# A prognostic nomogram based on competing endogenous RNA network for clear-cell renal cell carcinoma

**Yun Peng**
Department of Urology, Tianjin Medical University Second Hospital, Hexi, Tianjin, 300000, China
**Shangrong Wu**
Department of Urology, Tianjin Medical University Second Hospital, Hexi, Tianjin, 300000, China
**Zihan Xu**
Department of Urology, Tianjin Medical University Second Hospital, Hexi, Tianjin, 300000, China
**Dingkun Hou**
Department of Urology, Tianjin Medical University Second Hospital, Hexi, Tianjin, 300000, China
**Nan Li**
Department of Urology, Tianjin Medical University Second Hospital, Hexi, Tianjin, 300000, China
**Zheyu Zhang**
Department of Urology, Tianjin Medical University Second Hospital, Hexi, Tianjin, 300000, China
**Lili Wang**
Department of Oncology, Tianjin Medical University Second Hospital, Hexi, Tianjin, 300000, China
**Yangyi Zhang**
Department of Urology, Tianjin Medical University Second Hospital, Hexi, Tianjin, 300000, China
**Haitao Wang** ( haitao_peterrock@outlook.com )
Department of Oncology, Tianjin Medical University Second Hospital, Pingjiang Road, Hexi, Tianjin, China https://orcid.org/0000-0003-4345-9375

**Research**

**Keywords:** Renal Cell Carcinoma, ccRCC, competing endogenous RNA, ceRNA, Nomograms, Survival, Prognosis

**Posted Date:** March 31st, 2020

**DOI:** https://doi.org/10.21203/rs.3.rs-19720/v1

# Abstract

Backgroud

Clear-cell renal cell carcinoma (ccRCC) is stubborn to traditional chemotherapy and radiation treatment, which makes its clinical management a major challenge. Recently, we have made efforts to understand the etiology of ccRCC. Increasing evidence revealed that the competing endogenous RNA (ceRNA) were involved in the development of various tumor. However, it's scant for studying on ccRCC, and a comprehensive analysis of prognostic model based on lncRNA-miRNA-mRNA ceRNA regulatory network of ccRCC with large-scale sample size and RNA-sequencing expression data is still limited.

Methods

RNA-sequencing expression data were taken out from GTEx database and TCGA database, A total of 354 samples with ccRCC and 157 normal controlled samples were included in our study. The ccRCC-specific genes were obtained from WGCNA and differential expression analysis. Following, the communication between mRNAs and lncRNAs and target miRNAs were predicted by MiRcode, starBase, miRTarBase, and TargetScan. A gene signature of eight genes was constructed by univariate Cox regression, lasso methods and multivariate Cox regression analysis.

Results

A total of 2191 mRNAs and 1377 lncRNAs was identified, and a dys-regulated ceRNA network for ccRCC was established using 7 mRNAs, 363 lncRNAs, and 3 miRNAs. Further, a gene signature in cluding 8 genes based on this ceRNA was constructed, meanwhile, a nomogram predicting 1-, 3-, 5-year survival probability containing both clinical characteristics and ccRCC-specific gene signatures was developed.

Conclusion

It could contribute to a better understanding of ccRCC tumorigenesis mechanism and guide clinicians to make a more accurate treatment decision.


# Background

Kidney cancer which is the third most prevalent malignant tumor in the urogenital system of women and the second in men accounts to about 144,000 deaths annually worldwide(1). The most common subtype of renal cell carcinoma (RCC) is clear-cell renal cell carcinoma (ccRCC). However, since a lack of external tumor factors such as age, nuclear grading and microscopic tumor nectosis, it remains controversial for the optimum stratification of patients with ccRCC to use the TNM staging system(2), hence, identification of prognostic predictive system for ccRCC containing both tumor anatomical features and other clinical and genetic variables deserves increasing attention.

Recently the hypothesis of competing endogenous RNA (ceRNA) states that the pool of long non-coding RNAs (lncRNAs) can regulate messenger RNAs (mRNAs) activity by binding to and compete microRNAs (miRNAs)(3)(4). Combining with the miRNA response elements (MREs) on the target mRNAs, MiRNA can regulate the expression level of the target gene and lncRNA can serve as a molecular sponge to interact with miRNAs, which results in different kinds of human diseases process(5). At present, a lot of evidences showed that the ceRNA hypothesis was involved in the development of different kinds of tumors, such as gastric, colon, liver, breast, bladder and pancreatic cancer which makes it possible to construct a prognostic prediction system on the basis of ceRNA network. Nevertheless, there are limited prognosis-related researches conducted based on the dys-regulated ceRNA network in ccRCC. In this way, the present study pointed to explore the prognostic significance of gene contained in the ccRCC-specific dys-regulated ceRNA network.

In the current study, ccRCC-specific genes were obtained by employing weighted correlation network analysis (WGCNA)(6) and differential expression analysis to RNA-Seq data from Genotype-Tissue Expression (GTEx)(7) and The Cancer Genome Atlas (TCGA)(8), then miRNA database was used to predict the interaction between mRNAs or lncRNAs and miRNAs. Following, 7 mRNAs, 363 lncRNAs, and 3 miRNAs were found to assemble a dys-regulated ceRNA network for ccRCC. Further, A gene signature of one mRNA (MPP5) and seven lncRNAs (WT1-AS, AC114316.1, AC103719.1, AL162377.1, HS1BP3-IT1, LINC02657, AC015909.1) was constructed by univariate Cox regression, lasso methods and multivariate Cox regression analysis based on the regression coefficient, consequently, a prognostic nomogram assessment system predicting 1-, 3-, 5-year survival probability was constructed by including the gene signature and related clinical characteristics using a stepwise cox regression for ccRCC.

## Methods

# Gene Expression and Clinical Data

The High-throughput RNA sequencing data of 539 kidney samples with ccRCC and 72 normal controlled samples were obtained from TCGA data repository using TCGAbiolinks(9) R package. After filtering samples with tumor purity below 0.6(10), Formalin-Fixed Paraffin-Embedded (FFPE) tissue and duplicated samples, a total of 354 ccRCC samples and 72 normal controls were included in our analysis. miRNA sequencing data were also retrieved from TCGA(8). Subsequently, A total of 85 normal kidney cortex samples were obtained from GTEx(7) (version V8). No further approval was required from the Ethics Committee as the data comes from the TCGA and GTEx database. lncRNAs and mRNAs were recognized by the Ensembl(11) database (version GRCh38.98). lncRNAs and mRNAs that were not included in the database were excluded in this study. The batch effect was removed by limma(12). We mainly used the R program(version 3.6.1)(13) for analysis of RNA data.

# ccRCC-specific mRNAs And lncRNAs

WGCNA was used to identify coexpression network in lncRNAs or mRNAs expression profiles with different traits. the biweight midcorrelation efficiently analysis(14) was used to assess Weighted coexpression relationship. In this study, we obtained the most related mRNAs or lncRNAs with ccRCC patients by using WGCNA. significant differentially expressed genes between ccRCC and normal samples were identified by DESeq2 with threshold of |log2 fold change| > 1 and adjusted P value < 0.01. Consequently, genes most positive correlated with ccRCC intersected with up-regulated genes and genes most negative correlated with ccRCC intersected with down-regulated genes in both mRNA and lncRNA respectively to obtain ccRCC-specific mRNAs and lncRNAs.

# Construction Of A Dys-regulated ceRNA Network

Interactions between lncRNAs and miRNAs were identified by MiRcode(15)(16). The interactions between mRNAs and miRNAs were explored by StarBase(version 3.9)(17)(18), TargetScan(version 7.2)(19) and miRTarBase (version 8.0)(20) databases and we used Cytoscape(version 3.7.1) to depict the ceRNA network.

# Establishment Of Prognostic Gene Signature

To construct a risk assessment gene signature, a whole of 344 TCGA cases whose follow-up > 30 days with all clinical characteristics were randomly divided into a discovery group and a validation group, which were used to identify model and validate the efficacy of the model respectively. Both mRNA and lncRNA in the ceRNA network were applied to univariable Cox proportional-hazards model. Then, we selected gene meeting the statistical significance (P value < 0.01) to conduct Lasso penalized Cox regression analysis, thus finding a prognostic gene signature for patients with ccRCC on the basis of the lasso penalized Cox regression model coefficients (β) multiplied by gene expression levels normalized and transformed by DESeq2. To assess the prognostic gene signature, we conducted a time-dependent receiver operating characteristic (ROC) curve analysis and calculated Harrell's concordance index (C-index) among discovery group, validation group and the entire group separately.

# Construction and Validation of a Prognostic Nomogram

A prognostic nomogram predicting 1-, 3- and 5- year survival probability for ccRCC patients in the entire group was constructed by applying prognostic gene signature and relevant clinical characteristics to a stepwise Cox proportional-hazards model. Of course, we tested the discrimination and the calibration of the nomogram by Harrell's concordance index (C-index) analysis and the calibration curves analysis of the nomogram.

# Functional Enrichment Analysis

we only applied functional enrichment analysis to mRNA as mRNA is the main functional molecule in ceRNA network. The clusterProfiler(21) package was employed to investigate both Gene Ontology(22) (23) (GO) which was used to describe gene functions along three aspects: biological process (BP), cellular component (CC) and molecular function (MF) and Kyoto Encyclopedia of Genes and Genomes(24) (KEGG) pathway which was performed for functional pathways.

## Results

# Data Source And Cancer-specific lncRNAs

A total of 354 samples with ccRCC (all from TCGA database) and 157 normal controlled samples (85 from GTEX and 72 from TCGA) were collected from TCGA and GTEX database including 19538 mRNA expression values and 13511 lncRNA expression values. The flow chart of the ceRNA network construction was depicted (Additional file 1). we investigated co-expression network of 8332 lncRNAs by WGCNA with softpower 7, minModuleSize 15 and mergeCutHeight 0.20 as the threshold after filtering lncRNAs with median absolute deviation (MAD) zero. Finally, we identified a total of 7 coexpression modules and found that brown module with 744 lncRNAs was most positively correlated with ccRCC and blue module including 1424 lncRNAs displayed highest negative relationship with ccRCC (Additional file 2: figure A). the differentially expressed lncRNAs were also analyzed by DESeq2(25). We identified 3654 up-regulated lncRNAs and 1223 down-regulated lncRNAs (Additional file 2: figure B). By interacting them respectively with lncRNAs in WGNCA module most related with ccRCC, we obtained 610 ccRCC-specific lncRNAs both positive correlated to ccRCC and up-regulated in ccRCC and 767 ccRCC-specific lncRNAs displaying both negative relation with ccRCC and down-regulation in ccRCC.

# mRNA Modules Correlated With ccRCC

WGCNA was employed to analyze gene modules among the top 10000 mRNAs with maximum median absolute deviation (MAD) with softpower 14, minModuleSize 25 and mergeCutHeight 0.20 as the threshold. Consequently, we identified 17 gene color modules and topological overlap matrix (TOM) was depicted in Fig. 1A, As shown in Fig. 1B, the association between gene co-expression modules and particular clinical trait ccRCC was explored, A total of 4005 mRNAs which showed the highest relationship with ccRCC was found in Brown module, black module and turquoise module. As biological function often is displayed in the gene modules network, we conducted GO analysis to reveal these mRNAs functions in BP, thus finding that these genes were most related to angiogenesis, extracellular matrix organization and response to hypoxia (Fig. 1C). In addition, genes were highly enriched in HIF-1 signaling pathway, MAPK signaling pathway and PI3K-Akt signaling pathway by KEGG analysis (Fig. 1D).

# Identification Of Differential Expression mRNAs (DEmRNAs)

DESeq2 was used to explore the differential expression mRNAs between 354 samples with ccRCC and 157 normal control samples, revealing that there were 3679 significantly up-regulated and 1944 significantly down-regulated mRNAs which was depicted in volcano map (Additional file 3: Figure A). we used Gene Set Enrichment Analysis (GSEA) to demonstrate the biological function behind these identified differential expression genes, including both GO analysis and KEGG analysis. DEmRNAs were enriched in neutrophil mediated immunity, immune response-activating signal transduction, and neutrophil activation in biological process (BP) (Additional file 3: Figure B and Figure C). Moreover, Cytokine-cytokine receptor interaction, Human T-cell leukemia virus 1 infection and Viral carcinogenesis related genes were up-regulated while Collecting duct acid secretion, Proximal tubule bicarbonate reclamation and Glyoxylate and dicarboxylate metabolism pathways were downregulated by KEGG-GSEA (Additional file 3: Figure D). The same with lncRNA, we obtained ccRCC-specific mRNAs by collapsing lncRNAs from WGNCA modules and differential expression analysis, consequently we obtained 2191 mRNAs.

# Establishment Of Dys-regulated ceRNA Network

On the basis of ccRCC-specific mRNAs and lncRNAs, we construct a dys-regulated ceRNA network of lncRNA–miRNA–mRNA in ccRCC. There were 2191 mRNAs and 1377 lncRNAs included in the construction of dys-regulated ceRNA network. Then the interaction between cancer-specific lncRNAs and miRNAs was identificated using miRcode, following StarBase, miRTarBase and TargetScan databases were applied to demonstrate the target miRNA of cancer-specific mRNAs. At the same time, miRNA sequencing data from TCGA was employed to review the top 10% expressed miRNA since the implementation of ceRNA function depends on abundant of miRNAs, we only included triple lncRNAs-miRNAs-mRNAs with miRNA in above miRNAs. Consequently, a dys-regulated ceRNA network for ccRCC was established using 7 mRNAs, 363 lncRNAs, and 3 miRNAs, ceRNA network was depicted by Cytoscape (Fig. 2).

# Construction of gene signature using genes in the ceRNA network

A total of 344 ccRCC patients with the expression levels of genes in the dys-regulated ceRNA network were included in the construction of gene signature. We randomly classified patients into two group, respectively a discovery group (n = 210) and a validation group (n = 134). Meanwhile, the Univariate Cox proportional hazards model was applied to both lncRNAs and mRNAs in the dys-regulated ceRNA network to screen for genes as biomarkers which significantly influence overall survival and prognosis in discovery group. Consequently, we obtained twenty-one genes including one mRNA and twenty lncRNAs with the threshold of P value < 0.01, subsequently these genes were applied into Lasso penalized Cox regression analysis with a lambda based on Cross-Validation using glmnet (26). Finally, we obtained eight genes including one mRNA, namely MPP5, and seven lncRNAs namely WT1-AS, AC114316.1, AC103719.1, AL162377.1, HS1BP3-IT1, LINC02657, AC015909.1. the ceRNA network of these genes was

depicted in a Sankey diagram (Fig. 3A) and the expression level of these genes associated with clinical characteristics was also depicted in a heatmap (Fig. 3B). Furthermore, we also depicted the expression boxplot (Fig. 3C). A risk score was constructed on the basis of lasso cox model coefficients and the gene expression levels normalized and transformed by DESeq2. We calculated the C-index for the prognostic model in discovery group (0.74; 95%CI, 0.68 ~ 0.81), a time-dependent receiver operating characteristic curve (ROC) analysis was also performed in discovery group among 1-, 3-, and 5- year which showed the area under ROC (AUC) was 0.786, 0.796, and 0.786 respectively (Fig. 4A). To further test the gene signature, we divided patients from discovery group into predicted low- and high- risk groups based on the median risk scores, following a establishment of Kaplan-Meier survival curve in both groups which revealed that patients with predicted high risk displayed shorter OS than those with low risk ($p < 0.0001$, Fig. 4D).

# Estimation And Validation Of Gene Signature

We depicted the risk scores, overall survival time, and gene expression data of genes in the gene signature in Fig. 5A-5C. The C-index in validation group and entire group are 0.75 (95%CI, 0.68 ~ 0.83) and 0.74 (95%CI, 0.69 ~ 0.79) respectively. To further validate above finding, ROC analysis was conducted in both validation group (AUC at 1-, 3-, 5-year: 0.89, 0.77, 0.797) and entire group (AUC at 1-, 3-, 5-year: 0.808, 0.78, 0.789) (Fig. 4B-4C). Meanwhile, we also employed the validation group and the entire group with a Kaplan-Meier survival curve which both verified that the high-risk group displayed a worse overall survival compared with low-risk group (Fig. 4E-4F). In conclusion, this finding further confirmed that the exclusive mRNA and the seven lncRNAs in the gene signature could influence the prognosis of patients with ccRCC which are vital biomarkers.

# Building and Validation of a Prognostic Nomogram

In order to identified prognostic factors related to the OS in ccRCC, the gene signature based on the ceRNA network with complete clinical information including gender, ages, TNM stage, AJCC stage (Table 1) were firstly applied to a univariate and multivariate Cox proportional hazards regression model. Univariate cox regression analysis presented that gene signature, ages, T stage, M stage, N stage, and AJCC stage were significantly related to the overall survival ($p < 0.05$), and multivariate cox analysis further revealed that ages, N stage and gene signature were independent risk factors. Meanwhile, we employed a stepwise Cox regression model to establish a nomogram predicting the 1-, 3-, and 5-year OS of ccRCC patients (Fig. 6A) with a C-index 0.81 (95%CI, 0.70 ~ 0.89), we also depicted the calibration curve for 1-, 3-, and 5-year survival probability (Fig. 6B) which collectively indicated a good accuracy of the prognostic nomogram on the basis of the eight-gene signature in predicting the OS of ccRCC patients.

Table 1
Relationship between clinicopathological characteristics and risk score calculated by using the 8-genes signature

| | Level | Low risk | High risk | p |
|---|---|---|---|---|
| N | | 172 | 172 | |
| Vital status (%) | Alive | 151 (87.8) | 93 (54.1) | < 0.001 |
| | Dead | 21 (12.2) | 79 (45.9) | |
| OS (mean (SD)) | | 1515.10 (1008.39) | 1308.04 (971.94) | 0.053 |
| gender (%) | female | 57 (33.1) | 68 (39.5) | 0.262 |
| | male | 115 (66.9) | 104 (60.5) | |
| ages (mean (SD)) | | 60.40 (12.48) | 60.88 (11.64) | 0.711 |
| T stage (%) | T1 | 114 (66.3) | 81 (47.1) | 0.005 |
| | T2 | 20 (11.6) | 29 (16.9) | |
| | T3 | 36 (20.9) | 58 (33.7) | |
| | T4 | 2 (1.2) | 4 (2.3) | |
| N stage (%) | N0&NX | 172 (100.0) | 168 (97.7) | 0.131 |
| | N1 | 0 (0.0) | 4 (2.3) | |
| M stage (%) | M0&MX | 158 (91.9) | 139 (80.8) | 0.005 |
| | M1 | 14 (8.1) | 33 (19.2) | |
| AJCC stage (%) | Stage I | 114 (66.3) | 79 (45.9) | 0.001 |
| | Stage II | 18 (10.5) | 21 (12.2) | |
| | Stage III | 25 (14.5) | 36 (20.9) | |
| | Stage IV | 15 (8.7) | 36 (20.9) | |
| Gene signature (mean (SD)) | | -0.21 (0.15) | 0.24 (0.25) | < 0.001 |

# Discussion

The incidence of kidney cancer among which renal cell carcinoma accounts for 90% worldwide occupies a portion of 2.4%, accounting to 338,000 new cases and 144,000 deaths each year in total(1). ccRCC is a complex tumor with different clinical and pathological features, genetic variation, DNA methylation profiles, and RNA and proteomic signatures, which are closely related to the prognosis of ccRCC patients(27). Nevertheless, TNM staging system, the most used risk assessment system for ccRCC

patients, failed to take these genomic variation of ccRCC in to consideration which makes it not perfect for accurately predicting prognosis of ccRCC patients(28).

The novel hypothesis of gene expression regulation has been confirmed that each RNA combining with the same MREs could interact with or compete each other, which could help to further identify the mechanism of different kind of diseases especially cancer. The disturbance of the equipoise of ceRNA network was of vital importance for tumorigenesis. In gallbladder cancer, the lncRNA PVT1 which was positively related to malignancies and worse overall survival time was up-regulated in gallbladder cells. PVT1 and HK2 act as a ceRNA of miR-143 which could regulate aerobic glucose metabolism in GBC cells, promoting cell proliferation and metastasis (29). PTAR acts as a ceRNA of miR-101 which promotes tumorigenicity and metastasis in vivo in Ovarian cancer(30). LncRNA DANCR functions as a ceRNA in osteosarcoma which could promote cell proliferation and metastasis(31). And MT1JP, a down-regulated lncRNA in gastric cancer, was associated with malignant tumor phenotypes and survival of gastric cancer. MT1JP which severs as a ceRNA regulating FBXW7 expression could influence the progression of gastric cancer(32). Thus, ceRNA network containing crucial biomarkers was of vital importance in tumorigenesis.

Importantly, lncRNA-miRNA-mRNA dys-regulated ceRNA network displayed vital role in predicting disease prognosis. For example, in pancreatic cancer, 11 lncRNAs have been found and validated to function good in predicting prognosis(33). In the study of Soft tissue sarcoma (STS), seven genes (LPP-AS2, MUC1, GAB2, hsa-let-7i-5p, hsa-let-7f-5p, hsa-miR-101-3p and hsa-miR-1226-3p) in a recurrent STS-specific ceRNA network associated with recurrence and survival was identified based on the TCGA database containing 259 primary sarcomas and 3 local recurrence samples(34). In Hepatocellular carcinoma (HCC), MCM3AP-AS1 functioned as a ceRNA of miR-194-5p was a novel oncogenic lncRNA, which indicated poor clinical outcomes in patients with HCC, MCM3AP-AS1 could be a potential prognostic biomarker and therapeutic target for HCC(35). In glioblastoma multiforme, lung cancer, ovarian cancer and prostate cancer, based on the networks, the target mRNAs are normally up-regulated by the sponge lncRNAs after being released from miRNA control, and only a fraction of sponge lncRNA-mRNA regulatory relationships and hub lncRNAs are shared by the four cancers. Moreover, most sponge lncRNA-mRNA regulatory relationships show a rewired mode between different cancers, suggesting that different cancers had varied ceRNA networks(36). Although there are many studies on ceRNA networks in numerous cancers; however, few of them are on ccRCC. In addition, the sample sizes have not been large enough and most of them only focus on only lncRNAs.

In this study, the trait correlated and differential expression mRNAs and lncRNAs in ccRCC were identified by WGCNA and DESeq2 including data from TCGA and GTEx, we combined the differential expression genes with trait correlated genes and obtained ccRCC-specific genes including 2191 mRNAs and 1377 lncRNAs, followed by a construction of ceRNA network including 363 lncRNAs, 3 miRNAs and 7 mRNAs by MiRcode, StarBase, miRTarBase and TargetScan databases. To promote explore the relationships with prognosis of these 370 genes (mRNA and lncRNA in the ceRNA network), a gene signature with eight genes, namely MPP5, WT1-AS, AC114316.1, AC103719.1, AL162377.1, HS1BP3-IT1, LINC02657,

AC015909.1, was constructed by univariate cox proportional hazard regression, lasso and multivariate cox proportional hazard regression analysis. we also ensued the discriminations and accuracy of the gene signature with C-index and time-dependent ROC curve which all suggesting that the eight genes in the model could act as biomarkers based on the patients' prognosis.

Among the eight genes in the gene signature, the exclusive mRNA MPP5 which is associated with the membrane-associated guanylate kinase family helping the construction of cell polarity had been validated to be related to maintenance of cell polarity, invasion, cell division in prostate cancer(37), meanwhile, disruption of apical protein MPP5 which could negatively regulate YAP/TAZ abundance and activity might promote enrichment of oncogenic YAP and TAZ in HCC(38). MPP5 whose loss is a hallmark of cancer is crucial for tissue organization corresponding to the down-regulated expression in ccRCC. Long noncoding RNA WT1-AS which functioned as a potential tumor suppressor is related to poor survival in cervical squamous cell carcinoma(39)(40)(41) and triple-negative breast cancer (TNBC)(42). For lncRNA HS1BP3-IT1, it may be a prognosis biomarker for cholangiocarcinoma(43), laryngeal cancer(44) respectively. Therefore, our prediction of the ceRNA network had great confirmation of previous studies, meanwhile, by conducting Kaplan-Meier survival curves analysis for our gene signature, we confirmed that the high-risk group displayed a worse overall survival compared with low-risk group and Time-dependent ROC curve of our gene signature displayed a good performance in all group.

Nomograms are widely used as prognostic tools in oncology and medicine. By including various prognosis-associated variables and generating survival probability nomogram can help clinicians make a better treatment decision.(45). In present study, by including gene signature based on the ceRNA network and other related clinical characteristics into a stepwise cox model, a concise nomogram for prognostic prediction of ccRCC patients based on the eight-gene signature, ages, N stage, and AJCC stage was established, of course, an accurate estimation of the nomogram is validated by C-index and calibration curve with bootstrap methods which suggested its perfect discriminations and calibrations.

This study identified a eight-gene signature on the basis of dys-regulated ceRNA network, which will help us better understand dys-regulated ceRNA network mediated ccRCC and suggest therapeutics to ccRCC. Consequently, a nomogram with a good accuracy (C-index 0.81; 95% CI 0.70 ~ 0.89) based on genes in the dys-regulated ceRNA network in ccRCC is developed to predict prognosis of ccRCC patients compared with Jiang's nomogram (C-index 0.79;95% CI 0.75–0.82)(46). Although few novel prognosis-prediction systems based on gene expression have been developed for ccRCC patients, each previous study mainly focused on lncRNA molecular where we don't know whether there exist a regulation function in ccRCC, such as Jiang's (46) and Zhang's (47) nomogram, it's necessary to develop a comprehensive analysis to prognostic prediction on the basis of ceRNA network.

However, there are still some limitations in our study. First, although the ceRNA network was constructed based on TCGA and GTEX cohort, the development of the prognostic nomogram only using a cohort got from TCGA, and there were not a external validated dataset. Second, we ignored some other potential clinical characteristics. Although these limitations, the nomogram based on a comprehensive dys-

regulated ceRNA network analysis in ccRCC can promote clinicians to make a better and accurate treatment decision.

In conclusion, a dys-regulated ceRNA network based on ccRCC-specific genes was constructed followed by a development of nomogram predicting 1-, 3- and 5-year OS of ccRCC patient. The nomogram included both some clinical characteristics and ccRCC-specific gene signatures, so it will help clinicians make a better and accurate treatment decision more.

# Conclusions

We constructed a nomogram predicting 1-, 3-, and 5- OS of ccRCC patients with a gene signature based on the ccRCC-specific dys-regulated ceRNA network and some other clinical factors. It could contribute to a better understanding of ccRCC tumorigenesis mechanism and guide clinicians to make a more accurate treatment decision.

# Abbreviations

ccRCC
clear-cell renal cell carcinoma
ceRNA
competing endogenous RNA
lncRNAs
long non-coding RNAs
mRNAs
messenger RNAs
miRNAs
microRNAs
MREs
miRNA response elements
WGCNA
weighted correlation network analysis
GTEx
Genotype-Tissue Expression
TCGA
The Cancer Genome Atlas
FFPE
Formalin-Fixed Paraffin-Embedded
C-index
concordance index
CI
concordance index

BP
biological process
CC
cellular component
MF
molecular function
KEGG
Kyoto Encyclopedia of Genes and Genomes
MAD
median absolute deviation
TOM
topological overlap matrix
GSEA
Gene Set Enrichment Analysis
ROC
receiver operating characteristic curve
AUC
area under receiver operating characteristic curve
HCC
Hepatocellular carcinoma

# Declarations

### Ethics approval and consent to participate

No further approval was required from the Ethics Committee as the data comes from the TCGA and GTEx database

### Consent for publication

No applicable

### Availability of data and materials

The data that support the findings of this study are available in GTEx and TCGA databases.

### Competing interests

The authors declare that they have no competing interests

### Funding

This work was supported by the National Natural Science Foundation of China (program no. 81572543) and the Science and Technology Support Program of Tianjin, China (Grant No. 17ZXMFSY00040).

### Authors' contributions

HW designed and supervised the experiments. YP and SW participated in sample collection, sample processing, clinical information collection. ZX, DH, NL, ZZ, YP, SW analyzed and interpreted data. LW and YZ polished the manuscript. All authors read and approved the final manuscript.

### Acknowledgements

Not applicable.

## 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. Int J Cancer. 2015 Mar 1;136(5):E359-386.
2. Ficarra V, Galfano A, Mancini M, Martignoni G, Artibani W. TNM staging system for renal-cell carcinoma: current status and future perspectives. Lancet Oncol. 2007 Jun;8(6):554–8.
3. ceRNA in cancer: possible functions and clinical implications. - PubMed - NCBI [Internet]. [cited 2019 Nov 20]. Available from: https://www.ncbi.nlm.nih.gov/pubmed/26358722/
4. Ala U, Karreth FA, Bosia C, Pagnani A, Taulli R, Léopold V, et al. Integrated transcriptional and competitive endogenous RNA networks are cross-regulated in permissive molecular environments. Proc Natl Acad Sci U S A. 2013 Apr 30;110(18):7154–9.
5. The emergence of lncRNAs in cancer biology. - PubMed - NCBI [Internet]. [cited 2019 Nov 20]. Available from: https://www.ncbi.nlm.nih.gov/pubmed/22096659/
6. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008 Dec 29;9(1):559.
7. GTEx Portal [Internet]. [cited 2020 Mar 27]. Available from: https://www.gtexportal.org/home/
8. GDC TCGA portal [Internet]. [cited 2020 Mar 27]. Available from: https://portal.gdc.cancer.gov/
9. TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. - PubMed - NCBI [Internet]. [cited 2020 Jan 17]. Available from: https://www.ncbi.nlm.nih.gov/pubmed/26704973
10. Mounir M, Lucchetta M, Silva TC, Olsen C, Bontempi G, Chen X, et al. New functionalities in the TCGAbiolinks package for the study and integration of cancer data from GDC and GTEx. Wang E, editor. PLOS Comput Biol. 2019 Mar 5;15(3):e1006701.
11. Zerbino DR, Achuthan P, Akanni W, Amode MR, Barrell D, Bhai J, et al. Ensembl 2018. Nucleic Acids Res. 2018 Jan 4;46(D1):D754–61.
12. 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 Res. 2015 Apr 20;43(7):e47–

e47.
13. R Core Team. R: A Language and Environment for Statistical Computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2020. Available from: https://www.R-project.org/
14. Langfelder P, Horvath S. Fast R Functions for Robust Correlations and Hierarchical Clustering. J Stat Softw. 2012 Mar;46(11).
15. miRcode - transcriptome-wide microRNA target prediction including lncRNAs [Internet]. [cited 2020 Mar 27]. Available from: http://www.mircode.org/
16. miRcode: a map of putative microRNA target sites in the long non-coding transcriptome | Bioinformatics | Oxford Academic [Internet]. [cited 2020 Mar 27]. Available from: https://academic.oup.com/bioinformatics/article/28/15/2062/238391
17. Gong J, Shao D, Xu K, Lu Z, Lu ZJ, Yang YT, et al. RISE: a database of RNA interactome from sequencing experiments. Nucleic Acids Res. 2018 04;46(D1):D194–201.
18. ENCORI: The Encyclopedia of RNA Interactomes. [Internet]. [cited 2020 Mar 27]. Available from: http://starbase.sysu.edu.cn/
19. TargetScanHuman 7.2 [Internet]. [cited 2020 Mar 27]. Available from: http://www.targetscan.org/vert_72/
20. miRTarBase: the experimentally validated microRNA-target interactions database [Internet]. [cited 2020 Mar 27]. Available from: http://mirtarbase.cuhk.edu.cn/php/download.php
21. Yu G, Wang L-G, Han Y, He Q-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS J Integr Biol. 2012;16(5):284–7.
22. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene Ontology: tool for the unification of biology. Nat Genet. 2000 May;25(1):25–9.
23. The Gene Ontology Resource: 20 years and still GOing strong. Nucleic Acids Res. 2019 Jan 8;47(D1):D330–8.
24. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000 Jan 1;28(1):27–30.
25. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014 Dec;15(12):550.
26. Simon N, Friedman J, Hastie T, Tibshirani R. Regularization Paths for Cox's Proportional Hazards Model via Coordinate Descent. J Stat Softw [Internet]. 2011 [cited 2019 Nov 10];39(5). Available from: http://www.jstatsoft.org/v39/i05/
27. COMPREHENSIVE MOLECULAR CHARACTERIZATION OF CLEAR CELL RENAL CELL CARCINOMA. Nature. 2013 Jul 4;499(7456):43–9.
28. Moch H, Artibani W, Delahunt B, Ficarra V, Knuechel R, Montorsi F, et al. Reassessing the Current UICC/AJCC TNM Staging for Renal Cell Carcinoma. Eur Urol. 2009 Oct;56(4):636–43.
29. Chen J, Yu Y, Li H, Hu Q, Chen X, He Y, et al. Long non-coding RNA PVT1 promotes tumor progression by regulating the miR-143/HK2 axis in gallbladder cancer. Mol Cancer. 2019 02;18(1):33.

30. Liang H, Yu T, Han Y, Jiang H, Wang C, You T, et al. LncRNA PTAR promotes EMT and invasion-metastasis in serous ovarian cancer by competitively binding miR-101-3p to regulate ZEB1 expression. Mol Cancer. 2018 11;17(1):119.
31. Wang Y, Zeng X, Wang N, Zhao W, Zhang X, Teng S, et al. Long noncoding RNA DANCR, working as a competitive endogenous RNA, promotes ROCK1-mediated proliferation and metastasis via decoying of miR-335-5p and miR-1972 in osteosarcoma. Mol Cancer. 2018 12;17(1):89.
32. Zhang G, Li S, Lu J, Ge Y, Wang Q, Ma G, et al. LncRNA MT1JP functions as a ceRNA in regulating FBXW7 through competitively binding to miR-92a-3p in gastric cancer. Mol Cancer. 2018 02;17(1):87.
33. Giulietti M, Righetti A, Principato G, Piva F. LncRNA co-expression network analysis reveals novel biomarkers for pancreatic cancer. Carcinogenesis. 2018 30;39(8):1016–25.
34. Huang R, Meng T, Chen R, Yan P, Zhang J, Hu P, et al. The construction and analysis of tumor-infiltrating immune cell and ceRNA networks in recurrent soft tissue sarcoma. Aging. 2019 Nov 18;11.
35. Wang Y, Yang L, Chen T, Liu X, Guo Y, Zhu Q, et al. A novel lncRNA MCM3AP-AS1 promotes the growth of hepatocellular carcinoma by targeting miR-194-5p/FOXA1 axis. Mol Cancer. 2019 19;18(1):28.
36. Zhang J, Liu L, Li J, Le TD. LncmiRSRN: identification and analysis of long non-coding RNA related miRNA sponge regulatory network in human cancer. Bioinforma Oxf Engl. 2018 15;34(24):4232–40.
37. Teles Alves I, Hartjes T, McClellan E, Hiltemann S, Böttcher R, Dits N, et al. Next-generation sequencing reveals novel rare fusion events with functional implication in prostate cancer. Oncogene. 2015 Jan;34(5):568–77.
38. Tóth M, Wan S, Weiler S, Schmitt J, Sticht C, Gretz N, et al. MPP5 (Membrane Palmitoylated Protein 5), a novel upstream regulator of yes-associated protein (YAP) and its paralogue TAZ. In: Zeitschrift für Gastroenterologie [Internet]. Georg Thieme Verlag KG; 2018 [cited 2019 Nov 19]. p. A4.61. Available from: http://www.thieme-connect.de/DOI/DOI?10.1055/s-0037-1612820
39. Zhang Y, Na R, Wang X. LncRNA WT1-AS up-regulates p53 to inhibit the proliferation of cervical squamous carcinoma cells. BMC Cancer. 2019 Nov 6;19(1):1052.
40. Dai S-G, Guo L-L, Xia X, Pan Y. Long non-coding RNA WT1-AS inhibits cell aggressiveness via miR-203a-5p/FOXN2 axis and is associated with prognosis in cervical cancer. Eur Rev Med Pharmacol Sci. 2019 Jan;23(2):486–95.
41. Cui L, Nai M, Zhang K, Li L, Li R. lncRNA WT1-AS inhibits the aggressiveness of cervical cancer cell via regulating p53 expression via sponging miR-330-5p. Cancer Manag Res. 2019;11:651–67.
42. Wang J, Xi C, Yang X, Lu X, Yu K, Zhang Y, et al. LncRNA WT1-AS Inhibits Triple-Negative Breast Cancer Cell Migration and Invasion by Downregulating Transforming Growth Factor β1. Cancer Biother Radiopharm. 2019 Oct 17;
43. Song W, Miao D, Chen L. Comprehensive analysis of long noncoding RNA-associated competing endogenous RNA network in cholangiocarcinoma. Biochem Biophys Res Commun. 2018 Dec 2;506(4):1004–12.

44. Zhang G, Fan E, Zhong Q, Feng G, Shuai Y, Wu M, et al. Identification and potential mechanisms of a 4-lncRNA signature that predicts prognosis in patients with laryngeal cancer. Hum Genomics. 2019 15;13(1):36.
45. Balachandran VP, Gonen M, Smith JJ, DeMatteo RP. Nomograms in oncology: more than meets the eye. Lancet Oncol. 2015 Apr;16(4):e173-180.
46. Jiang W, Guo Q, Wang C, Zhu Y. A nomogram based on 9-lncRNAs signature for improving prognostic prediction of clear cell renal cell carcinoma. Cancer Cell Int [Internet]. 2019 Aug 5 [cited 2019 Dec 29];19. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6683339/
47. Zhang J, Zhang X, Piao C, Bi J, Zhang Z, Li Z, et al. A long non-coding RNA signature to improve prognostic prediction in clear cell renal cell carcinoma. Biomed Pharmacother. 2019 Oct 1;118:109079.

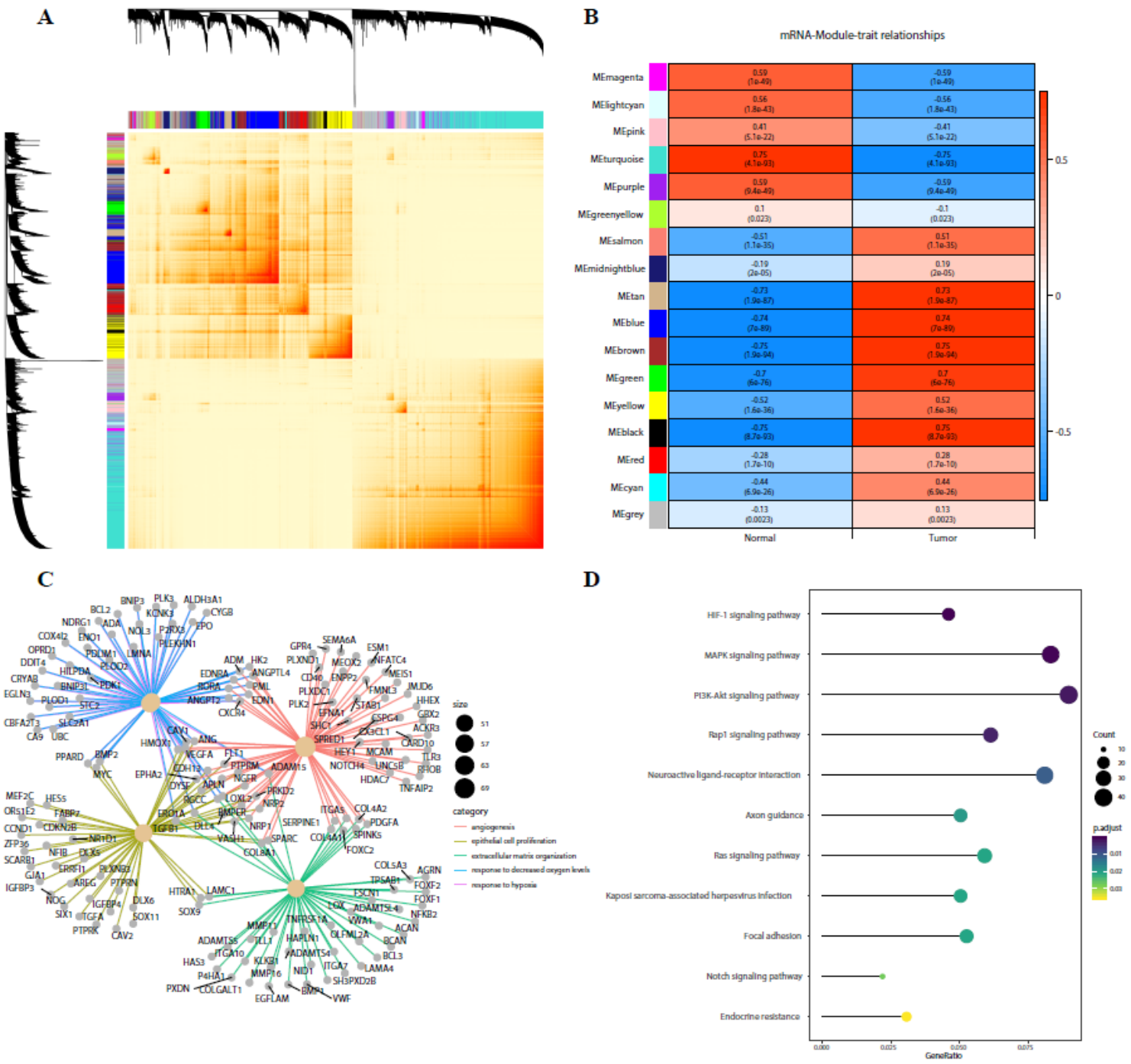
# Additional Files

**Additional file 1** (.pdf). The flow chart of the ceRNA network construction.

**Additional file 2** (.pdf). Identification of ccRCC-specific lncRNAs(A) The relationship of lncRNA in modules between ccRCC and normal samples was investigated. (B) the volcano of differential expression lncRNAs in ccRCC compared with normal controlled samples, red spots represent up-regulated genes, and blue spots represent down-regulated genes.

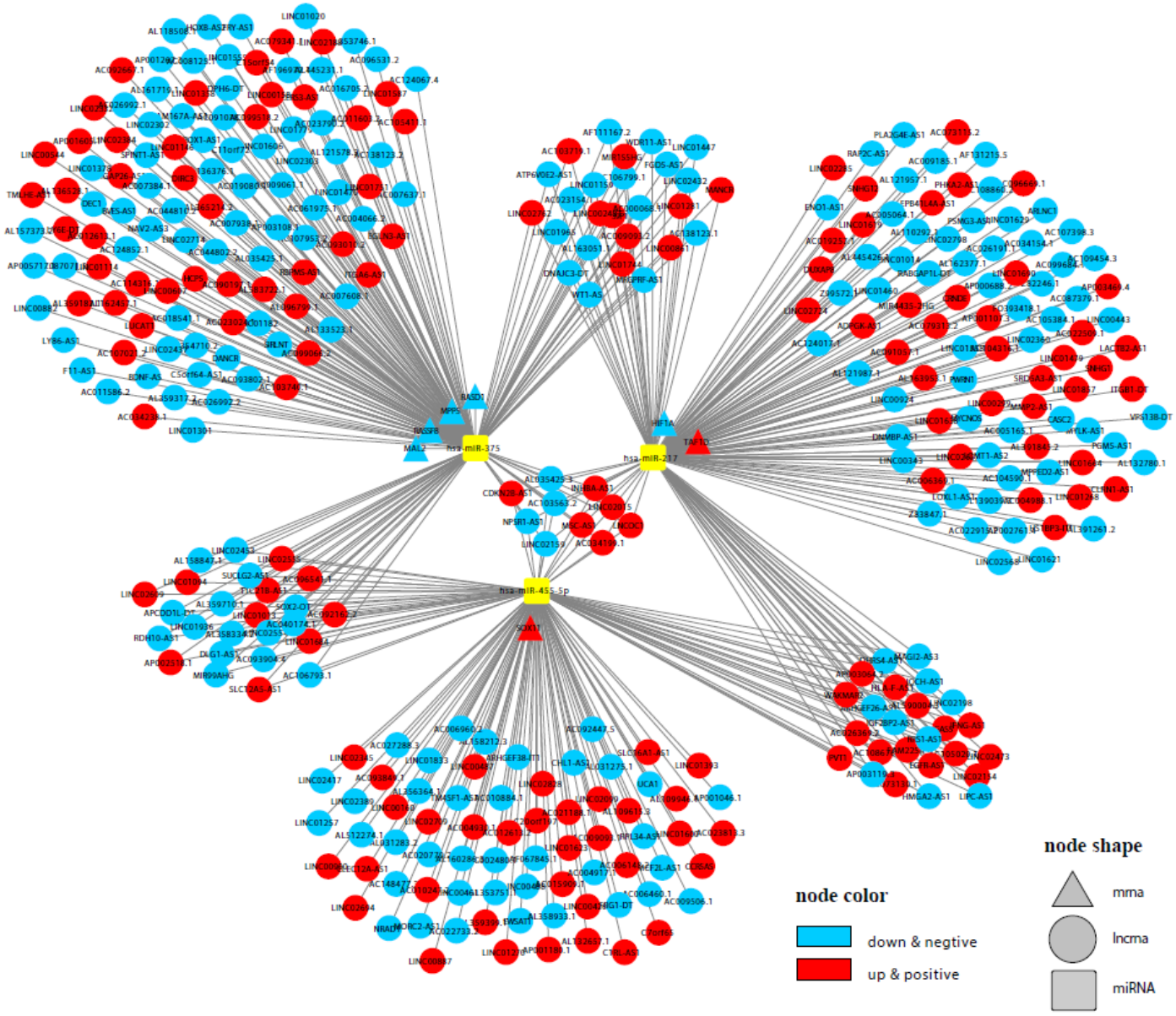
**Additional file 3**. (.pdf). Differential gene expression between ccRCC and normal controlled samples for mRNAs. (A) the volcano of differential expression mRNAs in ccRCC compared with normal controlled samples, red spots represent up-regulated genes, and blue spots represent down-regulated gens. (B) GSEA analysis was applied to differential expression genes for Go analysis in BP, CC and MF. (C) Gene symbols and interaction of the significantly mRNAs in BP were shown. (D) KEGG-GSEA was applied for signaling pathway analysis.

# Figures



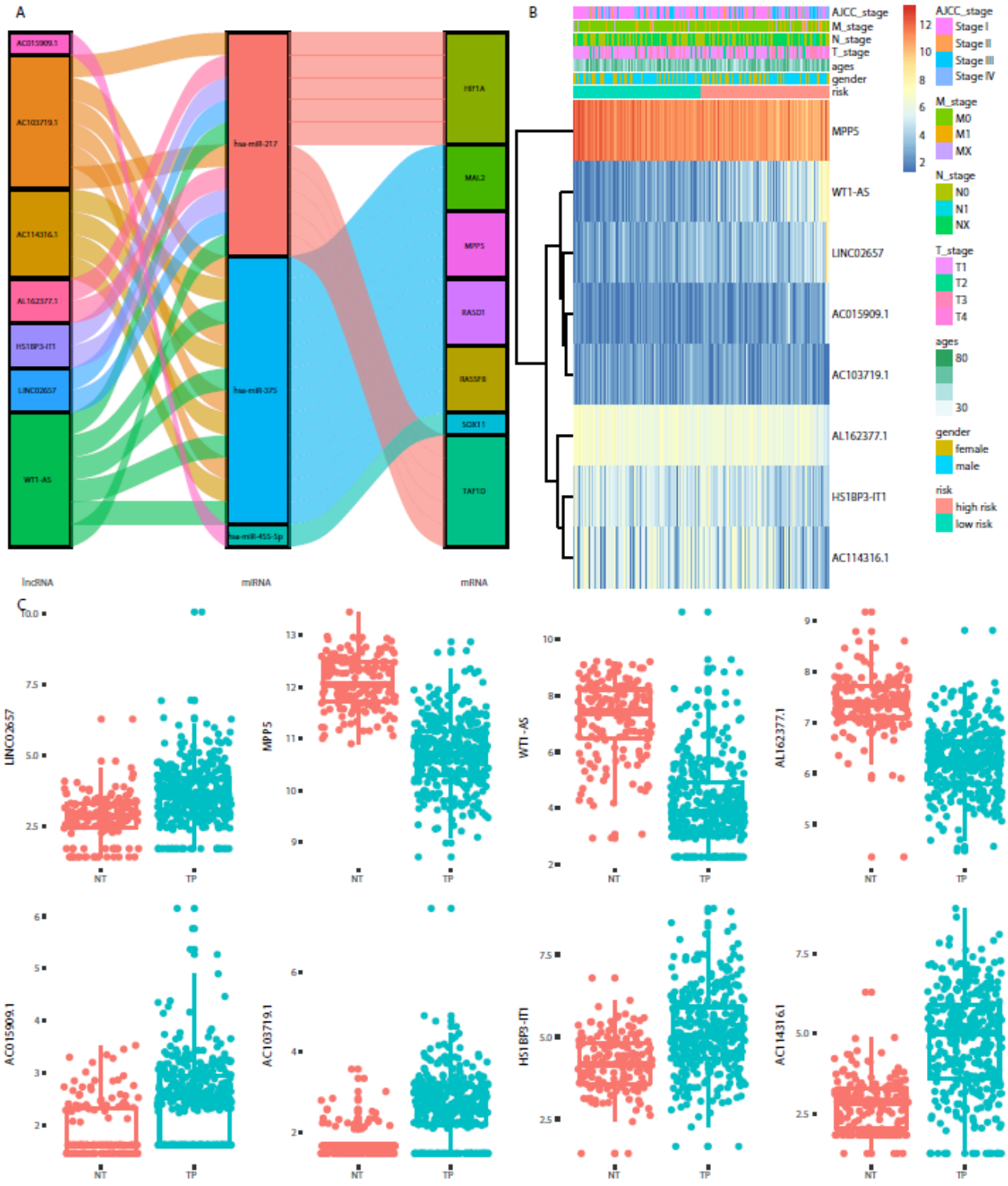

## Figure 1

WGCNA is applied to analyze mRNA module. (A) Heatmap plot of topological overlap in the gene network was shown. (B) The relationship of genes in modules between ccRCC and normal samples was investigated. (C) Gene symbols and interaction of the brown and black modules were shown. (D) KEGG analysis was used to investigate the pathway enrichment in brown and black modules.



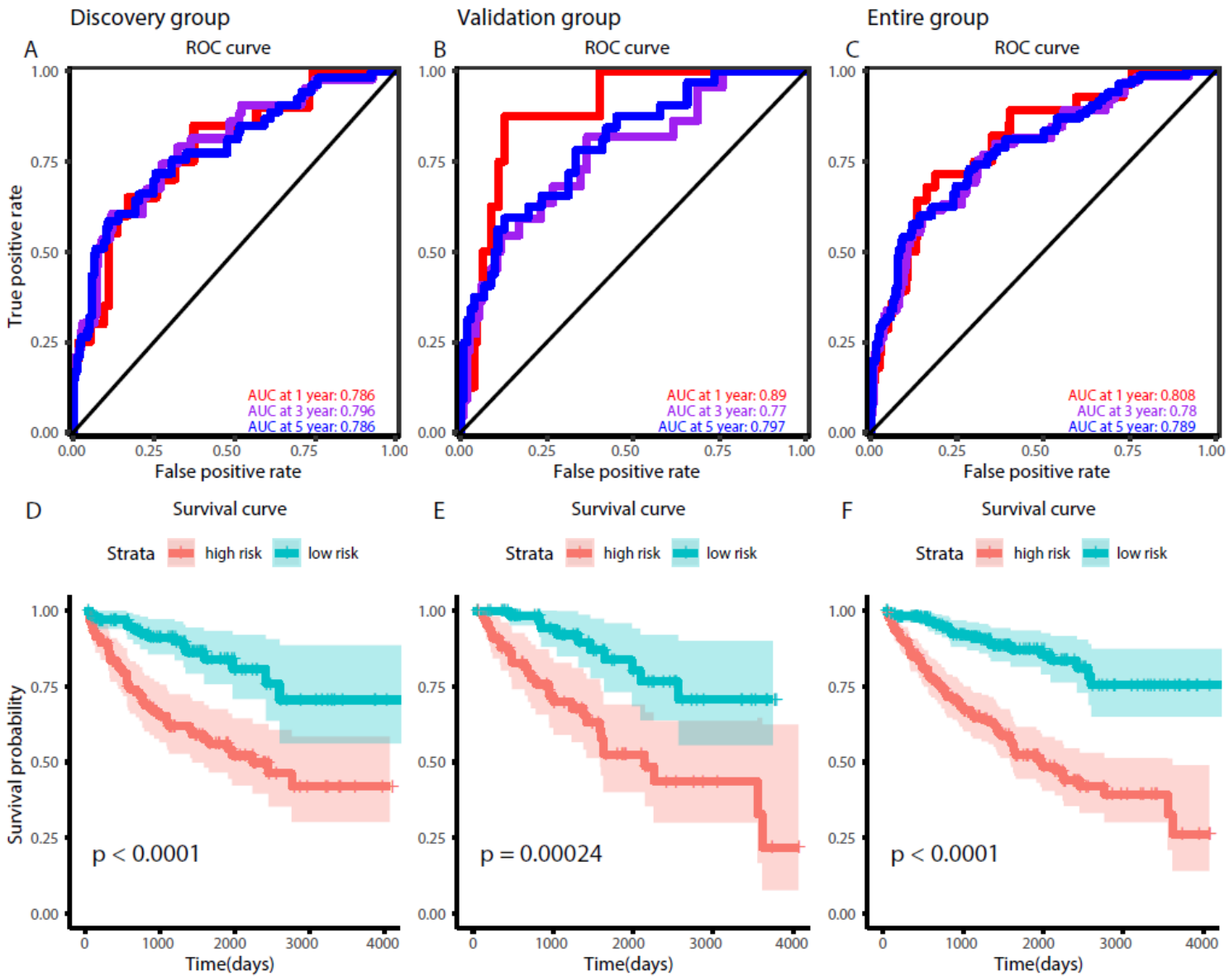

## Figure 2

A lncRNA-miRNA-mRNA ceRNA network was constructed by 363 lncRNAs, 3 miRNAs, and 7 mRNAs.



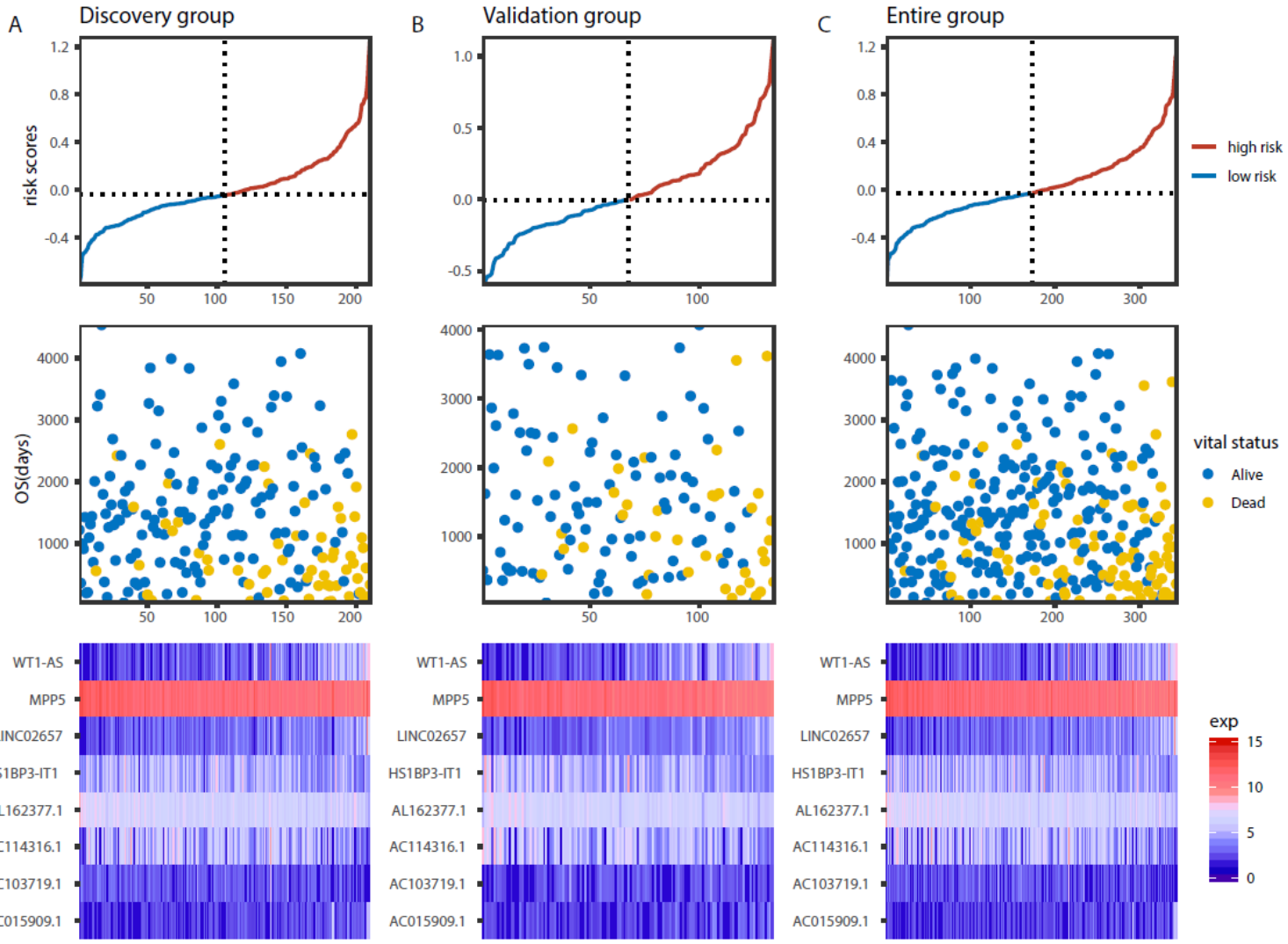

## Figure 3

A gene signature based on ccRCC-specific ceRNA network was developed. (A) The ceRNA network among gene in the gene signature is shown in a Sankey diagram. (B) the expression level of eight genes in gene signature associated with clinical characteristics was also depicted in a heatmap. (C) The expression boxplot of the eight genes in the gene signature.




## Figure 4

Estimation and validation of gene signature. Time-dependent ROC curve for 1-, 3-, and 5-year survival (A) discovery group (B) validation group (C) entire group; Kaplan–Meier survival curves showing overall survival outcomes for the high- and low-risk patients (D) discovery group (E) validation group (F) entire group.




## Figure 5

Analysis of gene signature risk score in ccRCC patients. (A) discovery group (B) validation group (C) entire group. Each panel consists of three rows: top row, the low- and high-score group for the gene signature in ccRCC patients; middle row, the survival status and duration of ccRCC cases; bottom row, heatmap showing the expression of the eight key genes. The color, from blue to red shows, low to high expression, respectively.

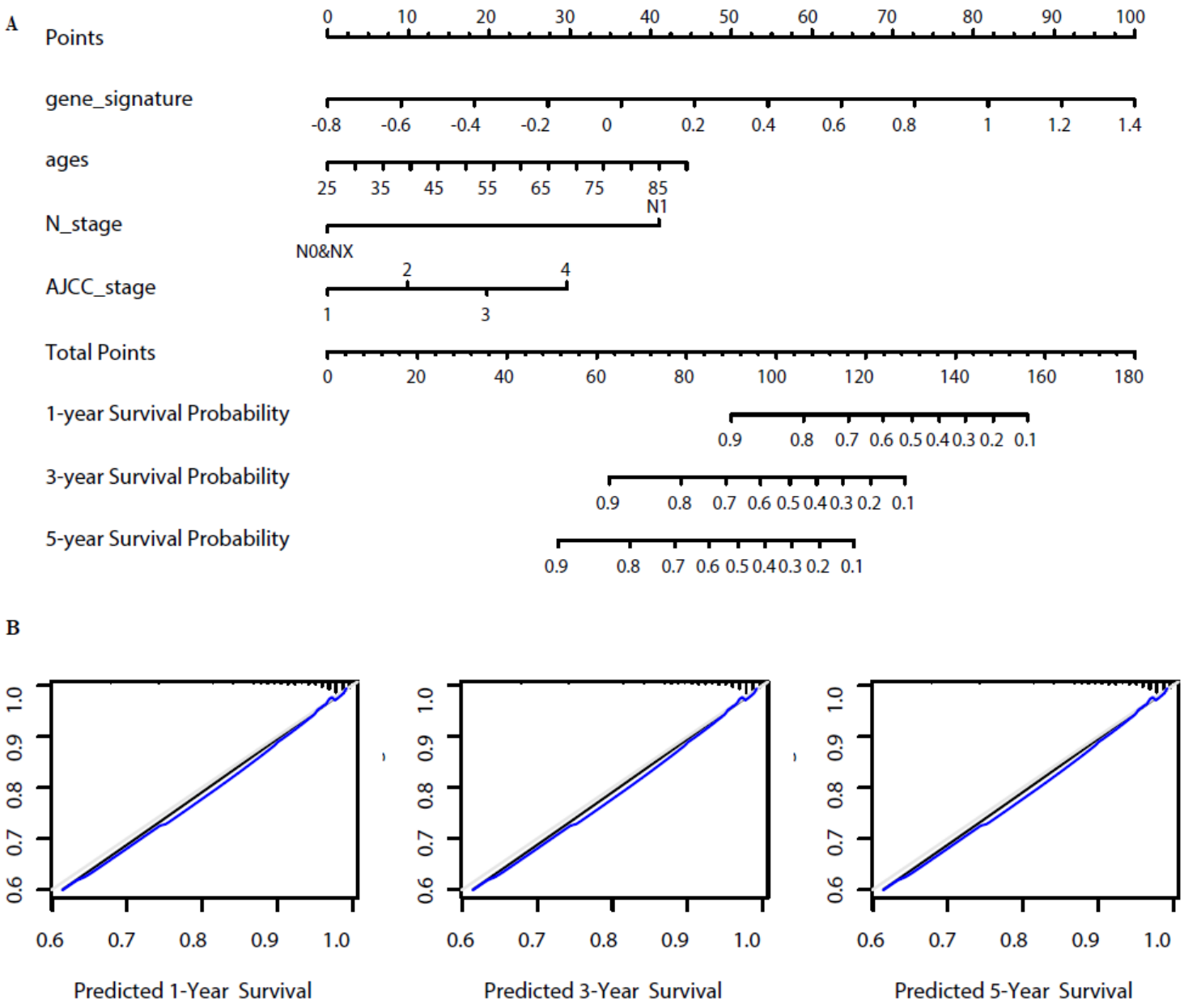



**Figure 6**

Building and Validation of a Prognostic Nomogram. (A) nomogram based on gene signatures, ages, N stage and AJCC stage for 1-, 3- and 5-year OS prediction. The number from 1 to 4 in AJCC stage represents AJCC stage I, II, III, and IV respectively (B) Calibration plot for agreement test between 1-, 3- and 5-year OS prediction and actual observation.

# Supplementary Files

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

- Additionalfile2.pdf
- Additionalfile3.pdf

- Additionalfile1.pdf