA, we analyzed lncRNA by selecting samples based on the database (TCGA) and the associated disease (PRCC) in GEPIA. However, the expression analysis functions of GEPIA have not yet covered miRNAs, so we constructed the life curves of hub miRNAs by the "survival" package in R software. This analysis is based on Kaplan–Meier univariate survival analysis.[23] P-values of <0.05 were considered statistically significant.

Results
Identifying DEmRNAs, DEmiRNAs, and DElncRNAs

The expression profiles of PRCC patients and corresponding clinical information were downloaded from the Data Transfer Tool of TCGA database. After downloading TCGA data set, we found that the data were derived from 32 normal renal tissues and 289 PRCC tissues; the clinical data were derived

from 291 PRCC patient samples. However, there were two cases without complete clinical information, and thus, we deleted these two samples. Finally, we used 289 PRCC patient samples for further analysis. By comparing the normal group with the PRCC group, we identified the significant DEmRNAs and DEmiRNAs using the "DESeq" package in R software. A total of 2,197 DEmRNAs and 84 DEmiRNAs were confirmed by the "DESeq" package in R software. As a result, there were 1,218 (55.44%) upregulated and 979 (44.56%) downregulated DEGs. Moreover, a total of 58 (69.05%) upregulated and 26 (30.95%) downregulated DEmiRNAs were identified. The highest ten mRNAs with |log2FC| >2.0 and miRNAs with |log2FC| >2.0 are shown in Table 1.

According to the GENECARDS annotation, we identified a total of 79 aberrantly expressed lncRNAs in PRCC

tissues compared to normal tissues. The ten lncRNAs with the highest expression levels are listed in Table 1.

Construction of weighted gene co-expression modules

To find the functional clusters in PRCC patients, the gene co-expression network analysis of 2,197 DEGs was carried out by WGCNA. A total of four color modules (with the non-clustering DEmRNAs identified as gray) are shown in Figure 1. Moreover, we plotted the heatmap of module–trait relationships, which explored the relation between modules and two conditions (survival time and survival status). The heatmap is presented in Figure 2. We found that the gray module for survival time displayed the highest correlation coefficient of relational value (r)=0.45, P-value =0.01. The blue and turquoise modules were highly correlated to survival status (r=0.44, P=0.01), but the correlation with survival time was less pronounced. The yellow and brown modules show the correlation for survival status (yellow: r=0.39, P=0.02; brown: r=0.4, P=0.02).

PPI and functional annotation for the modules

To better understand the function and PPI of DEGs in different modules, we used the "clusterProfiler" package
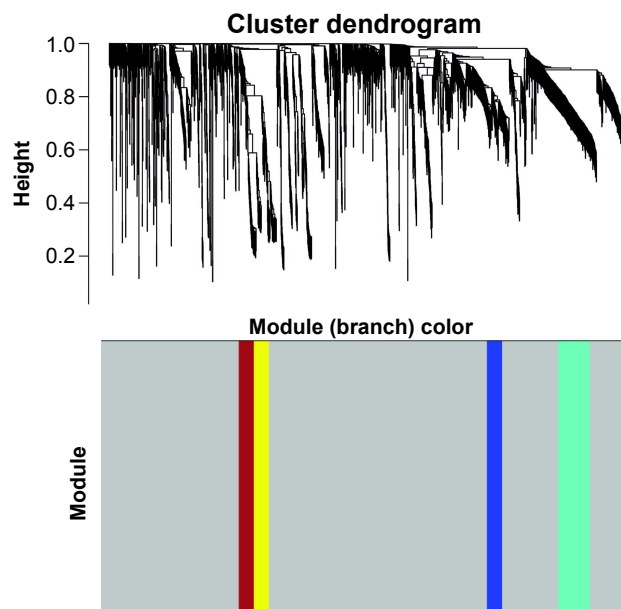
Table 1 The top 10 DEmRNAs, DEmiRNAs, and DElncRNAs

Gene symbol	logFC	Adjusted P-value	Stage
Top 10 DEmRNAs			
PAEP	9.699146376	2.21E–09	Up
KRT20	8.84162978	4.77E–10	Up
MAGEC2	8.476396027	1.79E–05	Up
MUC2	8.049151335	1.40E–08	Up
CPLX2	7.657755219	1.07E–07	Up
UMOD	−12.31646987	1.88E–25	Down
TMEM207	−8.713887124	2.02E–78	Down
CALB1	−8.346059544	4.66E–21	Down
CPNE6	−8.285605905	2.99E–12	Down
GP2	−8.264207662	4.65E–21	Down
Top 10 DEmiRNAs			
hsa-mir-519a	5.940804165	9.54E–06	Up
hsa-mir-875	5.910882711	5.26E–09	Up
hsa-mir-1297	5.68411564	5.99E–14	Up
hsa-mir-122	4.780014805	1.48E–04	Up
hsa-mir-211	4.52596231	1.15E–07	Up
hsa-mir-184	−5.299601274	2.65E–88	Down
hsa-mir-216b	−5.419488857	3.90E–67	Down
hsa-mir-33a	−2.577095453	3.97E–47	Down
hsa-mir-5708	−3.224770635	1.60E–44	Down
hsa-mir-145	−2.505152916	3.51E–39	Down
Top 10 DElncRNAs			
RP11-485M7.3	9.324918193	2.50E–05	Up
MUC2	8.049151335	1.40E–08	Up
LINC01029	6.789491286	2.95E–04	Up
LINC01234	6.492452763	1.79E–13	Up
HOTTIP	4.047013253	6.26E–06	Up
RP11-517M22.1	−7.584730102	3.00E–76	Down
LINC00982	−6.727620008	4.48E–74	Down
NPSR1-AS1	−5.395331033	6.73E–54	Down
TCL6	−4.55167931	1.41E–47	Down
WT1-AS	−3.300562406	2.63E–17	Down

Abbreviations: DEmRNAs, differentially expressed mRNAs; DEmiRNAs, differentially expressed microRNAs; DElncRNAs, differentially expressed long noncoding RNAs; FC, fold change.

Figure 1 Cluster dendrogram and color display of the co-expression network modules produced by average linkage hierarchical clustering of genes based on topological overlaps in the DEGs. Each branch in the dendrogram is a line that represents a single gene. Height indicates the Euclidean distance. Each color indicates a single module which contained genes with conservation closely in the data set. The area occupied by each color indicates the number of genes within the respective module.
Abbreviation: DEGs, differentially expressed genes.

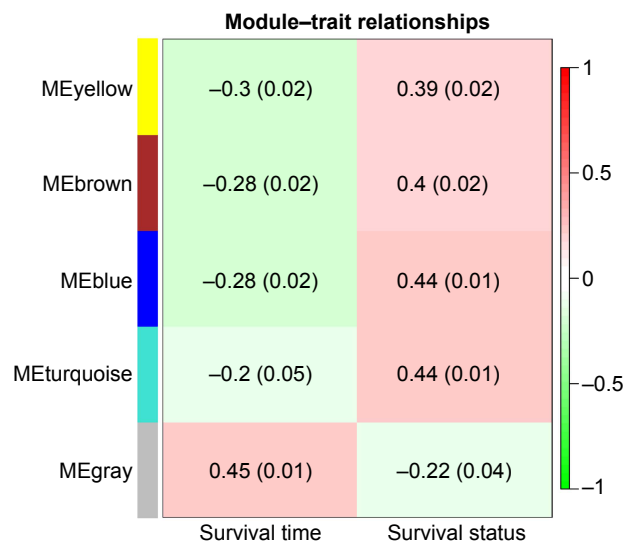**Module–trait relationships**

Figure 2 Module–trait relationships. Each row corresponds to a color module and column to clinical features (survival time and survival status). Each cell contains the corresponding relational value (r) in the first line and the P-value in the second line.
Abbreviation: ME, module eigengenes.

to show the enriched KEGG pathways (Table 2), where the Cytoscape software was applied for visualizing the PPI network (Figure 3). The blue module had seven nodes and eight interactions. All genes were downregulated DEmRNAs in this module. We found that solute carrier family 9 member A3 (SLC9A3) was central to the network for the blue module. In addition, we identified three enriched pathways in the blue module, and SLC9A3 as the hub gene was related to "Proximal tubule bicarbonate reclamation".

The brown module had five nodes and five interactions. In this module, we observed that tubulin alpha 3c (TUBA3C) was at the core of the PPI network, and five nodes were upregulated DEmRNAs. Moreover, there were five enriched pathways identified by clusterProfiler. We found that TUBA3C, as the hub gene, was involved in four pathways, namely, "Pathogenic *Escherichia coli* infection", "Gap junction", "Apoptosis", and "Phagosome".

Table 2 The KEGG pathways of WGCNA modules

Module names	KEGG pathways	Input number	Top genes	P-value
Blue module	hsa04915:Estrogen signaling pathway	2	HSPA2, KCNJ9	2.48E–03
	hsa04080:Neuroactive ligand–receptor interaction	2	NMUR2, PTGER1	1.74E–02
	hsa04964:Proximal tubule bicarbonate reclamation	2	SLC9A3, GADL1	1.77E–02
Brown module	hsa05130:Pathogenic *Escherichia coli* infection	2	TUBA3C, GABRG2	2.27E–03
	hsa04540:Gap junction	2	TUBA3C, GABRG2	5.55E–03
	hsa04210:Apoptosis	2	TUBA3C, GABRG2	1.33E–02
	hsa04145:Phagosome	2	TUBA3C, GABRG2	1.61E–02
	hsa04080:Neuroactive ligand–receptor interaction	2	CGA, GABRG2	4.67E–02
Gray module	hsa04080:Neuroactive ligand–receptor interaction	43	ADRA2B, ADRB1, CNR1, CCKBR, BDKRB2, NTSR2	1.38E–15
	hsa00830:Retinol metabolism	20	CYP1A2, UGT1A10, CYP1A1, DHRS9, RDH12, ADH1C	8.61E–12
	hsa04020:Calcium signaling pathway	30	CALML5, ADRB1, CCKBR, BDKRB2, PDGFRA, GNA14	8.61E–12
	hsa00980:Metabolism of xenobiotics by cytochrome P450	18	UGT2A1, AKR1C1, CYP1A2, CYP1A1, UGT1A10, DUX4	1.77E–09
	hsa01100:Metabolic pathways	77	PLA2G7, IDO2, ALAS2, AKR1D1, HMGCS2, PIK3C2G	7.59E–09
	hsa00982:Drug metabolism – cytochrome P450	16	UGT2A1, CYP1A2, UGT1A10, DUX4, ADH1C, ADH4	2.61E–08
	hsa00053:Ascorbate and aldarate metabolism	10	UGT1A10, UGT1A9, UGT1A8, UGT2A1, TSPY2, SLC14A2	6.57E–07
	hsa04261:Adrenergic signaling in cardiomyocytes	20	KCNQ1, CALML5, ADRB1, SLC8A1, PLN, CACNA1S	6.57E–07
	hsa04022:cGMP-PKG signaling pathway	21	ADRA2B, CALML5, ADRB1, SLC8A2, PLN, MYLK3	7.36E–07
	hsa00983:Drug metabolism – other enzymes	11	UGT1A10, UGT1A9, UGT1A8, UGT2A1, XDH, SLC14A2	4.49E–06
	hsa04960:Aldosterone-regulated sodium reabsorption	10	NAR4A2, PRKCG, SFN, SCNN1G, ATP1B2, ATP1A2	8.48E–06
	hsa04060:Cytokine-cytokine receptor interaction	24	IL19, IL12RB2, TNFSF9, IFNE, PDGFRA, IL20RB	1.71E–05
	hsa04966:Collecting duct acid secretion	8	SLC4A1, CLCNKB, ATP6V1B1, ATP6V1G3, ATP6V1C2, ATP6V0A4	3.80E–05
	hsa04961:Endocrine and other factor-regulated calcium reabsorption	9	SLC8A1, CALB1, BDKRB2, TRPV5, ATP1A2, DNM1	1.69E–04
	hsa04978:Mineral absorption	9	SLC8A1, TRPV6, TRPM6, MT1H, MT1G, ATP1A2	2.98E–04
	hsa04925:Aldosterone synthesis and secretion	11	CREB5, CYP21A2, PRKCG, LDLR, NR4A2, PDE2A	3.05E–04
	hsa04024:cAMP signaling pathway	17	CALML5, ADRB1, DRD1, PLN, SSTR2, CACNA1S	6.54E–04
	hsa04514:Cell adhesion molecules	14	ITGA8, NFASC, SELE, CLDN8, CNTN1, HLAG	8.62E–04
	hsa04145:Phagosome	14	FCGR2B, MBL2, TUBAL3, ATP6V1B1, FCGR1A, PLA2R1	1.45E–03
	hsa05202:Transcriptional misregulation in cancer	15	HPGD, BCL2A1, HMGA2, TLX3, PAX7, WT1	1.81E–03
	hsa04014:Ras signaling pathway	17	CALML5, PRKCG, PLA2G4F, FGF5, PLA2G3, KDR	2.25E–03

(Continued)

Table 2 (*Continued*)

Module names	KEGG pathways	Input number	Top genes	*P*-value
	hsa04151:PI3K–Akt signaling pathway	22	TNR, GHR, CREB5, RELN, PPP2R2C, NR4A1	2.27E–03
	hsa05200:Pathways in cancer	24	BDKRB2, PDGFRA, E2F2, DCC, PRKCG, WNT8B	7.48E–04
	hsa00350:Tyrosine metabolism	4	TYRP1, TYR, ADH1C, ADH4	1.09E–03
	hsa04110:Cell cycle	11	CDC25C, CDC6, PKMYT1, BUB1, CDKN2A, E2F2	6.79E–03
	hsa04010:MAPK signaling pathway	15	NTRK1, MECOM, PRKCG, IL1A, PLA2G4F, DUSP9	1.40E–02
	hsa05206:MicroRNAs in cancer	12	HMGA2, TNR, CDC25C, MIR34A, PRKCG, TIMP3	1.46E–02
	hsa04510:Focal adhesion	12	ITGA8, PRKCG, IBSP, KDR, PAK6, PDGFRA	1.56E–02
	hsa04115:p53 signaling pathway	6	BBC3, SFN, TP73, CDKN2A, SERPINB5, RRM2	1.57E–02
Turquoise module	hsa01100:Metabolic pathways	25	PNPLA3, HAO1, CYP4F2, ADH6, PAH, UPP2	4.71E–17
	hsa00010:Glycolysis/Gluconeogenesis	6	FBP1, G6PC, ADH6, ALDOB, PCK2	1.69E–08
	hsa04974:Protein digestion and absorption	5	SLC7A9, MME, SLC7A8, XPNPEP2, SLC6A19	2.59E–06
	hsa00260:Glycine, serine and threonine metabolism	4	AGXT, BHMT, DAO, PIPOX	3.08E–06
	hsa01200:Carbon metabolism	5	ALDH6A1, AGXT, FBP1, ALDOB, HAO1	7.51E–06
	hsa00590:Arachidonic acid metabolism	4	CYP4F2, GPX3, CYP2B6, CYP4A11	1.58E–05
	hsa00982:Drug metabolism – cytochrome P450	4	ADH6, CYP2B6, GSTA1, GSTA2	2.35E–05
	hsa03320:PPAR signaling pathway	4	CYP8B1, PCK2, FABP1, PCK1	2.76E–05
	hsa00980:Metabolism of xenobiotics by cytochrome P450	4	ADH6, CYP2B6, GSTA1, GSTA2	2.90E–05
	hsa04146:Peroxisome	4	AGXT, DAO, HAO1, PIPOX	4.69E–05
	hsa04152:AMPK signaling pathway	4	G6PC, FBP1, PCK2, PCK1	2.16E–04
	hsa04920:Adipocytokine signaling pathway	3	G6PC, PCK2, PCK1	6.18E–04
	hsa03013:RNA transport	2	RGPD3, PCDH15	6.99E–04
	hsa01230:Biosynthesis of amino acids	3	ASS1, ALDOB, PAH	7.50E–04
	hsa00360:Phenylalanine metabolism	2	HPD, PAH	8.41E–04
	hsa05204:Chemical carcinogenesis	3	ADH6, GSTA1, GSTA2	9.62E–04
	hsa04964:Proximal tubule bicarbonate reclamation	2	PCK2, PCK1	1.46E–03
	hsa04068:FoxO signaling pathway	3	G6PC, PCK2, PCK1	3.75E–03
	hsa04151:PI3K-Akt signaling pathway	3	G6PC, PCK2, PCK1	4.35E–02

Abbreviations: KEGG, Kyoto Encyclopedia of Genes and Genomes; WGCNA, weighted gene co-expression network analysis.

The turquoise module had 53 nodes and 76 interactions. We found some genes that were central to PPI, including patatin-like phospholipase domain-containing 3 (PNPLA3), albumin (ALB), microsomal triglyceride transfer protein (MTTP), and phosphoenolpyruvate carboxy-kinase 2 (PCK2), are the hub genes related to PRCC, and all the DEmRNAs were downregulated. In this module, a total of 19 pathways were enriched in the PPI network. PNPLA3 was enriched in the pathway of "Metabolic pathways". PCK2 was involved in the pathways of "Glycolysis/Gluconeogenesis" and the "PPAR signaling pathway". However, we found that there was no interaction between genes in the yellow module, and no significantly enriched KEGG pathway was identified.

The gray color cluster represented the non-clustering genes in WGCNA. Some non-clustering genes could serve as potential regulators that are associated with the progression of PRCC. Therefore, to better understand the functional annotation and the PPI, we constructed the PPI network and identified the KEGG pathways involved in these genes.

This PPI network showed 122 nodes and 962 interactions. The most significant hub genes were nuclear receptor subfamily 4 group A member 2 (NR4A2), dachshund family transcription factor 1 (DACH1), glypican 5 (GPC5), potassium voltage-gated channel subfamily H member 2 (KCNH2), specificity protein 1 (SP1), v-myc avian myelocytomatosis viral oncogene (MYCN), high mobility group AT-hook 2 (HMGA2), tenascin R (TNR), CAMP responsive element-binding protein 5 (CREB5), apoptosis regulator Bcl-2 (BCL2), cyclin-dependent kinase 2 (CDK2), and tissue inhibitor of metalloproteinases 3 (TIMP3). In this module, we found 29 significantly upregulated pathways and demonstrated that the crucial enriched genes were involved in biological processes and molecular metabolism, such as "Metabolism of xenobiotics by cytochrome P450", "Metabolic pathways", "Aldosterone-regulated sodium reabsorption", and "Cell cycle". The NR4A2 and CREB5 genes were identified as being related to "Aldosterone synthesis and secretion" pathways. In addition, CREB5 and TNR could
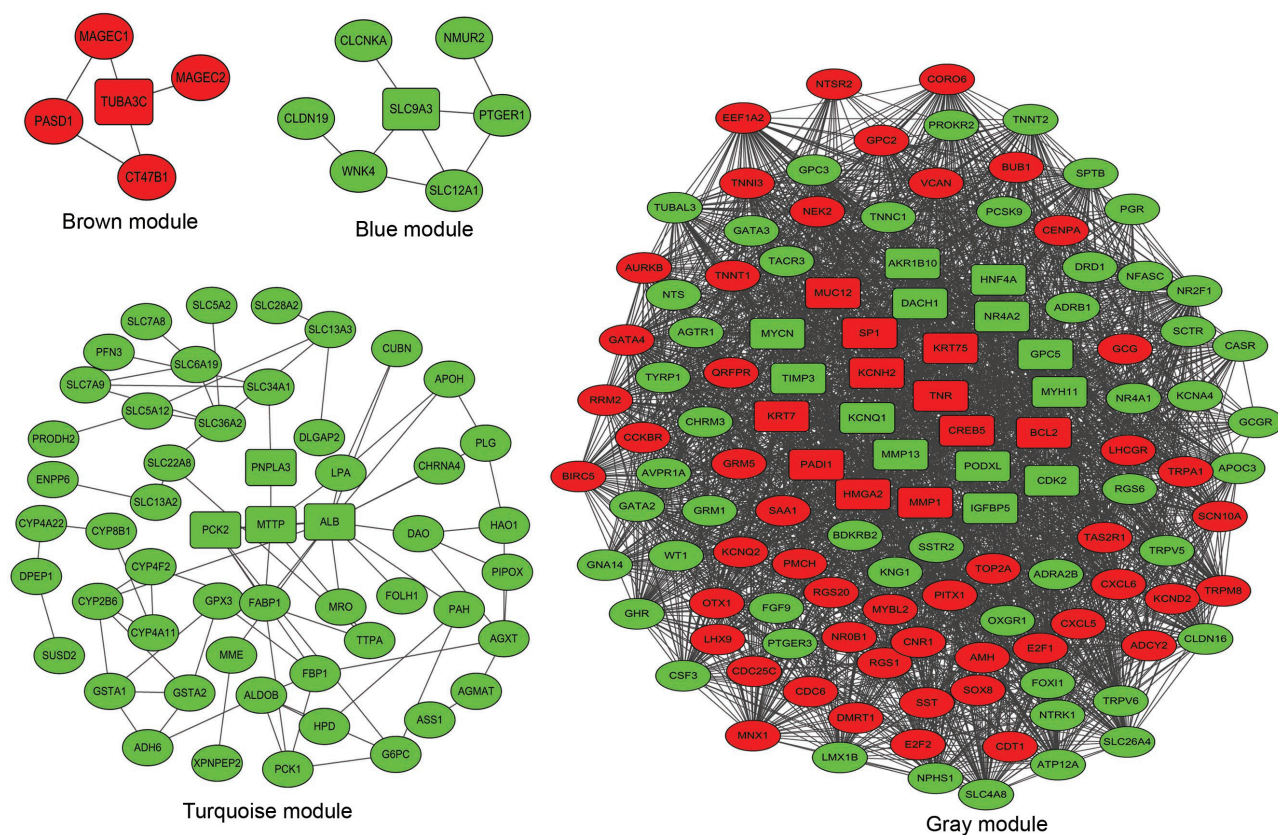
Figure 3 The PPI networks of DEmRNAs built by WGCNA modules in PRCC. Each ellipse corresponds to a protein-coding gene (mRNA). The key DEmRNAs at the core of modules are indicated as round rectangles. The green nodes denote that the genes are downregulated, while the red nodes denote that the genes are upregulated.
Abbreviations: PPI, protein–protein interaction; DEmRNAs, differentially expressed mRNAs; WGCNA, weighted gene co-expression network analysis; PRCC, papillary renal cell carcinoma.

play important roles in the "PI3K-Akt signaling pathway". TNR and HMGA2 were found to be enriched in "MicroRNAs in cancer". The pathway for "Transcriptional misregulation in cancer" included MYCN, HMGA2, and BCL2.

Construction of the ceRNA network in PRCC

To investigate DElncRNA mediation of PRCC in greater depth, we used Cytoscape software to construct the ceRNA network including lncRNA–miRNA–mRNA interactions (Figure 4A). We integrated the mRNASeq and miRNASeq data to identify the hub genes and DEmiRNAs, based on the above analytical methods. The DEmiRNA–mRNA pairs were constructed based on miRTarBase. A total of 28 hub DEmRNAs were identified as target genes by WGCNA. We found that hsa-mir-217, hsa-mir-133b, hsa-mir-216a, hsa-mir-133a, hsa-mir-145, hsa-mir-211, and hsa-mir-1297 were also closely related with these target DEmRNAs. Moreover, we found that 16 DElncRNAs were related with the key DEmiRNAs, based on retrieving miRcode database (Table 3). PVT1, as a key DElncRNA, had interactive associations with

mir-145, mir-211, mir-216a, mir-133a, and mir-133b in this network. The DElncRNA RP11-496D24.2 interacted with mir-145, mir-211, mir-216a, and mir-217. Taken together, we suggest that DElncRNA PVT1 and RP11-496D24.2 could be key regulators in the pathogenesis of PRCC by mediating the ceRNA network.

Survival analysis for the hub DEGs and DEmiRNAs

A total of 16 DElncRNAs and seven DEmiRNAs were identified as hub lncRNAs/miRNAs that were responsible for the tumorigenesis of PRCC. To investigate overall survival for these RNAs, we performed Kaplan–Meier curve analysis to obtain the prognostic signature. The survival analysis suggested that all key lncRNAs, except LINC00264, were related to the progression of PRCC (Figure S1). High expression of five DElncRNAs, including AC003092.1, PVT1, RP6-191P20.4, RP11-401P9.4, and RP11-496D24.2, was associated with poor prognosis (Figure 4B). Conversely, the other 10 DElncRNAs with highly expressed levels predicted a longer patient survival time. Similar to DEmiRNAs, survival
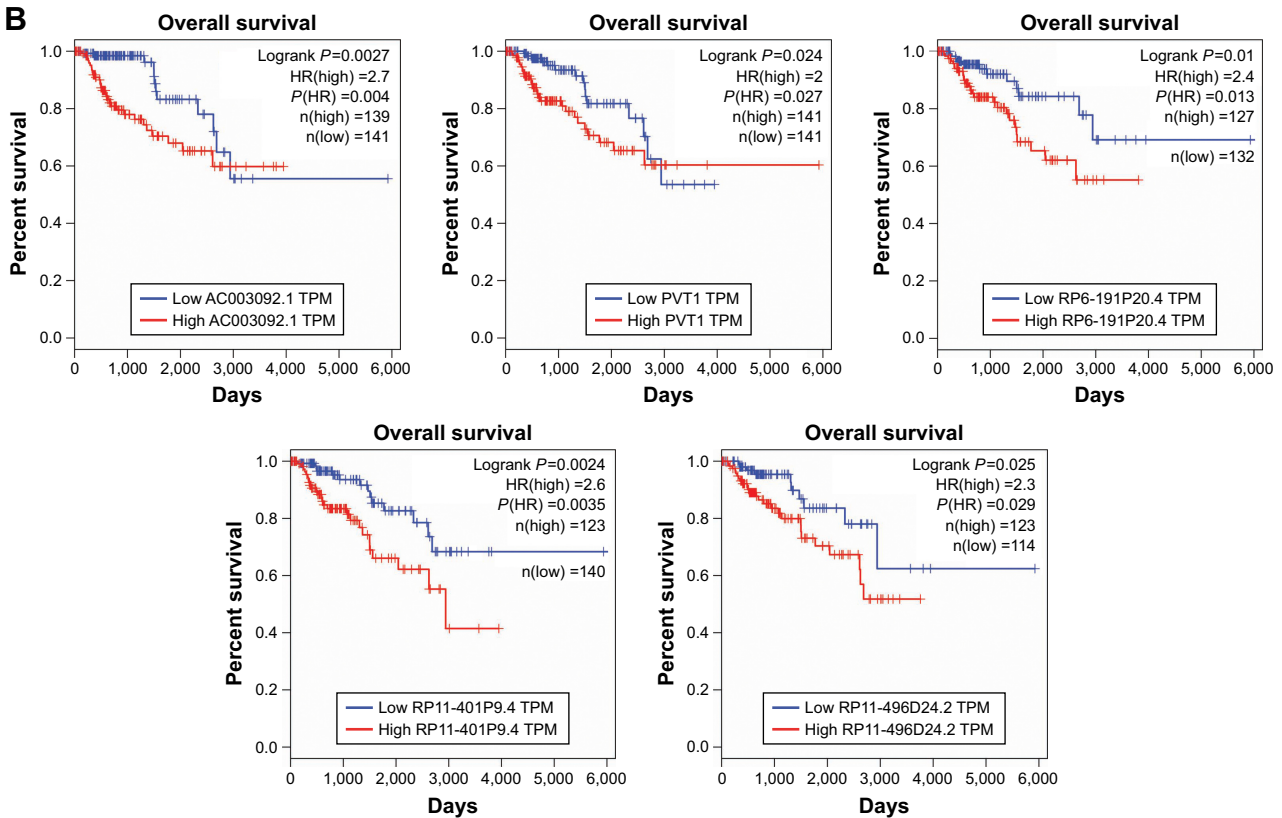
A
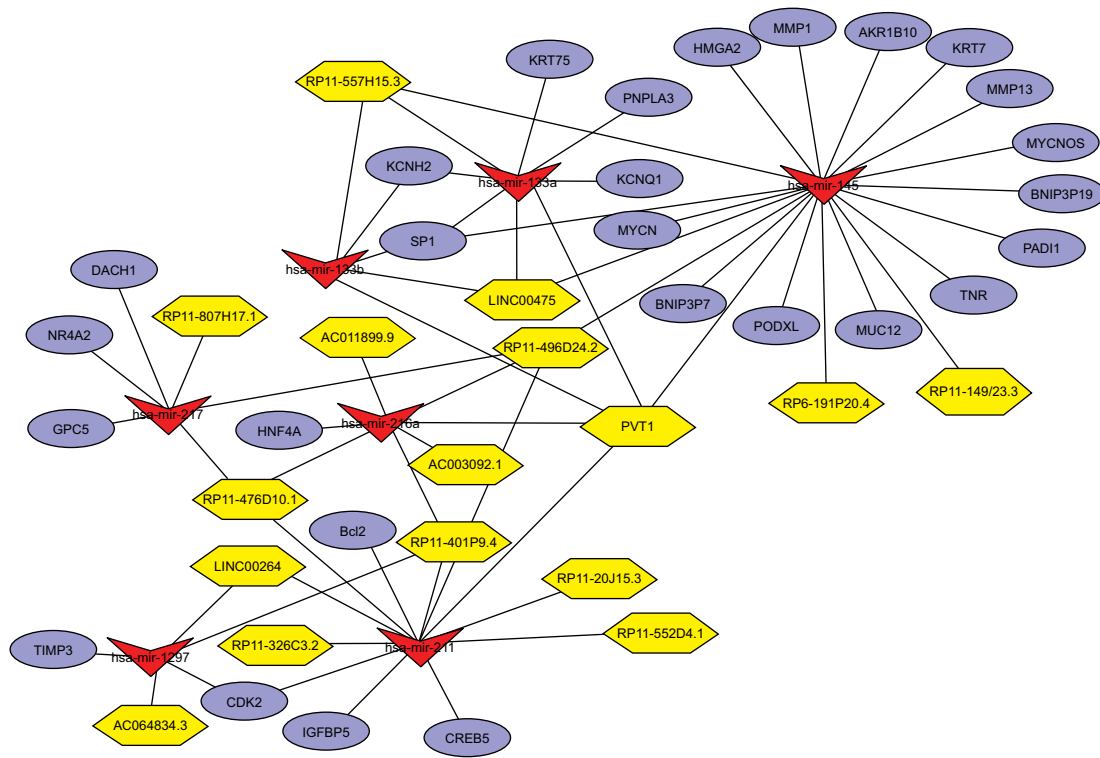
B

Figure 4 (**A**) The newly identified ceRNA network in PRCC. lncRNAs, miRNAs, and target mRNAs are indicated as yellow hexagons, red inverted triangles, and purple ellipses, respectively. (**B**) Kaplan–Meier curve analysis of five DElncRNAs for the overall survival in PRCC patients. High expression of five DElncRNAs, including AC003092.1, PVT1, RP6-191P20.4, RP11-401P9.4, and RP11-496D24.2, was associated with poor prognosis. Horizontal axis stands for overall survival time, while vertical axis stands for survival function.

Abbreviations: ceRNA, competing endogenous RNA; PRCC, papillary renal cell carcinoma; lncRNAs, long noncoding RNAs; miRNAs, microRNAs; DElncRNAs, differentially expressed lncRNAs.

Table 3 The DElncRNAs, DEmiRNAs, and DEmRNAs preserved in ceRNA network

The type of RNAs	Gene symbols
DElncRNAs	AC003092.1, AC011899.9, AC064834.3, LINC00264, LINC00475, PVT1, RP6-191P20.4, RP11-557H15.3, RP11-807H17.1 RP11-20J15.3, RP11-149I23.3, RP11-326C3.2, RP11-401P9.4, P11-476D10.1, RP11-496D24.2, RP11-552D4.1
DEmiRNAs	hsa-mir-217, hsa-mir-133b, hsa-mir-216a, hsa-mir-133a, hsa-mir-145, hsa-mir-211, hsa-mir-1297
DEmRNAs	NR4A2 (nuclear receptor subfamily 4 group A member 2), GPC5 (glypican 5), TIMP3 (tissue inhibitor of metalloproteinases 3), DACH1 (dachshund family transcription factor 1), CDK2 (cyclin-dependent kinase 2), HNF4A (hepatocyte nuclear factor 4 alpha), BCL2 (apoptosis regulator Bcl-2), IGFBP5 (insulin-like growth factor-binding protein 5), CREB5 (cAMP responsive element-binding protein 5), SP1 (specificity protein 1), KCNH2 (potassium voltage-gated channel subfamily H member 2), KRT75 (keratin 75), PNPLA3 (patatin-like phospholipase domain-containing 3), KCNQ1 (potassium voltage-gated channel subfamily Q member 1), MYCN (v-myc avian myelocytomatosis viral oncogene), BNIP3P7 (BCL2-interacting protein 3 pseudogene 7), PODXL (podocalyxin-like), MUC12 (mucin 12, cell surface associated), TNR (tenascin R), PADI1 (peptidyl arginine deiminase 1), MYCNOS (MYCN opposite strand), MMP13 (matrix metallopeptidase 13), AKR1B10 (aldo-keto reductase family 1 member B10), HMGA2 (high-mobility group AT-hook 2)

Abbreviations: DElncRNAs, differentially expressed long noncoding RNAs; DEmiRNAs, differentially expressed microRNAs; DEmRNAs, differentially expressed mRNAs; ceRNA, competing endogenous RNA.

analysis for the seven miRNAs showed that all DEmiRNAs, except mir-211, have negative effects on overall survival of patients (Figure 5).

Discussion

PRCC, as the second most common histological subtype neoplasm affecting the kidney, has been intensively investigated to identify the key regulatory genes and molecules. Most attention has been focused on the study of tumor progression mediated through dysregulated lncRNAs, and the influence that neoplastic metastasis has on prognosis. However, to our knowledge, few studies have been conducted to predict the prognosis of PRCC, and no reliable papillary renal cell cancer-specific lncRNAs, as biomarkers for the detection and risk stratification of PRCC, have been identified. Therefore, we comprehensively integrated mRNA and miRNA data from TCGA, using WGCNA to screen out the tumor-associated genes, and determined the lncRNA–miRNA–mRNA interactions involved in the ceRNA network to further investigate the regulatory mechanism of lncRNAs.

In this study, 16 DElncRNAs were identified as hub lncRNAs, and six of these (PVT1, PR11-401P9.4, PR11-496D24.2, PR11-476D10.1, PR11-557H15.3, and LINC00475), which were distinctly enriched in the ceRNA network, could be prognostic biomarkers for PRCC. After detailed observation of the ceRNA network, we noticed that two DElncRNAs, PVT1 and RP11-496D24.2, not only had more interactions but may also be involved in regulating the target genes by competing for three common miRNAs (mir-145, mir-211, and mir-216a).

lncRNAs are defined as non-protein-coding transcripts longer than 200 nucleotides.[24] However, very recent research

has indicated that some lncRNAs, playing widespread roles acting as gene regulators and in other cellular processes, can encode proteins. PVT1 has been shown to represent an lncRNA locus, across a genome interval of over 300 kb on the forward strand of chromosome 8, and this had been considered a candidate oncogene.[25] Overexpression of PVT1 has been confirmed to be associated with many types of cancers, including breast and ovarian cancers, acute myeloid leukemia, and Hodgkin lymphoma.[26–29] In this study, our analysis found that high expression of PVT1 may compete with three key DEmiRNAs (miR-133a, miR-133b, and miR-145) to mediate the expression of target genes. Moreover, patients with highly expressed PVT1 had shorter survival than those with low expression. It can be suggested that PVT1 knockdown may prevent cell proliferation in PRCC, and prolong survival time of patients. Furthermore, we noticed that high expression of RP11-496D24.2 also demonstrated a poor prognosis in patients, but there is little known about this lncRNA. Further investigations are required to clarify the function of RP11-496D24.2 in PRCC and other renal tumors.

miRNAs function as master regulators and signal modulators by fine-tuning gene expression in many complex pathways. Disrupting these miRNAs may result in a permissive tumorigenic state.[30] Many dysregulated miRNAs in adenocarcinoma have been reported to play various roles in carcinogenesis. To our knowledge, we found that seven miRNAs, namely, mir-217, mir-133b, mir-216a, mir-133a, mir-145, mir-211, and mir-1297, involved in the ceRNA network, can be supportive to further investigation. mir-145 and mir-211, at the core of the network, were particularly important. Wach et al showed that increased expression of mir-145 in PRCC subtypes was associated with worse outcome.[31]
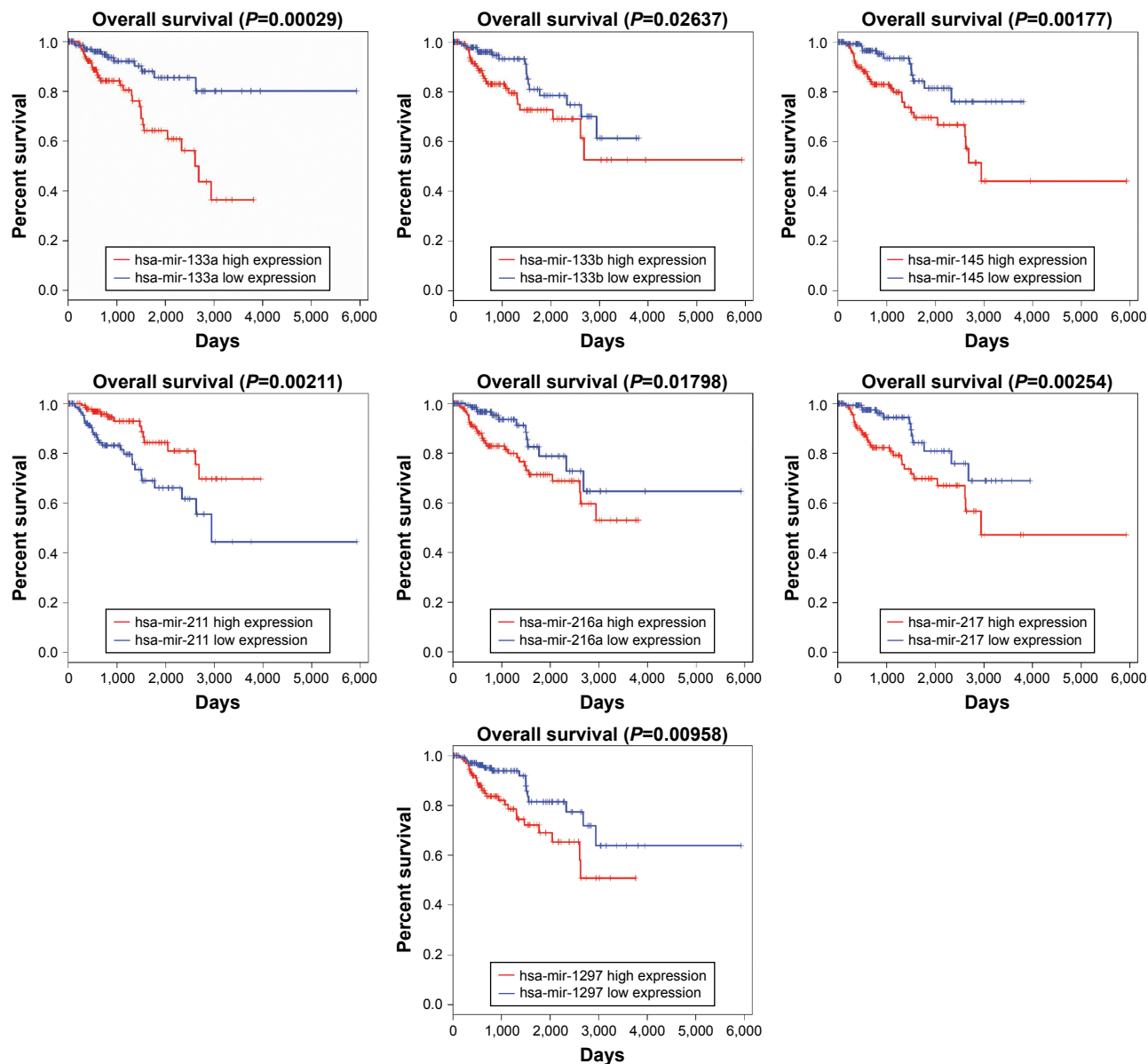
Figure 5 Kaplan–Meier curve analysis of all DEmiRNAs for the overall survival in PRCC patients. Survival analysis of seven miRNAs exhibited that all DEmiRNAs except mir-211 revealed negative effects on patient overall survival. Horizontal axis stands for overall survival time, while vertical axis stands for survival function.
Abbreviations: DEmiRNAs, differentially expressed miRNAs; PRCC, papillary renal cell carcinoma; miRNAs, microRNAs.

Furthermore, survival analysis also indicated that overexpression of mir-145 predicted a shorter lifetime compared to patients with lower mir-145 expression in our research. Regarding mir-211, previous reports have revealed that it can function as a metabolic switch in human melanoma cells, regulate CD44-dependent cell growth in colorectal cancer, and downregulate special AT-rich sequence-binding protein 2 (SATB2) to suppress hepatocellular carcinoma.[32–34] Our data have shown that there is a potential mir-211-mediated interaction with PRCC. In addition, survival analysis demonstrated that high mir-211 expression can prolong patient survival time. However, these findings require further

research to identify whether these miRNAs have a specific role in PRCC tumorigenesis.

According to the WGCNA, we found that many protein-coding genes from the gray module, such as SP1, CDK2, and MYCN, considered as important target genes, were involved in this newly identified ceRNA network. SP1 demonstrated the most interactions with DEmiRNA in the network. Previous studies have confirmed that SP1 is a specific transcription factor that can recognize GC-rich motifs and bind to DNAs or transcription activators by zinc fingers.[35] Moreover, the protein encoded by SP1 has been identified in various cellular processes, such as cell cycle, immune responses, response

to DNA damage, and chromatin remodeling. In addition, SP1 can also function as an activator or a repressor though posttranslational modifications (acetylation, phosphorylation, glycosylation, and proteolytic processing).[35,36] In the construction of the ceRNA network, SP1 can be a direct target of three hub DEmiRNAs, including miR-133a, miR-133b, and miR-145. Therefore, these DEmiRNAs may compete with a common DElncRNA PVT1. Therefore, we suggested that PVT1 may be a ceRNA to investigate tumor-suppressing effects in PRCC, via downregulation of SP1.

Although ncRNAs and miRNAs have received increasingly more attention in recent years, research focusing on the remarkable genes which code proteins is still ongoing. A similar study based on TCGA database has been conducted by He et al.[37] They used the WGCNA method to find out that 17 hub DEmRNAs may serve as potential biomarkers and associate with distinct various pathological stages of PRCC. However, this study only made use of the mRNASeq data, and there were no combinations with lncRNAs, miRNAs, or other types of biological molecules. Conversely, our research has revealed that these specific lncRNAs and miRNAs interact with the key DEmRNAs involved in successful construction of the lncRNA–miRNA–mRNA ceRNA network. Due to the differences in types of RNASeq gene expression, methods used for data processing, and cutoff criteria used for analyzing DEGs, there is no overlap between our research and He et al's study.

Conclusion

lncRNA-related studies have been increasingly conducted to investigate many biological processes, but few have characterized that lncRNAs, as ceRNAs, can regulate the tumorigenesis of PRCC. We constructed the ceRNA network to display a novel ceRNA regulatory mechanism for further investigation.

Acknowledgment

Disclosure

The authors declare that they have no conflicts of interest in this work.

References

1. Rini BI, Campbell SC, Escudier B. Renal cell carcinoma. *Lancet*. 2009;373(9669):1119–1132.
2. Moch H, Cubilla AL, Humphrey PA, Reuter VE, Ulbright TM. The 2016 WHO classification of tumours of the urinary system and male genital organs-part A: renal, penile, and testicular tumours. *Eur Urol*. 2016;70(1):93–105.
3. Low G, Huang G, Fu W, Moloo Z, Girgis S. Review of renal cell carcinoma and its common subtypes in radiology. *World J Radiol*. 2016;8(5):484–500.
4. Baltimore D. Our genome unveiled. *Nature*. 2001;409(6822): 814–816.
5. Salmena L, Poliseno L, Tay Y, Kats L, Pandolfi PP. A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? *Cell*. 2011; 146(3):353–358.
6. Cesana M, Cacchiarelli D, Legnini I, et al. A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. *Cell*. 2011;147(2):358–369.
7. Poliseno L, Salmena L, Zhang J, Carver B, Haveman WJ, Pandolfi PP. A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature*. 2010;465(7301):1033–1038.
8. Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol*. 2005;4:Article17.
9. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008;9:559.
10. Ruan X, Li H, Liu B, et al. A novel method to identify pathways associated with renal cell carcinoma based on a gene co-expression network. *Oncol Rep*. 2015;34(2):567–576.
11. Giulietti M, Occhipinti G, Principato G, Piva F. Weighted gene co-expression network analysis reveals key genes involved in pancreatic ductal adenocarcinoma development. *Cell Oncol (Dordr)*. 2016;39(4): 379–388.
12. Giulietti M, Occhipinti G, Principato G, Piva F. Identification of candidate miRNA biomarkers for pancreatic ductal adenocarcinoma by weighted gene co-expression network analysis. *Cell Oncol*. 2017;40(2): 181–192.
13. Tang RX, Chen WJ, He RQ, et al. Identification of a RNA-Seq based prognostic signature with five lncRNAs for lung squamous cell carcinoma. *Oncotarget*. Epub 2017 Apr 13.
14. Zhao W, Langfelder P, Fuller T, Dong J, Li A, Hovarth S. Weighted gene coexpression network analysis: state of the art. *J Biopharm Stat*. 2010;20(2):281–300.
15. Tomczak K, Czerwińska P, Wiznerowicz M. The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Contemp Oncol (Pozn)*. 2015;19(1A):A68–A77.
16. Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biol*. 2010;11(10):R106.
17. Stelzer G, Dalah I, Stein TI, et al. In-silico human genomics with GeneCards. *Hum Genomics*. 2011;5(6):709–717.
18. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. 2012; 16(5):284–287.
19. Szklarczyk D, Morris JH, Cook H, et al. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res*. 2017;45(Database issue):D362–D368.
20. Hsu SD, Tseng YT, Shrestha S, et al. miRTarBase update 2014: an information resource for experimentally validated miRNA-target interactions. *Nucleic Acids Res*. 2014;42(Database issue):D78–D85.
21. Jeggari A, Marks DS, Larsson E. miRcode: a map of putative microRNA target sites in the long non-coding transcriptome. *Bioinformatics*. 2012;28(15):2062–2063.
22. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res*. Epub 2017 Apr 12.
23. Goeman JJ, Oosting J, Cleton-Jansen AM, Anninga JK, van Houwelingen HC. Testing association of a pathway with survival using gene expression data. *Bioinformatics*. 2005;21(9):1950–1957.
24. Ponting CP, Oliver PL, Reik W. Evolution and functions of long non-coding RNAs. *Cell*. 2009;136(4):629–641.
25. Wang L, He JH, Han ZP. Characteristics of PVT1 and its roles in diseases. *Chin Med Sci J*. 2014;29(4):236–238.
26. Conte F, Fiscon G, Chiara M, Colombo T, Farina L, Paci P. Role of the long non-coding RNA PVT1 in the dysregulation of the ceRNA-ceRNA network in human breast cancer. *PLoS One*. 2017;12(2):e0171661.

27. Guan Y, Kuo WL, Stilwell JL, et al. Amplification of PVT1 contributes to the pathophysiology of ovarian and breast cancer. *Clin Cancer Res*. 2007;13(19):5745–5755.

28. Taniwaki M. [Recent advancements in molecular cytogenetics for hematological malignancies: identification of novel PVT1 fusion genes]. *Rinsho Ketsueki*. 2015;56(10):2056–2065. Japanese [with English abstract].

29. Ghesquières H, Larrabee BR, Casasnovas O, et al. A susceptibility locus for classical Hodgkin lymphoma at 8q24 near MYC/PVT1 predicts patient outcome in two independent cohorts. *Br J Haematol*. Epub 2016 Sep 9.

30. Adams BD, Kasinski AL, Slack FJ. Aberrant regulation and function of microRNAs in cancer. *Curr Biol*. 2014;24(16):R762–R776.

31. Wach S, Nolte E, Theil A, et al. MicroRNA profiles classify papillary renal cell carcinoma subtypes. *Br J Cancer*. 2013;109(3):714–722.

32. Mazar J, Qi F, Lee B, et al. MicroRNA 211 functions as a metabolic switch in human melanoma cells. *Mol Cell Biol*. 2016;36(7):1090–1108.

33. Wu X, He X, Li S, Xu X, Chen X, Zhu H. Long non-coding RNA ucoo2kmd.1 regulates CD44-dependent cell growth by competing for miR-211-3p in colorectal cancer. *PLoS One*. 2016;11(3):e0151287.

34. Jiang G, Cui Y, Yu X, Wu Z, Ding G, Cao L. miR-211 suppresses hepatocellular carcinoma by downregulating SATB2. *Oncotarget*. 2015;6(11):9457–9466.

35. Chu S. Transcriptional regulation by post-transcriptional modification – role of phosphorylation in Sp1 transcriptional activity. *Gene*. 2012; 508(1):1–8.

36. Chang WC, Hung JJ. Functional role of post-translational modifications of Sp1 in tumorigenesis. *J Biomed Sci*. 2012;19:94.

37. He Z, Sun M, Ke Y, et al. Identifying biomarkers of papillary renal cell carcinoma associated with pathological stage by weighted gene co-expression network analysis. *Oncotarget*. 2017;8(17):27904–27914.
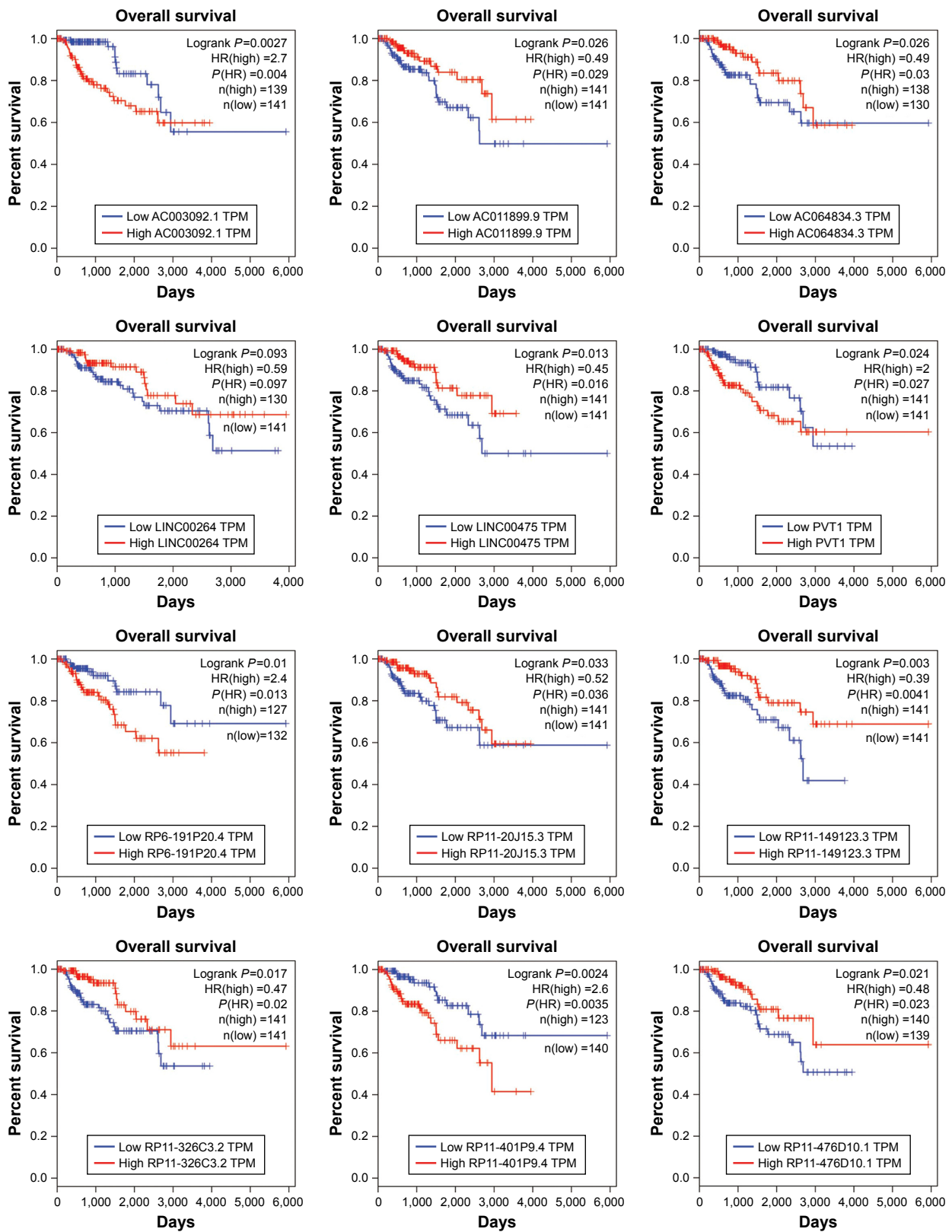
Supplementary material
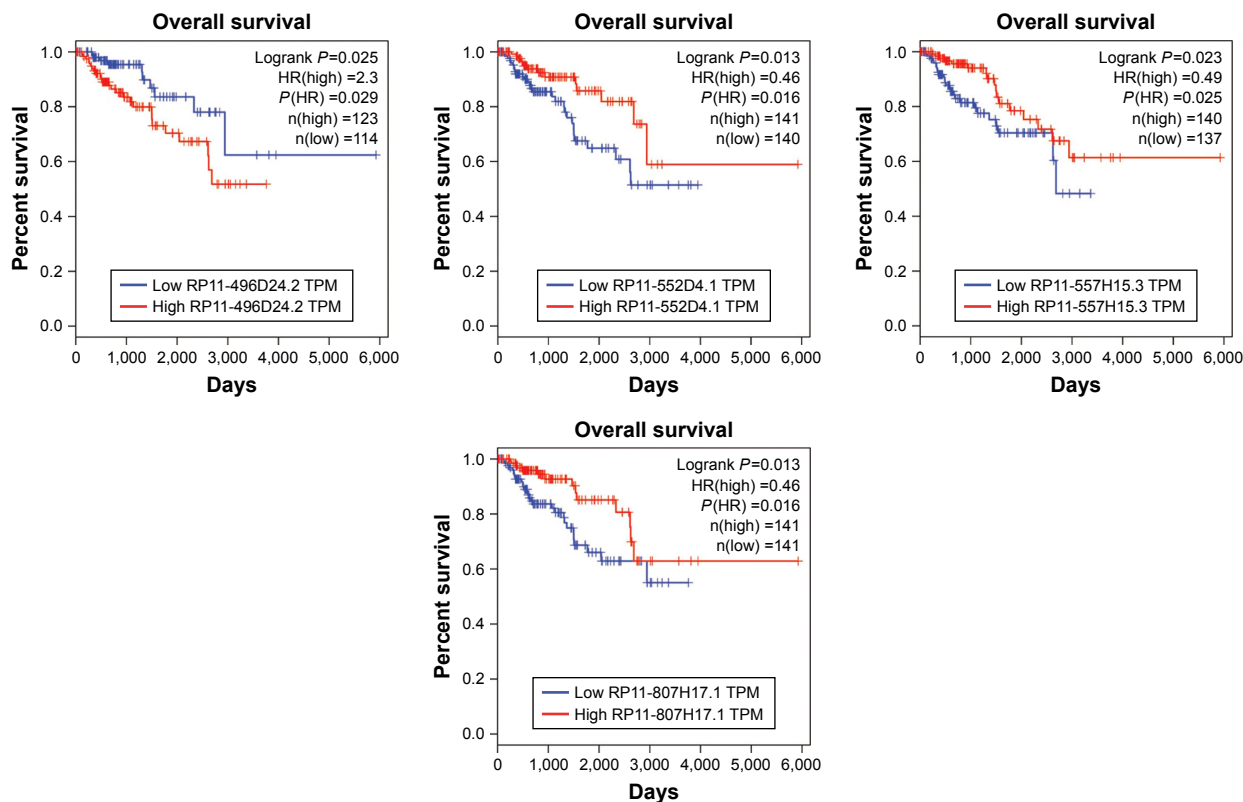

Figure S1 (*Continued*)

Figure S1 Kaplan–Meier curve analysis of all DElncRNAs for the overall survival in PRCC patients.
Abbreviations: DElncRNAs, differentially expressed lncRNAs; PRCC, papillary renal cell carcinoma.