

# A novel prognostic predictor for clear cell renal cell carcinoma constructed by weighted gene co-expression correlation network analysis

Feng Li

Hebei Medical University Fourth Affiliated Hospital and Hebei Provincial Tumor Hospital

Yi Jin

Hebei Medical University Fourth Affiliated Hospital and Hebei Provincial Tumor Hospital

Peng Guo

Hebei Medical University Fourth Affiliated Hospital and Hebei Provincial Tumor Hospital

Zhiyu Wang (✉ [pmzq573@163.com](mailto:pmzq573@163.com))

Hebei Medical University Fourth Affiliated Hospital and Hebei Provincial Tumor Hospital

---

## Research

**Keywords:** ccRCC, overall survival, WGCNA, prognostic signature

**Posted Date:** May 13th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-24719/v1>

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

[Read Full License](#)

---

# Abstract

## Background

Clear cell renal cell carcinoma (ccRCC) accounts the largest proportion of all types of renal tumor, it is highly invasive leading to shorter survival time of patients suffer from tumor. This study aims at constructing a new prognostic signature to provide a new reference for clinic work to adjust treatment protocols.

## Material and Methods

Comprehensive gene expression profiles of ccRCC were downloaded from The Cancer Genome Atlas (TCGA). Weighted gene co-expression network analysis (WGCNA), a genome-wide profiling analysis, was applied to investigate key genes associated with clinical traits. Cox regression analysis was conducted to screen prognostic gene in ccRCC. LASSO analysis was used to enhance the robust of correlated genes. Then, a prognostic signature was developed as an independent factor to predict outcomes of ccRCC patients.

## Results

The co-expression network was constructed by WGCNA, 26 modules were identified in total, including a module (grey) showed no significance. By associating the features with clinic features, we selected three modules (pink, purple and turquoise) significantly associated with the overall survival(OS) of ccRCC patients. Univariate Cox analysis for genes from each module separately and then we used LASSO regression to screen the significant genes for selecting potential prognosis element-genes. 8 genes (CAPRN2, IFI44, LTV1, ZNF320, MTHFR, XPOT, BCL3 and PAX2) finally reserved by multivariate COX regression. Score<sub>W</sub> was defined as a prognosis mode established by the final potential prognosis 8 genes. Multivariate regression suggested the Score<sub>W</sub> was an independent prognosis predicting element. ROC curve method was performed and the AUC value of Score<sub>W</sub> showed satisfied and attractive clinical prognosis importance.

## Conclusion

Our findings provided the framework of co-expression gene modules of ccRCC and identified valuable prognostic method might be detection for ccRCC patients.

## Introduction

Renal cell carcinoma (RCC), derived from different types of renal cell, were classified to clear cell renal carcinoma, papillary carcinoma and chromophobe renal carcinoma [1, 2]. Clear cell renal cell carcinoma (ccRCC) is the main histologic subtype of renal cancer, accounts for approximately 70% of RCC cases [3, 4]. ccRCC can be invasive, nearly 30% patients present metastases when the first time of diagnosis, and quite a few, about 30%, patients develop metastases eventually [5, 6]. Metastatic ccRCC patients exhibit dismal 8% five-year overall survival (OS) for most ccRCCs being radio-chemo-resistant [5, 7]. Targeting therapy represents the best option to cancer treatment except for surgery [8, 9]. Because of the radio-chemo-resistant, the prognosis of the ccRCC can also be unpredictable. So novel molecular biomarkers and new prognosis predicting method are urgently needed.

WGCNA is defined as an effective method frequently applied to investigate the complicate relationships between phenotypes and gene expression profiles [10, 11]. The WGCNA can group gene expression data into a model or network, known as modules, similar behaving genes with a common biological relationship or function were included in one module [12]. WGCNA has been successfully used to study various biological processes and was demonstrated to be quite valuable for exploring potential biomarkers and therapeutic targets [13, 14]. Topological overlap measure (TOM) was used to be a proximity measure to establish network modules that combine the adjacency of two genes and the connection strengths with which these two genes interact with other neighbor genes [15, 16]. The correlation between module eigengenes, which are defined as principal components of the expression profiles, and all clinical traits is calculated, it helps in the process of comparing differentially expressed genes and visualizing the interactions between different co-expression modules.

In this study, we aimed to provide new insight of therapeutic targets and provide a better prognosis method for ccRCC patients. Expression data and clinical profiles of ccRCC samples were downloaded from TCGA, WGCNA was used to identify key gene has relationship with clinical traits such as age, gender, survival time, stage. Then, we select three modules which contain multiple genes that most correlate with clinical traits. Functional and pathway enrichment was used to analysis genes selected from WGCNA. LASSO regression, univariate and multivariable Cox regression were used to select the potential prognostic genes to establish a prognosis signature for ccRCC.

## **Material And Methods**

### **Data collection**

RNA sequence data and clinic information of ccRCC patients were collected from The Cancer Genome Atlas (TCGA) database (<https://cancergenome.nih.gov/>). We filter out duplicate samples and unanalyzable samples, 507 samples remained in total.

### **Construction of co-expression modules of ccRCC**

We used R package “WGCNA” to establish gene co-expression network in ccRCC. The soft-thresholding power  $\beta$  was calculated in the construction of each module using the pickSoftThreshold function of the WGCNA. Power value was screened using gradient algorithm to test the independence and the power value of different modules ranges from 1 to 20. After determining suitable power value when the index value for the reference dataset exceeded 0.8, the gene modules were constructed. The minimum number at 40 was set for each module, and the heatmap tool package was used to analyze and visualized the strength of correlation between each module. In order to generate dendrogram plot, a cut-line (0.25) was chosen.

### **Construction of module-clinic trait relationships**

Modules from WGCNA were identified according to gene expression similarities in samples. In order to acquire the module we are interested in, the significance between clinical traits (survival time, status, age, gender, historical stage) and each module was calculated. The gene modules significantly correlated with “fuptime” namely survival time in TCGA dataset were retained for next step.

### **Functional enrichment analysis of co-expression modules**

The intra-modular connectivity indicates the gene-fit of modules. The genes inside the same co-expression module have high connectivity. The genes within one module may have similar function. To identify the biological function of the significant modules, the corresponding gene annotation information was mapped to the database for Annotation, Visualization, and Integrated Discovery. The “ClusterProfiler” function of R was used to conduct Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis for each module.  $P < 0.05$  after correction was chosen as the thresholds. The R package “GOplot” were used to visualize the gene function identified by GO and KEGG analysis.

### **Univariate Cox analysis and LASSO analysis in certain modules.**

The genes in the module whose P-value between the gene module and “fuptime” was less than 0.01 were selected to receive univariate Cox analysis, which removes genes that have no significant correlation with OS. The results of univariate analysis was included in least absolute shrinkage and selection operator (LASSO) regression to eliminate the effects of collinearity. The LASSO scales the size of the parameter vector,  $\beta$ , and that unimportant variables (variables whose  $\beta$  coefficients are close to zero) are deleted from the model. The LASSO is usually applied to reduce the false discovery rate in high-dimensional Cox regression models.

### **Construction of the Prognostic Signature**

Multivariable Cox analysis was performed to develop a predictive model of OS, basing on the prognostic genes from LASSO regression analysis of certain modules. Combine the expression values of the prognostic genes weight by their regression coefficients to establish a prognosis predicting risk score for ccRCC patients as follows:

$$\text{Score } W = \sum_{i=1}^n \text{exp}_i * \beta_i$$

$n$  was the number of prognostic genes,  $\text{exp}_i$  was the expression of gene  $i$ ,  $\beta_i$  the regression coefficient of gene  $i$  in the multivariable analysis. Median score value was defined as a cutoff value, classifying ccRCC patients into low risk and high risk groups.

Combine the  $\text{Score}_W$  with clinic factors by using univariate and multivariable COX regression to identify whether it can be the independent factor to predict survival time.

## Statistical Analysis

$\text{Score}_W$  was included in multivariable Cox regression together with other clinic information to evaluate whether it can be an independent feature for survival prediction. The Kaplan-Meier curve was used to estimate the differences in survival time of low- and high-risk ccRCC patients. The R package "survivalROC" was used to estimate the prognostic signature. Statistical analysis was conducted by SPSS 25.0 (Chicago, IL, USA) and R 3.6.1 (<https://www.r-project.org/>). Independent sample t-tests were used to analysis correlation between clinical traits and predicting factors. All statistical significance was defined as a P-value less than 0.05 unless otherwise stated.

## Results

### Co-expression gene modules form WGCNA

R package of "WGCNA" was used to construct co-expression network. All genes were included in cluster analysis to generate gene modules. We chose the power of  $\beta = 6$  (scale free  $R^2 = 0.92$ ) as the soft thresholding to guarantee scale-free network (Fig.1A). The result of hierarchical clustering analysis is presented in Fig.1B. We merged some modules whose height is lower than 0.25, because these modules are not considered to be capable of acting as a standalone module. Finally, we identified 25 modules ranging from 845 genes (Turquoise) to 40 genes (White) in size (Fig.2A). Meanwhile, genes not classified into any one module were included in an insignificant module (Grey), and it not belongs to any other 25 modules. Interactions of the 25 co-expression modules were analyzed (Fig.2B), light color represents low overlap while darker red color indicates higher overlap.

### Gene co-expression modules associated with clinic traits

Download clinic traits dataset from TCGA database. Associations between clinic traits we were interested in and modules were identity based on correlation between modules eigengene and the clinic traits (Fig.2C). We investigated correlation coefficient and P value between each module and each clinical feature to select modules we are interested in. As we chose  $P < 0.01$  as a significant threshold, results showed that there are 3 modules (pink, purple, turquoise) were positively correlated to both survival time

and survival state at the same time. Furthermore, the dendrogram and heatmap was used to quantify groups of correlated eigengenes(Fig.3).

### **Functional enrichment analysis**

GO enrichment analysis was conducted in 3 positively correlated modules (Fig.4). Genes in pink module mainly enriched in regulation of mRNA metabolic process, RNA splicing, nuclear speck, spliceosomal complex, catalytic step 2 spliceosome and pre-mRNA binding. Genes in purple module mainly enriched in nuclear chromosome segregation, DNA replication, chromosome segregation, mitotic nuclear division, condensed chromosome, chromosomal region, chromosome, centromeric region and DNA-dependent ATPase activity. Genes in turquoise module mainly enriched in membrane docking, organelle localization by membrane tethering, RNA splicing, golgi vesicle transport, nuclear speck, SNARE complex, SNAP receptor activity and spliceosomal complex.

### **Univariate Cox proportional hazards regression and LASSO regression analysis**

For next step to establish a prognosis model to predict the outcome of ccRCC patients, the genes inside these 3 modules were included in univariate Cox proportional hazards regression. In order to improve the robustness and eliminate collinearity between genes as much as possible, we select LASSO regression method to process OS correlated genes data of three modules. When the partial likelihood deviance reached its minimum, remove the gene if its coefficient is 0. In pink module, the log (Lambda) reached its minimum of -3.45 (Fig.5A). In purple module, the log (Lambda) reached its minimum of -3.05 (Fig.5B). In turquoise module, the log (Lambda) reached its minimum of -3.175 (Fig.5C).

### **Construction of signature**

The application of LASSO reduces the dimension of multivariable Cox regression analysis to an acceptable level. We merged all the correlated genes of 3 modules, and put them into multivariable Cox regression analysis. The regression coefficients from the multivariable Cox analysis were used to construct  $Score_W$  and develop a prognostic model to predict OS. 8 genes (CAPRN2, coefficient =1.16E-05; IFI44, coefficient = 1.33E-06; LTV1, coefficient =3.10E-06; ZNF320, coefficient =-3.85E-06; MTHFR, coefficient =-5.75E-06; XPOT, coefficient =2.02E-06; BCL3, coefficient =8.62E-07; PAX2, coefficient =7.89E-07) were included in the signature. The information of 8 genes was showed in Table 1. The K-M survival curves of 8 genes were plotted in Fig.6, the high expression of 6 genes (CAPRN2, IFI44, LTV1, MTHFR, XPOT and BCL3) can significantly influence OS of ccRCC patients. We used the median score value as the cutoff value to classify patients into low or high risk. The  $Score_W$  distribution, gene expression, survival status of each patient was ranked by the  $Score_W$  values of prognostic signature (Fig.7A). K-M curve was plotted to indicate the correlation between  $Score_W$  and survival time of patients with ccRCC( $P<0.001$ , Fig.7B). The time-dependent ROC curve suggest a considerable effective performance of the signature for OS prediction ( $P<0.05$ ,  $AUC=0.796$ , Fig.7C). Correlation between  $Score_W$  and clinical

traits was showed in Fig.7D. Results indicated that higher Score<sub>W</sub> often appeared in the advanced cancer stage, higher Score<sub>W</sub> means the worse status of patients.

We bring Score<sub>W</sub> as one element which was similar to other clinic traits (age, gender, histological grade, and pathological stage) in univariable and multivariable Cox analysis. The results showed that predictive ability of Score<sub>W</sub> has significance in univariable and multivariable Cox analysis for OS of ccRCC patients (Table 2). The results of univariable analysis indicate Score<sub>W</sub> was significantly associated with OS [HR=1.089(1.060-1.118), P<0.001, Fig.7E]. And the results of multivariable analysis indicate that the Score<sub>W</sub> can be an independent predicting factor from other clinical elements [HR=1.054(1.032-1.076), P<0.001, Fig.7F]. The details of each element by time-dependent ROC analysis were presented in Fig.7G.

Only 6 genes (CAPRIN2, IFI44, LTV1, MTHFR, XPOT and BCL3) were significant with clinical pathological parameters (Fig.8). Results indicated CAPRIN2, IFI44, LTV1, XPOT and BCL3 were the risky genes for progression of ccRCC while MTHFR was a protective gene.

## Discussion

In recent decades, ccRCC accounts for about 3% of all malignant tumors and 85% of all primary kidney cancer, reaching the most proportion in all types of renal cancer<sup>[17, 18]</sup>. It is the third most common cancer in the urinary system, and it has the maximum fatality rate. ccRCC is highly invasive, nearly 60% patients present metastasis at the procedure of diagnosis and therapy<sup>[19]</sup>. Meanwhile, metastatic cancer is not sensitive to chemotherapy and radiotherapy<sup>[20]</sup>. So, 5-year survival rate of ccRCC becomes negative and is of great significance to the therapeutic effect of tumor. As we already know, histologic grade, pathological stage, local scope of the tumor, presence of regional nodal metastases showed significant effects on 5-year survival rate and occurrence of metastatic cancer<sup>[21]</sup>. However, these factors were not able to be detected early for ccRCC patients, resulting in 20–30% of the ccRCC patients were diagnosed late in advanced and metastasis cancer. Investigating potential molecular markers and construct an effective prognostic module is necessary for the diagnosis, treatment and prognosis of patients with ccRCC.

WGCNA has been widely used to investigate potential biomarkers of various diseases. Previous studies only focus on other clinic traits such as Clarke et al. associated gene module with histologic grade<sup>[22]</sup>, Tian et al. preferred the relationship of metastasis<sup>[23]</sup> and He et al. showed interest in pathological stage, few studies combine the module eigengene with survival time<sup>[24]</sup>. In this study, WGCNA was used to form gene co-expression network of ccRCC gene expression profiles downloaded from TCGA, and then calculated the stability of gene module. The relationship between stable modules and clinical features were assessed after gene modules were identified. Three modules (pink, purple and turquoise) were selected out for the significant correlation with survival time. The results of functional enrichment analysis indicated that the module pink was mainly enriched in the regulation of RNA metabolic process

and different way of RNA splicing. The module purple was mainly associated with chromosome and DNA replication. The module turquoise was more enrich in the function of membrane of organelle in cell.

Previous studies have identified many molecular signatures that classify patients into different prognostic groups. However, those studies were lack of stability. In this study, LASSO regression was used to enhance the robustness of prognosis model. Of note, the result of the AUC analyses in our study showed that the AUC values of the combination of 8 genes were more 0.796 in the OS, suggesting that the  $Score_W$  could be regarded as a novel factor that may serve as a prognosis indicator for ccRCC patients

All the elements in our model of potential prognosis 8 genes (CAPRIN2, IFI44, LTV1, ZNF320, MTHFR, XPOT, BCL3 and PAX2) were widely explored in previous literature. Jia et al. said that CAPRIN2 activate the Wnt pathway to influence cancer cell growth [25]. Huang et al. demonstrated that IFI44 is a novel tumor suppressor, and affect cancer stemness, metastasis, and drug resistance via regulating met/Src signaling pathway [26]. Meanwhile, Puig-Butille et al. showed the over-expression of IFI44 influence the procession of cancers [27]. Ghalei et al. showed LTV1 is necessary for ribosome assembly and cell growth [28]. In Yu's study, ZNF was identified as a key gene in microsatellite instability in certain cancer [29]. Liew et al. focuses on the recent evidence-based reports on the associations of the MTHFR C677T polymorphism and the various diseases globally, showed that MTHFR is associated with various diseases with the epidemiology [30]. Gonzales et al. present the most comprehensive analyses to date of MTHFR polymorphism-mutations and breast cancer risk [31]. Lin et al. demonstrate that XPOT Down-regulation of XPOT significantly inhibited tumor proliferation and invasion in vitro and vivo, and XPOT may affect tumor progression through cell cycle and ubiquitin-mediated proteolysis [32]. Wang et al. links Akt and MAPK pathways to NF- $\kappa$ B through Bcl3 and provides mechanistic insight into how Bcl3 functions as an oncoprotein through collaboration with IKK1/2, Akt, and Erk2 [33]. Boudjadi et al. investigated that PAX3 contributes to the pathogenesis of the soft tissue cancers alveolar rhabdomyosarcoma and biphenotypic sinonasal sarcoma [34]. However, CAPRIN2, IFI44, LTV1, ZNF320, XPOT, are rarely explored, therefore, we believe these genes mentioned in the context have more experimental significance in further studies.

In conclusion, by combining molecular signature and clinical characteristics, we constructed a novel risk score model,  $Score_W$ , for ccRCC patients' survival. The risk score showed significantly correlation with survival time and acts as an independent predicting factor. We provided the framework of co-expression gene modules of ccRCC and identified valuable prognostic method might be detection for ccRCC patients. The  $Score_W$  can be used as a new prognostic method which may conduce to treatment decisions for individuals. However, further verification in a large clinic samples are needed and that is what we plan to do next step and hold promise for clinical practice in the near future.

# Declarations

## Ethics approval and consent to participate

The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. This study complied with the Declaration of Helsinki and was approved by Ethics Committees of The Fourth Hospital of Hebei Medical University.

## Conflicts of interest

All the authors declare that they have no competing interests.

## Consent for publication

The authors have no interests ethical, legal and financial conflicts related to the article. All authors read and approved the manuscript to publish.

## Authors' contributions

- (I) Conception and design: Zhiyu Wang and Feng Li.
- (II) Administrative support: Zhiyu Wang.
- (III) Collection and assembly of data: Feng Li and Yi Jin.
- (IV) Data analysis and interpretation: Yi Jin.
- (V) Manuscript writing: Feng Li and Peng Guo.
- (VI) Final approval of manuscript: All authors.

## Funding

There is no Funding.

## Acknowledgements

Not available.

# References

1. Verhagen P, Boeve ER. The European Association of Urology Guideline on Renal Cell Carcinoma (RCC) is Not Concise in its Recommendation to Perform Partial Nephrectomy in T1b RCC. *Eur Urol.* 2019;76(1):132–3.
2. Allison SJ. Kidney cancer: Targeting metabolism in RCC. *Nat Rev Nephrol.* 2017;13(6):320.

3. Ricketts CJ, Linehan WM. The origin, evolution and route to metastasis of clear cell RCC. *Nat Rev Nephrol.* 2018;14(9):538–40.
4. Wang M. A central anti-oncogenic pathway in ccRCC. *Nat Rev Nephrol.* 2019;15(1):2.
5. Wang Y, Chen L, Wang G, et al. Fifteen hub genes associated with progression and prognosis of clear cell renal cell carcinoma identified by coexpression analysis. *J Cell Physiol.* 2019;234(7):10225–37.
6. Yang H, Li W, Lv Y, et al. Exploring the mechanism of clear cell renal cell carcinoma metastasis and key genes based on multi-tool joint analysis. *Gene.* 2019;720:144103.
7. Yuan L, Chen L, Qian K, et al. Co-expression network analysis identified six hub genes in association with progression and prognosis in human clear cell renal cell carcinoma (ccRCC). *Genom Data.* 2017;14:132–40.
8. Fu Q, Xu L, Wang Y, et al. Tumor-associated Macrophage-derived Interleukin-23 Interlinks Kidney Cancer Glutamine Addiction with Immune Evasion. *Eur Urol.* 2019;75(5):752–63.
9. Yao X, Tan J, Lim KJ, et al. VHL Deficiency Drives Enhancer Activation of Oncogenes in Clear Cell Renal Cell Carcinoma. *Cancer Discov.* 2017;7(11):1284–305.
10. Qu L, Wang ZL, Chen Q, et al. Prognostic Value of a Long Non-coding RNA Signature in Localized Clear Cell Renal Cell Carcinoma. *Eur Urol.* 2018;74(6):756–63.
11. Abu-Jamous B, Kelly S. 0000-0001-8583-5362 AO. Clust: automatic extraction of optimal co-expressed gene clusters from gene expression data. *Genome Biol.* 2018;19(1):172.
12. Sezin T, Vorobyev A, Sadik CD, Zillikens D, Gupta Y, Ludwig RJ. Gene Expression Analysis Reveals Novel Shared Gene Signatures and Candidate Molecular Mechanisms between Pemphigus and Systemic Lupus Erythematosus in CD4(+) T Cells. *Front Immunol.* 2017;8:1992.
13. Filteau M, Pavey SA, St-Cyr J, Bernatchez L. Gene coexpression networks reveal key drivers of phenotypic divergence in lake whitefish. *Mol Biol Evol.* 2013;30(6):1384–96.
14. Zheng X, Xue C, Luo G, Hu Y, Luo W, Sun X. Identification of crucial genes in intracranial aneurysm based on weighted gene coexpression network analysis. *Cancer Gene Ther.* 2015;22(5):238–45.
15. Fatima A, Connaughton RM, Weiser A, et al. Weighted Gene Co-Expression Network Analysis Identifies Gender Specific Modules and Hub Genes Related to Metabolism and Inflammation in Response to an Acute Lipid Challenge. *Mol Nutr Food Res.* 2018. 62(2).
16. Radulescu E, Jaffe AE, Oho AUID-, et al. Identification and prioritization of gene sets associated with schizophrenia risk by co-expression network analysis in human brain. *Mol Psychiatry.* 2018.
17. Feng X, Yan N, Sun W, et al. miR-4521-FAM129A axial regulation on ccRCC progression through TIMP-1/MMP2/MMP9 and MDM2/p53/Bcl2/Bax pathways. *Cell Death Discov.* 2019;5:89.
18. Wang Y, Guo X, Bray MJ, Ding Z, Zhao Z. An integrative genomics approach for identifying novel functional consequences of PBRM1 truncated mutations in clear cell renal cell carcinoma (ccRCC). *BMC Genom.* 2016;17 Suppl 7:515.
19. Zhu WK, Xu WH, Oho AUID-, et al. Decreased SPTLC1 expression predicts worse outcomes in ccRCC patients. *J Cell Biochem.* 2019.

20. Jingushi K, Kashiwagi Y, Ueda Y, et al. High miR-122 expression promotes malignant phenotypes in ccRCC by targeting occludin. *Int J Oncol*. 2017;51(1):289–97.
21. Ficarra V, Brunelli M, Novara G, et al. Accuracy of on-bench biopsies in the evaluation of the histological subtype, grade, and necrosis of renal tumours. *Pathology*. 2011;43(2):149–55.
22. Clarke C, Madden SF, Doolan P, et al. Correlating transcriptional networks to breast cancer survival: a large-scale coexpression analysis. *Carcinogenesis*. 2013;34(10):2300–8.
23. Tian H, Guan D, Li J. Identifying osteosarcoma metastasis associated genes by weighted gene co-expression network analysis (WGCNA). *Medicine (Baltimore)*. 2018. 97(24): e10781.
24. He YH, Chen C, Shi Z. 0000-0001-6486-3663 AO. The biological roles and clinical implications of microRNAs in clear cell renal cell carcinoma. *J Cell Physiol*. 2018;233(6):4458–65.
25. Jia D, Dong R, Jing Y, et al. Exome sequencing of hepatoblastoma reveals novel mutations and cancer genes in the Wnt pathway and ubiquitin ligase complex. *Hepatology*. 2014;60(5):1686–96.
26. Huang WC, Tung SL, Chen YL, Chen PM, Chu PY. IFI44L is a novel tumor suppressor in human hepatocellular carcinoma affecting cancer stemness, metastasis, and drug resistance via regulating met/Src signaling pathway. *BMC Cancer*. 2018;18(1):609.
27. Puig-Butille JA, Escamez MJ, Garcia-Garcia F, et al. Capturing the biological impact of CDKN2A and MC1R genes as an early predisposing event in melanoma and non melanoma skin cancer. *Oncotarget*. 2014;5(6):1439–51.
28. Ghalei H, Schaub FX, Doherty JR, et al. Hrr25/CK1delta-directed release of Ltv1 from pre-40S ribosomes is necessary for ribosome assembly and cell growth. *J Cell Biol*. 2015;208(6):745–59.
29. Yu C, Hong H, Zhang S, et al. Identification of key genes and pathways involved in microsatellite instability in colorectal cancer. *Mol Med Rep*. 2019;19(3):2065–76.
30. Liew SC, Gupta ED. Methylenetetrahydrofolate reductase (MTHFR) C677T polymorphism: epidemiology, metabolism and the associated diseases. *Eur J Med Genet*. 2015;58(1):1–10.
31. Gonzales MC, Yu P, Shiao SP. MTHFR Gene Polymorphism-Mutations and Air Pollution as Risk Factors for Breast Cancer: A Metaprediction Study. *Nurs Res*. 2017;66(2):152–63.
32. Lin J, Hou Y, 0000-0003-2569-1680 AO, et al. Exportin-T promotes tumor proliferation and invasion in hepatocellular carcinoma. *Mol Carcinog*. 2019. 58(2): 293–304.
33. Wang VY, Li Y, Kim D, et al. Bcl3 Phosphorylation by Akt, Erk2, and IKK Is Required for Its Transcriptional Activity. *Mol Cell*. 2017;67(3):484–97.e5.
34. Boudjadi S, Chatterjee B, Sun W, Vemu P, Barr FG. The expression and function of PAX3 in development and disease. *Gene*. 2018;666:145–57.

## Tables

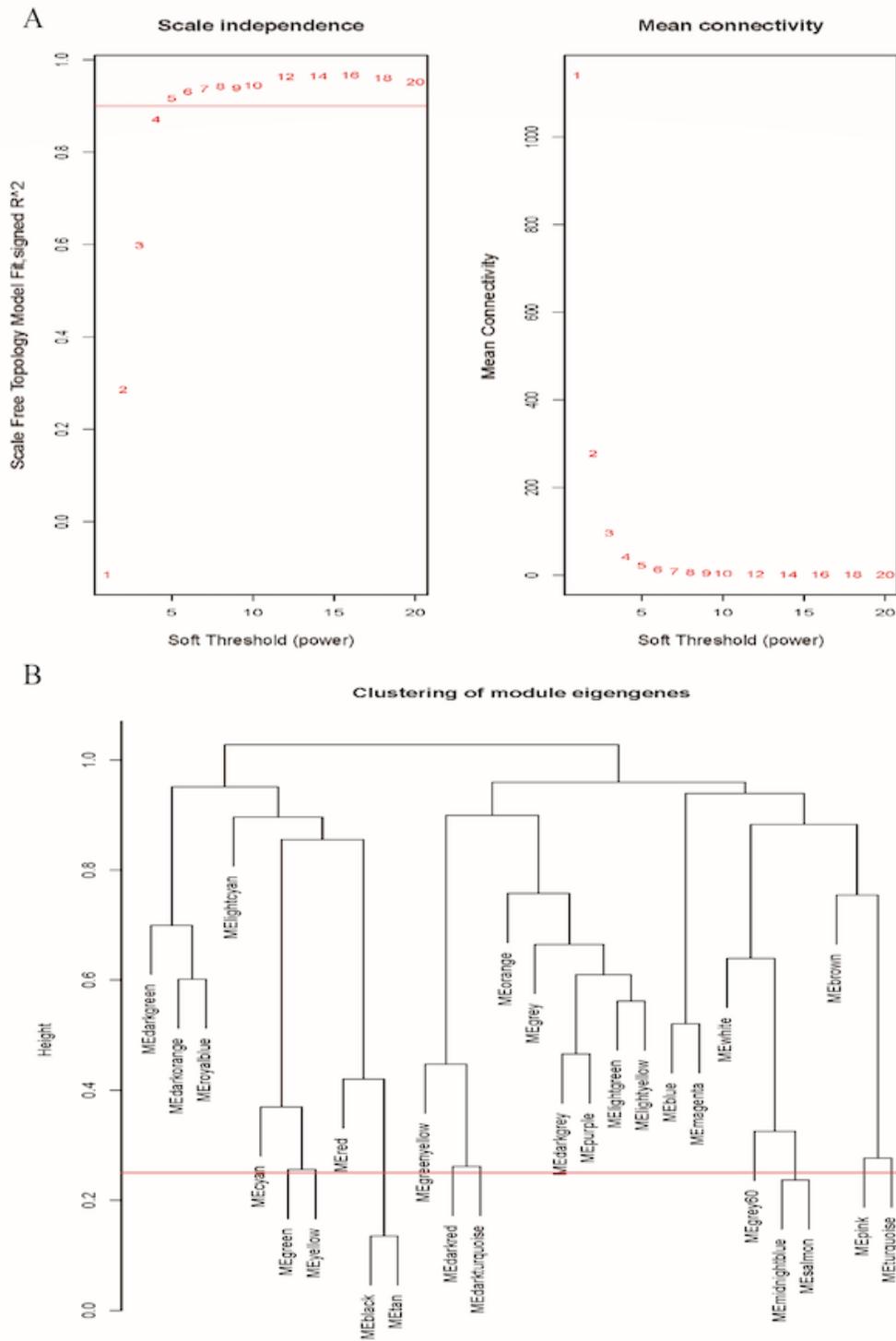
Table1 General information of the 8 genes for constructing the prognostic signature

Gene stable ID	Gene name	Gene type	Chromosome	Gene start (bp)	Gene end (bp)
ENSG00000110888	CAPRIN2	protein coding	12	30709552	30754993
ENSG00000137965	IFI44	protein coding	1	78648943	78664078
ENSG00000135521	LTV1	protein coding	6	143843273	143863812
ENSG00000204604	ZNF320	protein coding	19	52859491	52897727
ENSG00000177000	MTHFR	protein coding	1	11785723	11806103
ENSG00000184575	XPOT	protein coding	12	64404392	64451125
ENSG00000069399	BCL3	protein coding	19	44742620	44760044
ENSG00000075891	PAX2	protein coding	10	100735396	100829944

Table2 Univariate and multivariate Cox regression analyses of the gene signature and overall survival of ccRCC patients.

		N	Univariate analysis			Multivariate analysis		
			HR	95% CI	P	HR	95% CI	P
age	<=60	260	1		<0.001	1		0.045
	>60	247	1.694	1.236-2.322		1.384	1.008-1.906	
gender	female	174	1		0.787	1		0.543
	male	333	0.957	0.694-1.318		1.106	0.799-1.532	
stage	I/II	307	1		<0.001	1		<0.001
	III/IV	200	1.637	1.495-1.792		3.522	2.510-4.941	
Risk score	low	254	1		<0.001	1		<0.001
	high	253	1.089	1.060-1.118		2.911	2.017-4.202	

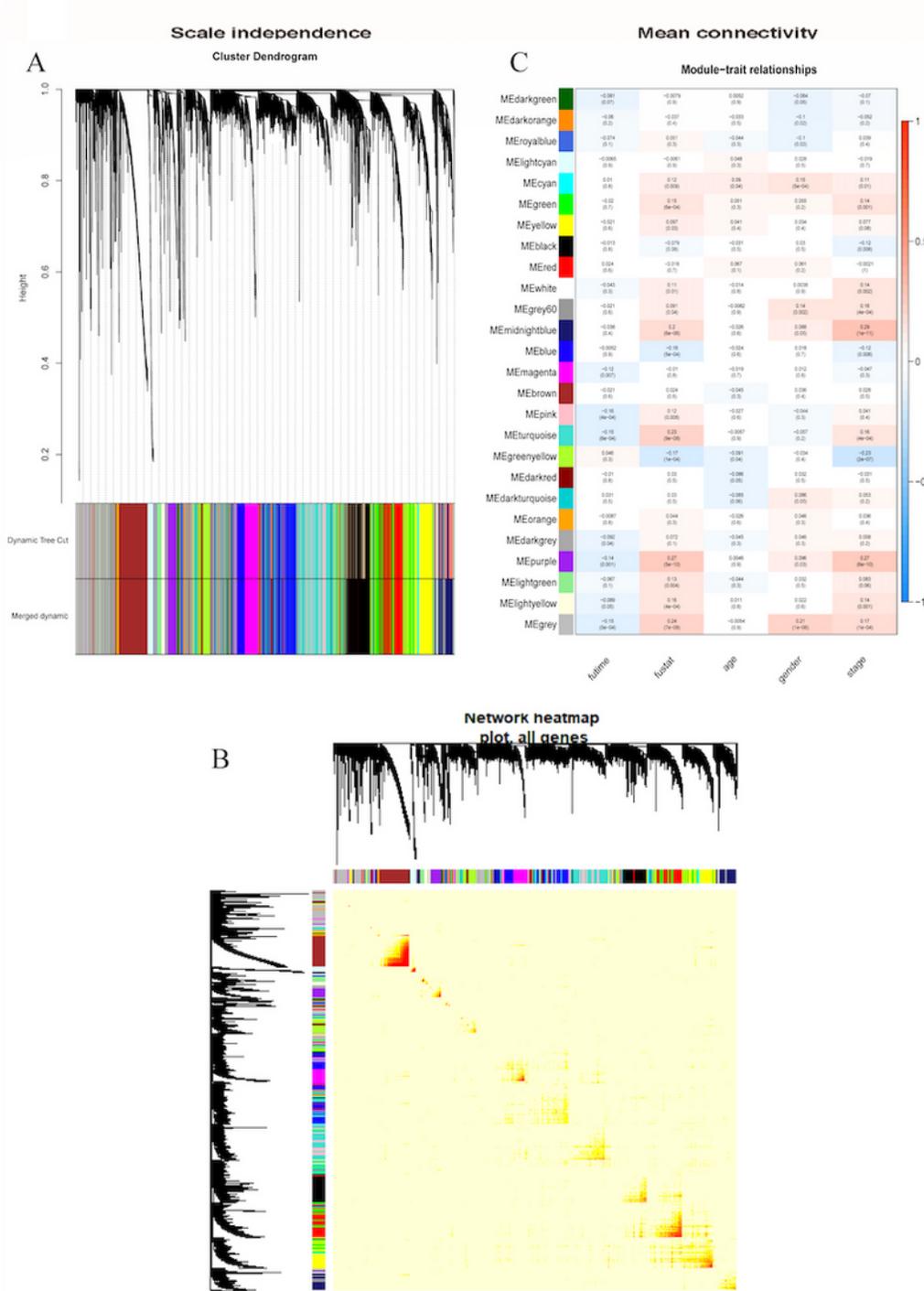
## Figures



**Figure 1**

Network topology for different soft-thresholding powers (A) Analysis of the scale-free fit index for various soft-thresholding powers ( $\beta$ ) and the mean connectivity for various soft-thresholding powers. The approximate scale-free topology can be attained at the soft-thresholding power of 6. (B) Dendrogram of consensus module eigengenes obtained by WGCNA on the consensus correlation. The red line is the

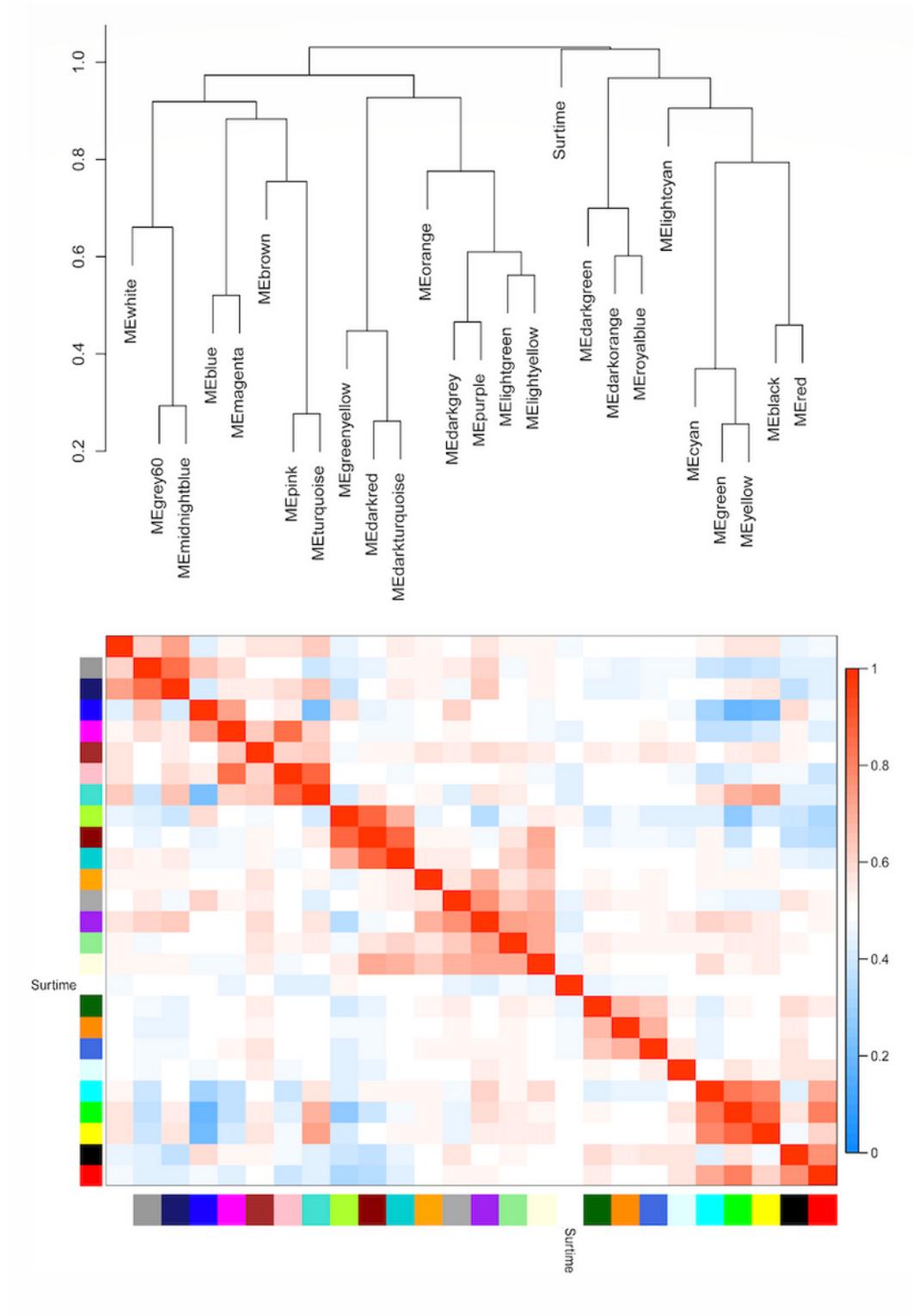
merging threshold, and groups of eigengenes below the threshold represent modules whose expressions profiles should be merged due to their similarity.



**Figure 2**

Gene modules identified by WGCNA (A) Dendrogram of all differentially expressed genes clustered based on a dissimilarity measure (1-TOM). TOM, topological overlap matrix. (B) Visualizing the gene network using a heatmap plot. The heatmap describes the Topological Overlap Matrix (TOM) among all genes in

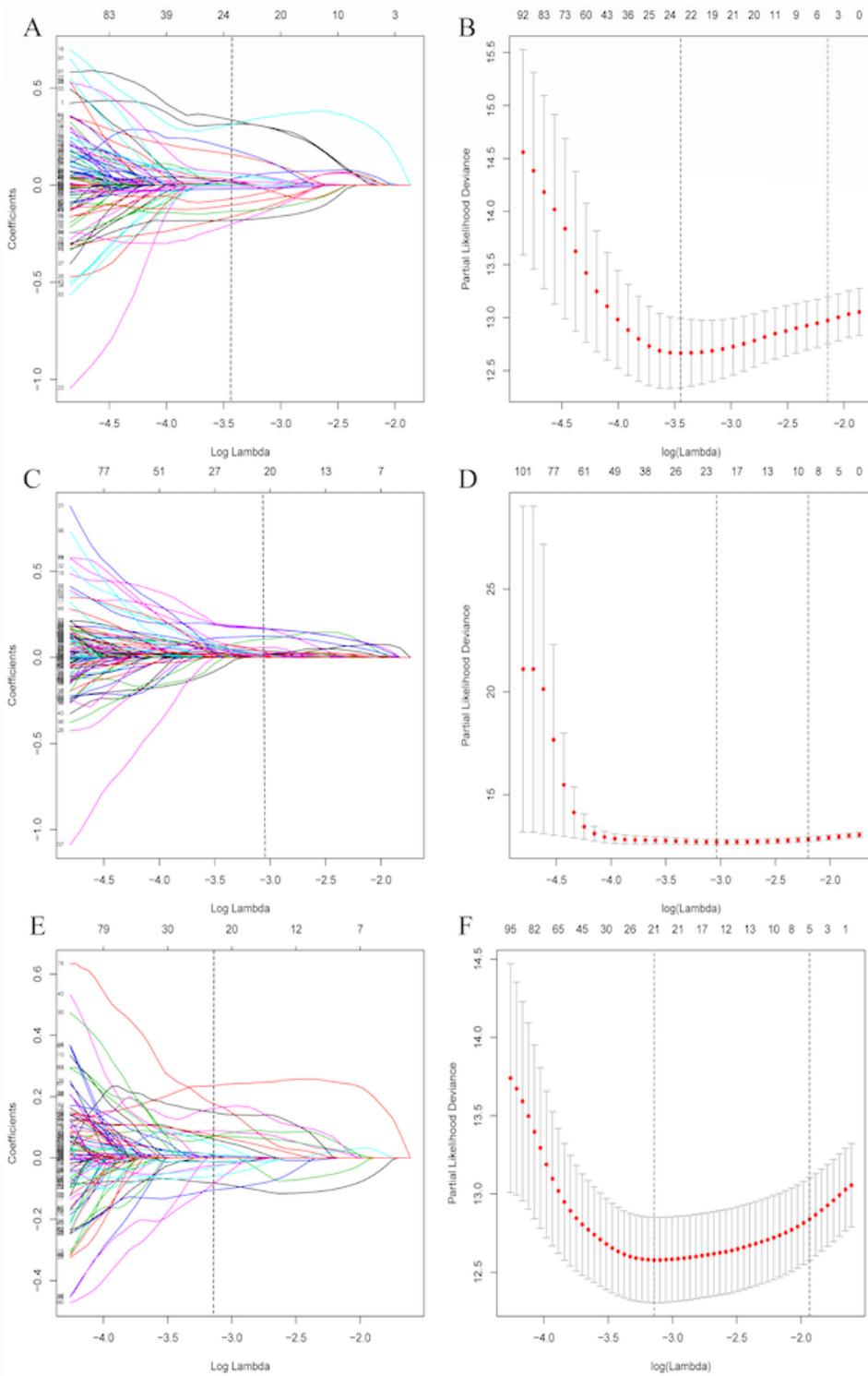
the analysis. The gene dendrogram and module assignment are also shown along the left side and the top. (C) Module-trait relationships. Each row presents a module eigengene, column to a clinic trait. Each cell contains the corresponding correlation and P value. The table is color-coded by correlation according to the color legend.



**Figure 3**

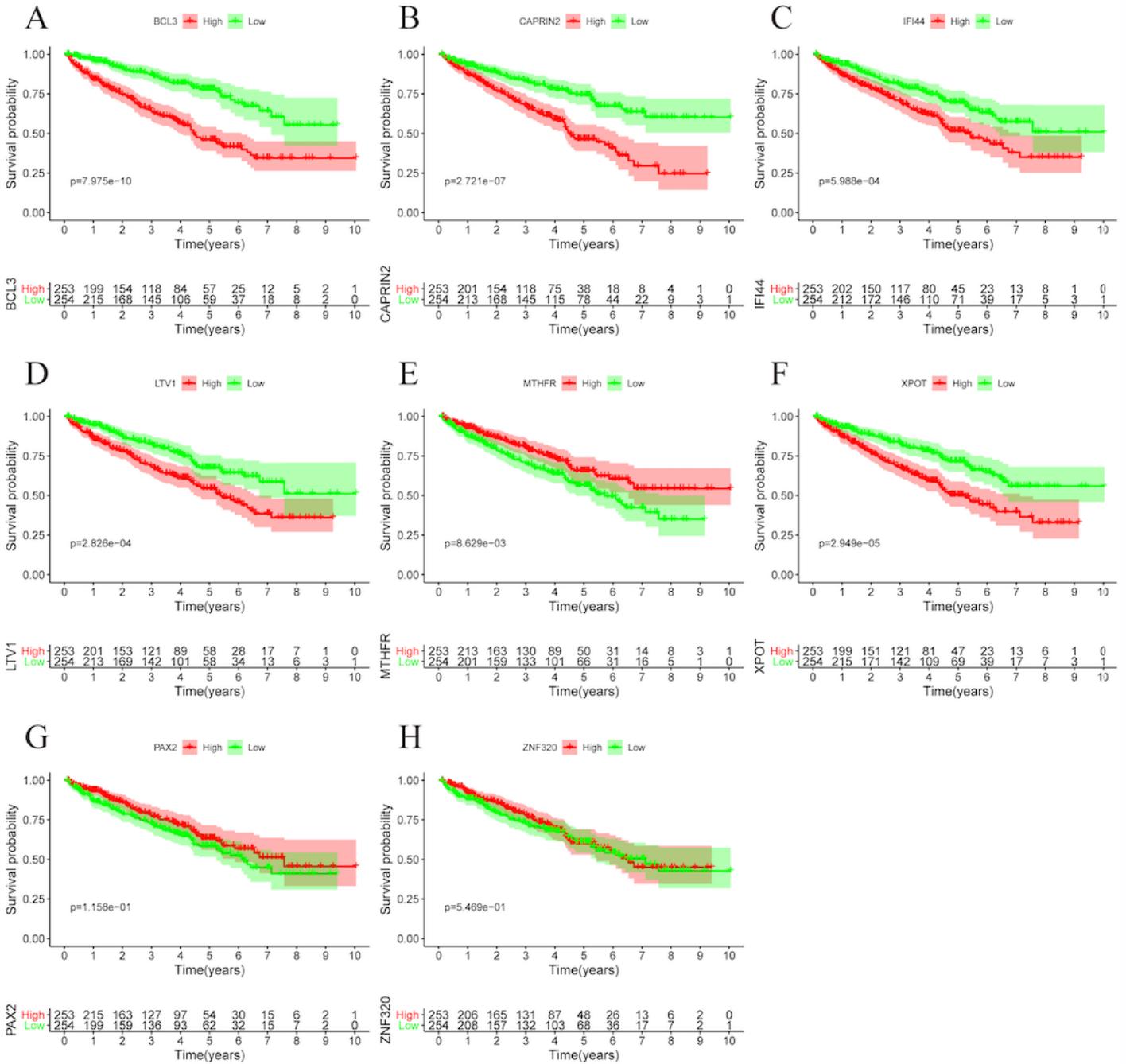
The eigengene dendrogram and heatmap identify groups of correlated eigengenes.





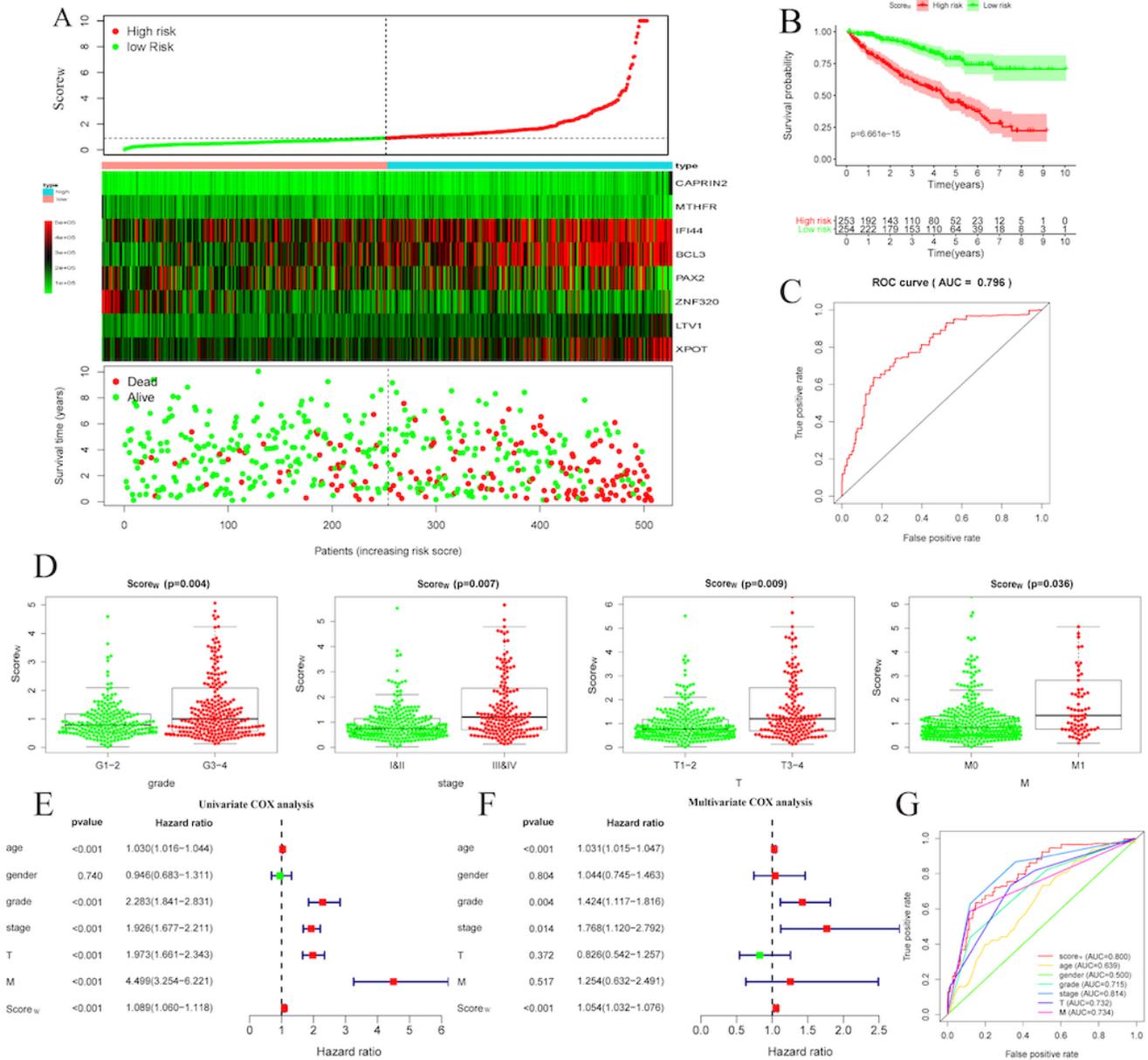
**Figure 5**

LASSO coefficient profiles of 3 modules. A dotted vertical line is drawn at the value identified by ten-fold cross-validation, where optimal  $\lambda$  results in nonzero coefficients. (A) Pink module. (B) Purple module. (C) Turquoise module.



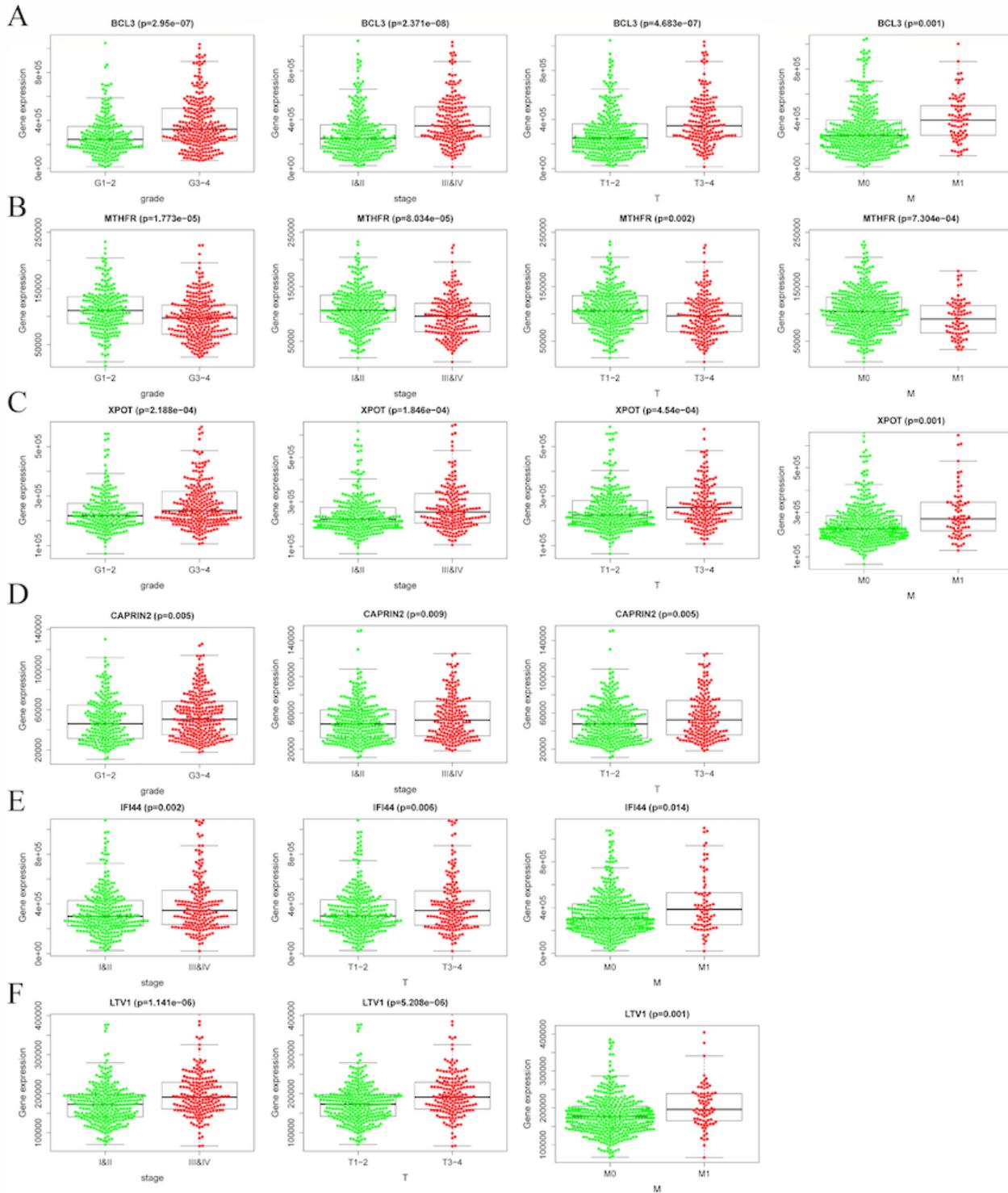
**Figure 6**

Kaplan-Meier curves of 8 genes. (A) BCL3 (B) CAPRN2 (C) IFI44 (D) LTV1 (E) MTHFR (F) XPOT (G) PAX2 (H) ZNF320



**Figure 7**

Construction of ScoreW. (A) The ScoreW distribution, gene expression profiles, and survival status of ccRCC patients. The black-dotted line represents the median cut-off, dividing patients into high- and low-risk groups. (B) Kaplan-Meier for the signature. Patients with high ScoreW had poor outcome in terms of overall survival. (C) ROC curves of signature. AUC=0.796. (D) Box plot visualization of relationship between ScoreW and clinical pathological parameters. (E) Univariate Cox regression analysis of clinical parameters and ScoreW. (F) Multivariate Cox regression analysis of clinical parameters and ScoreW. (G) ROC curves of clinical parameters and ScoreW



**Figure 8**

Correlation between clinical traits and genes in ScoreW. (A) Clinical relevance analysis of BCL3. (B) Clinical relevance analysis of MTHFR. (C) Clinical relevance analysis of XPOT. (D) Clinical relevance analysis of CAPRN2. (E) Clinical relevance analysis of IFI44. (F) Clinical relevance analysis of LTV1.

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

- [Dataandmaterials.rar](#)