

# Molecular Characterization, Clinical Significance, Tumor Microenvironment Association and Drug Susceptibility of DNA Methylation-Driven Genes in Renal Cell Carcinoma

**Jinpeng Wang**

The Second Affiliated Hospital of Harbin Medical University

**Pengfei Wu**

The Second Affiliated Hospital of Harbin Medical University

**Wei Zhang**

The Second Affiliated Hospital of Harbin Medical University

**Wenbin Hou**

The Second Affiliated Hospital of Harbin Medical University

**Enyang Zhao**

The Second Affiliated Hospital of Harbin Medical University

**Xuedong Li** (✉ [h06370@hrbmu.edu.cn](mailto:h06370@hrbmu.edu.cn))

The Second Affiliated Hospital of Harbin Medical University <https://orcid.org/0000-0002-1872-4213>

---

## Primary research

**Keywords:** renal cell carcinoma, molecular characterization, methylation-driven genes, tumor microenvironment, drug susceptibility

**Posted Date:** June 23rd, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-621150/v1>

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

[Read Full License](#)

---

# Abstract

**Background:** Accumulating evidence suggests that DNA methylation has essential roles in the development of renal cell carcinoma (RCC). Aberrant DNA methylation acts as a vital role in RCC progression through regulating gene expression, yet little is known about the role of methylation and its association with prognosis in RCC. The purpose of this study is to explore the DNA methylation-driven genes for establishing prognostic-related molecular clusters and providing a basis for survival prediction.

**Methods:** Differentially expressed genes (DEGs) were calculated by the “limma” R package. DNA methylation-driven genes (MDGs) were identified using the “methylMix” R package. The differentially expressed DNA methylation-driven genes (DEMDGs) were obtained by intersecting MDGs and DEGs. RCC data sets were divided into two groups using “ConsensusClusterPlus” R packet. The prognostic model was constructed based on multivariate Cox regression analysis. Based on the optimal cut-off value of the risk score, the patients were divided into high-risk group and low-risk group. The potential molecular mechanisms of prognostic-related DEMDGs were detected by GO, KEGG and GSEA analysis.

**Results:** In this study, 146 DEMDGs were selected and two clusters were distinguished by consensus clustering. We further evaluated the immune status of two clusters and selected 106 DEGs in cluster 1. Functional enrichment analysis of 106 DEGs and cluster-based immune status analysis provided new insights for the development of RCC. The prognostic model based on 17 DEMDGs was constructed to predict prognosis of RCC patients. The predictive nomogram and the web-based survival rate calculator (<http://127.0.0.1:7634/>) were built to validate predictive accuracy of prognostic model. Furthermore, the risk scores were strongly associated with clinical features, immune status and drug susceptibility.

**Conclusion:** Novel prognosis-related clusters based on the 146 DEMDGs were constructed and might contribute to precision medicine development. The 17 DEMDGs were utilized to construct a prognostic model for RCC prediction, which was significantly correlated with prognosis, immune infiltration, clinical features and drug sensitivity.

## Background

Renal cell carcinoma (RCC) is the most common urologic cancer types. There were an estimated nearly 65,340 new cases and 14,970 deaths worldwide in 2020[1]. With the advancement of diagnostic approaches, an increasing number of RCC could be diagnosed at early-stage[2]. Early diagnosis of RCC are essential for prolonging overall survival (OS) of patients[3]. Currently, the most curative treatment for localized RCC is still considered to be surgical resection[4]. Nowadays, there are already many surgical methods to remove tumors, including nephron-sparing surgery, radical nephrectomy and laparoscopic surgery[4]. However, the treatment options are still limited for unresectable and metastatic RCC[5, 6]. Early and accurate diagnosis of RCC have been regarded as a research priority[3]. Recently, with the biomedical research progresses, molecular prognostic biomarkers have become one of the basic ideas of precision medicine. Unfortunately, early-stage diagnosis for RCC by molecular prognostic biomarkers has many

challenges due to the lack of biomarkers for prediction of progression[7]. For this reason, more molecular biomarkers are urgently needed to screen for RCC diagnosis.

Modifications of the epigenome, such as DNA methylation, play a crucial role in development of many diseases[8]. The correct methylation pattern is very important for normal biological functions, and aberrant methylation is one of the drivers for progression of several diseases, especially cancer[9]. Numerous prior studies suggested that hypermethylated and hypomethylated DNA always show different activities[10]. Abnormal methylation always occurs in cancer cells, leading to some genes aberrantly activated and some genes aberrantly silenced[11]. Hypomethylation of proto-oncogenes or tumor suppressor gene methylation is considered one of the leading mechanisms of tumorigenesis in many cancer types[12, 13]. Therefore, the detection of the methylation patterns alteration of specific gene can aid the cancer diagnosis. Silencing of tumor suppressor genes caused by promoter hypermethylation provides new ideas for inquiring the molecular mechanisms of RCC[14]. The aberrant methylation is involved in the progression of RCC. Some studies found that the DNA methylation in RCC silenced the von Hippel-Lindau (VHL) tumor-suppressor gene[15]. In addition, RCC can be genotyped based on DNA methylation mutations[16].

There have been many studies focused on DNA methylation or gene expression. However, RCC prognostic models based on DNA methylation driven genes have barely been explored. In this study, we established a prognostic model to accurately predict patient survival. In addition, we divided RCC samples into two clusters according to 146 DEMDGs and further explored the relationship between the tumor immune status and clusters of RCC. The results we distilled will ultimately contribute to improving diagnostic accuracy and efficacy in immunotherapy.

## Materials And Methods

### Date collection

A total of 28 RNA-seq transcriptional profiling of normal samples were downloaded from GTEx dataset (<https://gtexportal.org/>). A total of 1021 RNA-seq transcriptome profiling (128 normal samples and 893 RCC samples), 872 DNA methylation data (205 normal samples and 667 tumor samples) and corresponding clinical information of RCC were downloaded from TCGA dataset (<https://gdc.cancer.gov/>).

### Differentially expressed genes (DEGs) screening in RCC and heatmaps plotting

We standardized RNA-seq transcriptional profiling by using “limma” R package, and Wilcoxon rank-sum test was utilized to identify DEGs[17, 18]. The false discovery rate (FDR) < 0.05 and  $|\log_2$  fold change (FC)| > 2 were taken advantage of as cutoff criteria. The “pheatmap” R package was used to plot the heatmaps[19].

### Integrated analysis of gene methylation data and gene expression data

Gene expression data and DNA methylation data were standardized by using “limma” R package[17]. DNA methylation-driven genes (MDGs) were identified using the “methylMix” R package[18, 20]. DNA methylation data and paired gene expression data were integrated and analyzed jointly to identify DNA methylation status negatively correlated with gene expression of a particular gene, indicating that the gene is a DNA methylation-driven gene[21]. Inclusion criteria were set to the correlation of methylation data and corresponding gene expression data of DEGs less than -0.3,  $|\log_2FC| > 0$  and adjust p < 0.05. The differentially expressed DNA methylation-driven genes (DEMDGs) were obtained by intersecting MDGs and DEGs for further analysis[18].

### **Construction of PPI network**

The PPI network was established by the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) database (<https://string-db.org/>) and visualized by Cytoscape software (v3.8.2)[22].

### **Evaluation of immune status and boxplots plotting**

The immune cells infiltration levels in the samples were evaluated by “CIBERSORT” R package, and the inclusion criteria was  $p < 0.05$  [23]. Stromal score and immune score also estimated by “estimate” R package[24]. Boxplots were plotted using the “ggpubr” R package [25].

### **Consensus clustering analysis**

We capitalized on k-means clustering algorithm in the “ConsensusClusterPlus” R packet to perform clustering [26]. The similarity in samples distance was measured by computing Euclidean distance and k-means algorithm to determine the credibility of our classification. The RCC datasets were grouped into distinct and nonoverlapping groups according to the consistent expression of 146 DEMDGs.

### **Construction of prognostic model**

Utilizing the “glmnet” R package, “survival” R package and “survminer” R package, the prognostic model was constructed [27-29]. The risk scores were calculated according to a linear combination of the gene expression levels weighted by the regression coefficients from the multivariate Cox regression analysis[30]. The “survivalROC” R package was utilized to validate the stability of the prognostic model[31]. The Kaplan-Meier survival curves were carried out to assess the survival time between high- and low-risk score RCC patients[32]. We validated the accuracy of optimal cut-off value by the principal component analysis[33].

### **Independence of the prognostic model from clinical features**

We evaluated the independence of the prognostic model from clinical features via univariate and multivariate Cox regression analysis[34]. The significant levels were set to  $p < 0.05$  and hazard ratios (HRs) with 95% CIs were also calculated.

## Construction of the nomogram and the dynamic nomogram

The nomogram was constructed utilizing the “rms” R package [35]. The web-based survival rate calculator was established using the “shiny” and “DynNom” R packages to predict cancer-specific survival rates dynamically[36, 37].

## Gene set enrichment analysis (GSEA) and column diagrams plotting

Gene set enrichment analysis (GSEA) was performed using GSEA4.0 (<https://www.gsea-msigdb.org/gsea/index.jsp/>). The annotated gene set files (c2.cp.kegg.v7.4.symbols.gmt gene set and c5.go.bp.v7.4.symbols.gmt) were considered as the reference gene set. The inclusion criteria were  $p < 0.05$  and  $FDR < 0.25$ . The column diagrams were plotted by GraphPad Prism 7 [38].

## Functional and pathway enrichment analysis

Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis were performed using “enrichplot” R package, “org.Hs.eg.db” R package and “clusterProfiler” R package[39, 40]. The inclusion criteria were set to  $p < 0.05$  and  $q < 1$ . The results were visualized by “ggplot2” R package[41].

## Statistical analysis

Differential methylation calculated from  $\text{mean}(\beta\text{-value-cancer}) - \text{mean}(\beta\text{-value-normal})$ . The differences in variables among different groups were compared by means of the Student’s t-test. All statistical analysis was conducted using R software (version 4.0.3) (<http://www.r-project.org/>), Rstudio software (version 1.4.1106) (<https://www.rstudio.com/>), Origin (version 2020) and GraphPad Prism7 (GraphPad Software Inc., La Jolla, CA).  $P < 0.05$  was considered to be statistically significant.

# Results

## Identification of 146 DEMDGs in RCC

The research process of the study was showed in Fig. 1. Based on Wilcoxon rank test, a total of 5198 DEGs were selected from 28 RNA-seq transcriptome profiling of normal samples in GTEx dataset and 1021 RNA-seq transcriptome profiling (893 RCC samples and 128 normal samples) in TCGA dataset ( $FDR < 0.05$  and  $|\log_2FC| > 2$ ) (Table S1). The heatmap shows the expression of DEGs between RCC samples and normal samples (Fig. 2A). We screened the 270 methylation-driven genes (MDGs), whose methylation status negatively correlated with expression levels ( $Cor < -0.3$ ,  $|\log_2FC| > 0$  and  $\text{adjust } p < 0.05$ ) (Table S2). The heatmap shows the expression of MDGs between RCC samples and normal samples (Fig. 2B). Then, 270 MDGs and 5198 DEGs were intersected to obtain 146 DEMDGs for further analysis (Fig. 2C). We further visualized the methylation levels (Fig. 2D) and gene expression levels (Fig. 2E) of 146 DEMDGs in RCC samples and normal samples. The comprehensive landscape of DEMDGs interactions and DEMDGs connection for RCC patients was depicted with the DEMDGs network (Fig. 2F).

The above results found 146 significantly DEMDGs in RCC and normal samples. These abnormal DEMDGs were interconnected and may be involved in the occurrence and development of RCC.

### **Consensus clustering based on 146 DEMDGs and immune status analysis of clusters**

To select optimized cluster number, we calculated k-means clustering algorithm with the ConsensusClusterPlus R packet. K = 2 was identified with optimal clustering stability (Fig. 3A-D). Then, we analyzed the methylation levels and gene expression levels of 146 DEMDGs, as well as the clinical features of paired patients. There were significant differences in the methylation levels and gene expression levels between cluster 1 and cluster 2, and the clinical features were evenly distributed in two clusters (Fig. 3E-F). The RCC patients in cluster 2 (n = 419) had a better overall survival (OS) than the patients in cluster 1 (n = 435,  $p < 0.001$ ) (Fig. 3G).

Immune checkpoint inhibitors (ICIs) are administered for the treatment of RCC. We investigated whether the two clusters were related to ICI-related biomarkers. The results showed that cluster 1 was positively correlated with high expression of LAG3 ( $p < 0.001$ ), CD160 ( $p < 0.001$ ), HAVCR2 ( $p < 0.001$ ), CTLA4 ( $p < 0.001$ ) and TIGIT ( $p < 0.001$ ), and the stromal score and immune score were significantly higher in cluster 1 compared with cluster 2 ( $p < 0.001$ ) (Fig. 4A). The abundance of B cells naive ( $p < 0.001$ ), T cells CD8 ( $p < 0.001$ ), T cells CD4 memory activated ( $p < 0.001$ ), T cells follicular helper ( $p < 0.001$ ), T cells gamma delta ( $p < 0.001$ ) and Macrophages M1 ( $p < 0.001$ ) was significantly higher in cluster 1 compared with cluster 2 (Fig. 4B). The higher immune infiltration level corresponded to cluster 1, and lower immune infiltration level corresponded to cluster 2 (Fig. 4C). The RCC patients in high-immune score group had a worse OS than the patients in low-immune score group ( $p < 0.001$ ) (Fig. 4D). Besides, we accessed the correlation of immune cells in cluster 1 and cluster 2. In cluster 1, the positive correlation between T cells CD8 and T cells follicular helper was the strongest, in which correlation coefficient was 0.55. The correlation coefficient between T cells CD8 and T cells CD4 memory resting was -0.66, the lowest negative correlation (Fig. 4E). However, in cluster 2, the cells with the strongest negative correlation were activated T cells CD8 and Macrophages M2, in which correlation coefficient was -0.46 (Fig. 4F). These results showed that the two clusters based on 146 DEMDGs were closely associated with prognosis and immune status in RCC patients.

To explore the possible reasons causing worse OS in cluster 1, we selected 106 DEGs from two clusters ( $Cor < -0.3$  and  $|\log_2FC| > 1$ ). We used heatmap visualizing the gene expression levels of 106 DEGs in RCC and normal samples (Fig. S1). We performed GO and KEGG analysis to analyze underlying functions and pathways of 106 DEGs ( $p < 0.05$ ). Results of GO analysis were significantly enriched in regulation of T cell activation, T cell proliferation, positive regulation of leukocyte proliferation, positive regulation of T cell proliferation, positive regulation of cell-cell adhesion, etc. (Fig. 4G, S2A and S2B). Results of the KEGG pathways were significantly enriched in pathogenic Escherichia coli infection, natural killer cell mediated cytotoxicity, viral myocarditis, one carbon pool by folate, cytokine-cytokine receptor interaction, etc. (Fig. 4H). The above results indicated that the 106 DEGs were in close contact with immune microenvironment, which may be the causing for OS difference between two clusters.

## Construction and evaluation of the prognostic model

To determine the prognostic value of 146 DEMDGs, univariate Cox regression analysis, LASSO and multivariate Cox regression analysis were used to identify them. Subsequently, the prognostic model was constructed based on 17 independent and prognostic DEMDGs (including TRIM4, TCF19, SH3BGR, PPP1R18, NMI, NCKAP1L, NAPSA, MYH14, MOB3A, FMNL1, FAXDC2, ESRP2, EMP3, CX3CL1, C1orf54, BST2, BDH1) (Fig. 5A-D). We analyzed the association between the gene expression and the survival of the patients. The patients' OS with high-expression of TRIM4, TCF19, SH3BGR, PPP1R18, NMI, NAPSA, MYH14, MOB3A, FMNL1, FAXDC2, ESRP2, EMP3, CX3CL1, C1orf54, BST2 and BDH1 were worse than the low-expression group ( $p < 0.05$ ) (Fig. S3). The methylation levels of 17 DEMDGs were inversely correlated with their expression level ( $p < 0.001$ ) (Fig. S4). The risk score was calculated as follows = SH3BGR\* (-0.0896) + BDH1\*(-0.0667) + MOB3A\*(-0.0244) + CX3CL1\*(-0.0343) + NMI\*(0.0516) + NAPSA\*(-0.0068) + PPP1R18\*(0.0375) + TCF19\*(0.0231) + FMNL1\*(0.0946) + C1orf54\*(-0.0348) + FAXDC2\*(-0.03550) + BST2\*(0.0037) + MYH14\*(-0.0410) + ESRP2\*(-0.0836) + NCKAP1L\*(-0.1905) + TRIM4\*(-0.0840) + EMP3\*(-0.0132). The coefficients of each gene were showed in Table 1.

RCC patients were split into high- and low- risk groups according to optimal cut-off value of the risk score (Fig. 5E). The AUC of the ROC curves was 0.788, 0.743 and 0.747 within 1-, 2-, 3-year, which demonstrated that risk score had a good prognostic value (Fig. 5F). The distributions of the risk score in high- and low-risk groups were showed in Fig. 5G. Patients' mortality risk increased with increasing risk score (Fig. 5H). The survival curve was carried out to assess the survival time between high- and low-risk score groups. The survival time of high-risk group was significantly worse than the low-risk group ( $p < 0.001$ ) (Fig. 5I). RCC samples were clearly structured in two different groups by the principal component analysis, which suggested our study could significantly reflect the prognosis differences of RCC patients (Fig. 5J). To investigate the prognostic value of the signature for OS in RCC patients stratified by clinical features, RCC patients were stratified according to gender, grade, clinical stage, T stage, N stage and M stage. For all different stratifications, the OS time of the high-risk group was shorter than that of the low-risk group ( $p < 0.001$ ) (Fig. 6). These results showed that the 17 DEMDGs for OS could accurately predict the prognosis of RCC patients and that the prognostic model could accurately reflect the survival of patients under different clinical features.

## Clinical verification of prognostic model and enrichment analysis of prognostic model

We investigated the relationship between the risk score of RCC and several variables. We found that T stage ( $p < 0.001$ ), N stage ( $p < 0.05$ ), clinical stage ( $p < 0.0001$ ), tumor grade levels ( $p < 0.05$ ), cluster ( $p < 0.0001$ ) and immune score ( $p < 0.01$ ) were significantly associated to the risk score. The risk score level of the T1, T2 and N0 stage were clearly lower than that of the T3, T4 and N1, N2 stage (Fig. 7A-B). The higher level of clinical stage had higher level of risk score (Fig. 7C). The risk score increased with the grade of the tumor (Fig. 7D). The risk score of the cluster 1 was significantly higher than that of the cluster 2 (Fig. 7E). The low immune score group had lower risk score than high immune score group (Fig. 7F).

To further annotate functions enriched in the high- and low-risk groups, GSEA was queried to confirm the signaling pathways in which the genes are enriched. The results were represented in Fig. 7G-J ( $p < 0.05$ ). The following biological processes were enriched in the high-risk group: regulation of DNA damage response signal transduction by p53 class mediator, chondrocyte development, negative regulation of gene expression epigenetic, positive regulation of leukocyte proliferation and cardiac epithelial to mesenchymal transition. The following signaling pathways were enriched in the high-risk group: cytokine-cytokine receptor interaction, vascular smooth muscle contraction, cell cycle, nod-like receptor signaling pathway, jak-stat signaling pathway. To clarify the possible molecular mechanism of 17 prognosis-related DEMDGs, we also performed GO and KEGG pathway analysis. Results of GO analysis were significantly enriched in negative regulation of cytokine production, response to interferon-gamma, cortical actin cytoskeleton organization, multivesicular body, SCAR complex, Rac GTPase binding, actin binding, etc. (Fig. 7K, S5A and S5B). Results of KEGG analysis were significantly enriched in synthesis and degradation of ketone bodies, pathogenic *Escherichia coli* infection, regulation of actin cytoskeleton, butanoate metabolism (Fig. 7L). These results suggested that the risk score was correlated with clinical features, cluster and immune score and that the above biologic functions and signaling pathways might impact RCC patients' prognosis.

### **The predictive accuracy of prognostic model**

In order to determine whether the risk score could be presented as an independent prognostic factor for RCC patients, we employed univariate and multivariate Cox proportional hazards regression analysis. In the univariate analysis and multivariate analysis, the risk score, age and M stage showed pronounced effects on the RCC prognosis ( $p < 0.05$ ) (Fig. 8A-B). Beyond this, we constructed a nomogram with the significant variables in the multivariate analysis (Fig. 8C). Results suggested that risk score had the significant influence on survival prediction. The 1-, 2-, and 3-year predicted calibration curves also respectively suggested that the model had a good prediction accuracy (Fig. 8D-F). We also established a dynamic web-based survival rate calculator (<http://127.0.0.1:7634/>), which could individually predict the survival of patients according to their clinical features and risk score. For example, the 3-year cancer-specific survival rate was approximately 55% (95% CI 42–72%) for patients with low risk, M0 stage and aged < 65 years (Fig. 8G-H).

### **Prognostic model correlated with tumor-infiltrating immune cells and drug susceptibility**

To further analyze the relationship between prognostic model and tumor-infiltrating immune cells, we performed a detailed spearman correlation analysis and the result was presented with the lollipop shape (Fig. 9A). High-risk group were more positively correlated with tumor-infiltrating immune cells, including CD8<sup>+</sup> T cells, Macrophage M1, B cell, monocytes, Myeloid dendritic cell, T cell regulatory and myeloid dendritic cells, etc. The detailed results were showed in Table S3. We also attempted to identify associations between prognostic model and the efficacy of six common chemotherapeutic drugs for the treatment of RCC. The high-risk score was associated with the lower half maximal inhibitory concentration (IC50) of chemotherapeutics such as Temsirolimus ( $p = 1.1e-09$ ), Sunitinib ( $p < 2.22e-16$ ),

Pazopanib ( $p = 0.0041$ ), Doxorubicin ( $p = 0.0054$ ), Bleomycin ( $p = 0.012$ ), Axitinib ( $p = 7.1e-06$ ) (Fig. 9B-G). The above results showed that this model closely correlated with tumor-infiltrating immune cells and drug susceptibility.

## Discussion

Previous studies had described that RCC possesses a high number of genetic alterations and epigenetic alterations[42, 43]. As the major epigenetic modification, DNA methylation studies have become a research hotspot in many cancers, specifically in RCC[44]. In this paper, we identified 146 DEMDGs by integrating transcriptomic and DNA methylation profiles. Two clusters were identified by consensus clustering based on 146 DEMDGs, and we developed a reliable prognostic model consisting of 17 DEMDGs.

We selected 5198 DEGs from normal samples and RCC samples. Then, 270 MDGs of RCC were identified by using the MethyIMix algorithm. 146 DEMDGs were identified by intersecting 5198 DEGs and 270 MDGs. RCC patients were successfully divided into two clusters based on the 146 DEMDGs, and patients of different clusters had different clinical features, methylation level and gene expression level. The OS of patients in cluster 2 was significantly longer than those in cluster 1. Immune status, immune score, immune checkpoints and infiltrating percentage of immune cells in two clusters also showed significant differences. Two clusters had different survival rate for the following possible reasons. (1) Aberrant DNA methylation could contribute to tumor progression due to gene aberrant transcriptional responses[45]. Aberrant DNA methylation patterns are a feature of tumor development[46]. (2) Tumor-infiltrating immune cells of RCC could influence the prognosis and progression of tumor[47]. Additionally, cluster 1 had the higher immune score and immune cells infiltration. There were studies reporting that high immune score as well as high infiltration of immune cells were associated with poor prognosis, which was similar with our results[48]. We identified 106 DEGs from cluster 1 to perform further analysis. Many results of biological processes were significantly enriched in immunity, including positive regulation of I-kappaB kinase/NF-kappaB signaling, leukotriene D4 metabolic process, etc.. This provided further evidence that the different immune status of two clusters may be the possible cause for different survival status.

Subsequently, 17 DEMDGs (including SH3BGR, BDH1, MOB3A, CX3CL1, NMI, NAPSA, PPP1R18, TCF19, FMNL1, C1orf54, FAXDC2, BST2, MYH14, ESRP2, NCKAP1L, TRIM4, EMP3) prognostic model was constructed to act as a reliable predictor by univariate and multivariate Cox analysis. Risk score also was calculated based on the gene expression and regression coefficients of each gene. Patients were divided into high-risk group and low-risk group based on their risk scores. The AUC of 1-, 2- and 3-year survival was 0.788, 0.743 and 0.747, respectively. Patients in the low-risk group had a longer overall survival compared with those in the high-risk group ( $p < 0.001$ ). For all different clinical features, the high-risk group had worse survival than the low-risk group significantly with  $P < 0.001$ . We used GSEA to confirm the signaling pathways where the genes were enriched in the high- and low-risk groups. The high-risk group mainly enriched in immune-related processes. Then, we performed GO analysis and KEGG analysis on the 17 DEMDGs. We noted that BP group of GO analysis were mainly enriched in immune-related

processes. These results suggest a closely relation between prognosis and immune status in RCC. The multivariate Cox regression analysis results indicated that our prognostic model was unaffected by clinical features. Besides, we constructed a nomogram and a dynamic web-based survival rate calculator to verify prediction accuracy of prognostic model. The above outcomes indicate that our prognostic model has a good predictive ability and excellent predictive accuracy.

We further explored the relationship between tumor-infiltrating immune cells and risk score with seven common acceptable methods of estimating the immunoinfiltrating cell, including TIMER[49], CIBERSORT[49], XCELL[50], QUANTISEQ[51], MCPcounter[52], EPIC[53], and CIBERSORT-ABS[54]. The synthetical analysis showed that risk score was more positively related to tumor-infiltrating immune cells. Finally, we evaluated the relationship between risk score and the efficacy of six common anti-tumor drugs the treatment of RCC. The results showed IC50s of six drugs significantly inversely correlated to risk score. Here, this study demonstrated that a novel prognostic model constructed by 17 DEMDGs that could predict prognosis for patients with RCC and might help distinguish those who could benefit from anti-tumor immunotherapy. However, our findings still need to be demonstrated by experimental methods.

## Conclusions

To sum up, our study indentified 146 DEMDGs in RCC. The consensus clustering based on 146 DEMDGs could be used to predict prognosis of RCC, and the clusters were associated with the immune microenvironment of RCC. We utilized 17 DEMDGs to construct a new OS-related prognostic model for early diagnosis of RCC, and risk score was significantly correlated with prognosis, immune infiltration, clinical features and drug sensitivity. Different from previous studies, we taken TCGA dataset and GTEx dataset together to perform analysis, overcoming the short plank of smaller numbers of normal samples in TCGA dataset. Notably, we broke the limitations of previous studies by analyzing all pathological types of RCC instead of the main pathological type.

## Abbreviations

RCC, renal cell carcinoma; TCGA, The Cancer Genome Atlas; GTEx, Genotype-Tissue Expression; DEGs, differentially expressed genes; MDGs, DNA methylation-driven genes; DEMDGs, differentially expressed DNA methylation-driven genes; GSEA, Gene set enrichment analysis; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; LASSO, least absolute shrinkage and selection operator; OS, overall survival; FDR, false discovery rate; FC, fold change; CDF, cumulative distribution function; HRs, hazard ratios; IC50, half maximal inhibitory concentration.

## Declarations

### Acknowledgements

This work was funded by the Scientific Research Project of Heilongjiang Provincial Health and Family Planning Commission (2017-070), the Second Affiliated Hospital of Harbin Medical University First-class Discipline First-class Specialist Construction Project (100123) and the Harbin Medical University Scientific Research Innovation Fund (YJSKYCX2018-95HYD).

### **Author contributions**

JPW and PFW originally designed the study and wrote this article. EYZ, WZ and WBH contributed to methodology assistance. JPW, PFW, EYZ, WZ, WBH and XDL collected data and prepared tables and Figs. All authors read and approved the final manuscript.

### **Availability of data and materials**

The datasets analyzed during the current study were generated by TCGA Research Network (<https://gdc.cancer.gov/>) and GTEx dataset (<https://gtexportal.org/>).

### **Ethics approval and consent to participate**

Not applicable.

### **Consent for publication**

Not applicable

### **Conflict of interests**

The authors declare no conflict of interest.

### **Author details**

<sup>1</sup>Department of Urology, The Second Affiliated Hospital of Harbin Medical University, Harbin, 150086, China

<sup>2</sup>Department of Neurosurgery, The First Affiliated Hospital of USTC, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei, Anhui, 230001, P.R. China

<sup>3</sup>Anhui Key Laboratory of Brain Function and Diseases, Hefei, Anhui, 230001, P.R. China

## **References**

1. Creighton, JJ.H.V.L.D.C.E.H.C.C.J.. Genomic classifications of renal cell carcinoma: a critical step towards the future application of personalized kidney cancer care with pan-omics precision. *J Pathol.* 2018;244(5):525–37.

2. Talonso-Gordoa, M.L.G.-B., Enrique Grande 3, Pilar Garrido 4, Alfredo Carrato 5, Javier Molina-Cerrillo 6, *Targeting Tyrosine kinases in Renal Cell Carcinoma: "New Bullets against Old Guys"*. Int J Mol Sci, 2019. **20**(8): p. 1901.
3. Camillo Porta MEG 1, Brian I. Rini 3, Bernard Escudier 4, Subramanian Hariharan 5, Lorna P Charles 5, Liqiang Yang 5, Liza DeAnnuntis 5, Robert J Motzer 6, *Long-term Safety of Sunitinib in Metastatic Renal Cell Carcinoma*. Eur Urol. 2016;69(2):345–51.
4. Ljungberg B, Hanbury CN, Hora DC, Kuczyk M, Merseburger MA, Patard AS, Mulders JJ, Sinescu PF. IC, EAU guidelines on renal cell carcinoma: the 2010 update. Eur Urol. 2010;58(3):398–406.
5. Shinder BM, Farrell RK, Farber D, Stein NJ, Jang MN, Singer TL. EA, Surgical Management of Advanced and Metastatic Renal Cell Carcinoma: A Multidisciplinary Approach. Front Oncol. 2017;7:107.
6. Li P, Armstrong WY, Haas K, Subedi N, Davis-Cerone P, Doshi M. JA, Survival among patients with advanced renal cell carcinoma in the pretargeted versus targeted therapy eras. Cancer Med. 2016;5(2):169–81.
7. Li H, Yu WG, Jian Y, Zhang W, Wang D, Wang Y, Meng T, Yuan Y, Zhang C. C,  $\alpha$ -1,2-Mannosidase MAN1C1 Inhibits Proliferation and Invasion of Clear Cell Renal Cell Carcinoma. J Cancer. 2018;9(24):4618–26.
8. Shen L, Zhang GG, Zhang Y, Ye H, Huang Z, Huang S, Kang J. J., A single amino acid substitution confers enhanced methylation activity of mammalian Dnmt3b on chromatin DNA. Nucleic Acids Res. 2010;38(18):6054–64.
9. Jeltsch A. J.R., Allosteric control of mammalian DNA methyltransferases - a new regulatory paradigm. Nucleic Acids Res. 2016;44(18):8556–75.
10. Park J, Zeng PZ, Elango J, Park N, Wheeler T, Werren D, Yi JH. SV, Comparative analyses of DNA methylation and sequence evolution using Nasonia genomes. Mol Biol Evol. 2011;28(12):3345–54.
11. Vasanthakumar A, Zegarek LJ, Kocherginsky MH, Singh M, Davis M, Link EM, Anastasi PA, Le Beau J, Karpf MM, Godley AR. LA, Dnmt3b is a haploinsufficient tumor suppressor gene in Myc-induced lymphomagenesis. Blood. 2013;121(11):2059–63.
12. Hayslip J. M.A., Tumor suppressor gene methylation in follicular lymphoma: a comprehensive review. Mol Cancer. 2006;6(5):44.
13. Ding LY, Kuo HY, Hsu IY, Tsai TY, Chang TC, Hsu HW, Tsao WY, Tian CC, Wang CC, Wang PS, Lee HC, Wang CT, Lin YC, Hughes SH, Chuang MW, Lu WJ, Shan PJ, Huang YS. PH., Epigenetic silencing of AATK in acinar to ductal metaplasia in murine model of pancreatic cancer. Clin Epigenetics. 2020;12(1):87.
14. Liu Q, Ying JJ, Cui J, Sun Y, Zhang M, Fan L, Xu Y, Zhang B. Q, Epigenetic inactivation of the candidate tumor suppressor gene ASC/TMS1 in human renal cell carcinoma and its role as a potential therapeutic target. Oncotarget. 2015;6(26):22706–23.
15. Grech G, Yoo ZX, Bubnov BC, Hagan R, Danesi S, Vittadini R, Desiderio G. DM, EPMA position paper in cancer: current overview and future perspectives. EPMA J. 2015;6(1):9.

16. Tian Y, Gotoh AE, Komiyama M, Fujimoto M, Kanai H. Y, Prognostication of patients with clear cell renal cell carcinomas based on quantification of DNA methylation levels of CpG island methylator phenotype marker genes. *BMC Cancer*. 2014;14:772.
17. Ritchie ME, Wu PB, Hu D, Law Y, Shi CW, Smyth W. GK, limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43(7):e47.
18. Zhang D, Hu WY. X., Identification and Comprehensive Validation of a DNA Methylation-Driven Gene-Based Prognostic Model for Clear Cell Renal Cell Carcinoma. *DNA Cell Biol*. 2020;39(10):1799–812.
19. Szekely G, Rizzo M. *Hierarchical Clustering via Joint Between-Within Distances: Extending Ward's Minimum Variance Method*. *Journal of Classification*, 2005(22): p. 151–183.
20. O G. MethylMix: an R package for identifying DNA methylation-driven genes. *Bioinformatics*. *Bioinformatics*. 2015;31(11):1839–41.
21. Cedoz PL, Brennan PM, Gevaert K. O, MethylMix 2.0: an R package for identifying DNA methylation genes. *Bioinformatics*. 2018;34(17):3044–6.
22. Shannon P, Ozier MA, Baliga O, Wang NS, Ramage JT, Amin D, Schwikowski N, Ideker B. T, Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003;13(11):2498–504.
23. Newman AM, Green LC, Gentles MR, Feng AJ, Xu W, Hoang Y, Diehn CD, Alizadeh M. AA, Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods*. 2015;12(5):453–7.
24. Jia D, Li LS, Xue D, Yang H, Liu D, Mining Y. TCGA database for genes of prognostic value in glioblastoma microenvironment. *Aging*. 2018;10(4):592–605.
25. Whitehead MJ, Willison MG, Barnett HJ. SC., MyelinJ: an ImageJ macro for high throughput analysis of myelinating cultures. *Bioinformatics*. 2019;35(21):4528–30.
26. Wilkerson MD. H.D., ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics*. 2010;26(12):1572–3.
27. Armbruster M, Seifert RM, Frommherz S, Weinert L, Mack C, Roth C, Merz A, Bunzel B, Krüger D, Kulling R, Watzl S, Bub B. A, Metabolite profiles evaluated, according to sex, do not predict resting energy expenditure and lean body mass in healthy non-obese subjects. *Eur J Nutr*. 2019;58(6):2207–17.
28. Xi J, Liang YJ, Zhan J, Jiang C, Lin W, Xu Z, Wang S. Q, *Prognostic Impact of Radiological Consolidation Tumor Ratio in Clinical Stage IA Pulmonary Ground Glass Opacities*. *Front Oncol*. 11: p. 616149.
29. Wang L, Liu CF, Shi R, Zhao L, Yan G. Z, Gene expression and immune infiltration in melanoma patients with different mutation burden. *BMC Cancer*. 2021;21(1):379.
30. Wei X, Wang WW, Wang H, Wang Y, Li Y, Ji G, Ren C, Song X, Qin N. C, Identification of an independent autophagy-gene prognostic index for papillary renal cell carcinoma. *Transl Androl Urol*. 2020;9(5):1945–56.

31. Huang R, Li LX. Q, Identification and validation of potential prognostic gene biomarkers for predicting survival in patients with acute myeloid leukemia. *Onco Targets Ther.* 2017;10:5243–54.
32. Lawrie GM, Calhoun MGJ, Safi JH, Zamora H, Beltengady JL, Baron M, Silvers A A, Chapman DW., *linical results of coronary bypass in 500 patients at least 10 years after operation.* *Circulation*, 1982. **125**(6): p. 1432-7.
33. Kim S, Rhyu PJ, Nam S, Lee J. K, Quantitative analysis of piano performance proficiency focusing on difference between hands. *PLoS One.* 2021;16(5):e0250299.
34. Qi LW, Jiang JJ, Hu CH. JM, Contributions and Prognostic Values of N6-Methyladenosine RNA Methylation Regulators in Hepatocellular Carcinoma. *Front Genet.* 2021;11:614566.
35. Pond GR, Bellmunt AN, Choueiri J, Qu TK, Fougeray A, Vaughn R, James D, Salhi ND, Albers Y, Niegisch P, Galsky G, Wong MD, Ko YN, Stadler YJ, O'Donnell WM, Sridhar PH, Vogelzang SS, Necchi NJ, Di Lorenzo A, Sternberg G, Mehta CN, Sonpavde A. G, A nomogram including baseline prognostic factors to estimate the activity of second-line therapy for advanced urothelial carcinoma. *BJU Int.* 2014;113(5b):E137-43.
36. Bakin EA, Danilenko SO, Lioznov DM, Kulikov DA. AN, Fast prototyping of a local fuzzy search system for decision support and retraining of hospital staff during pandemic. *Health Inf Sci Syst.* 2021;9(1):21.
37. Sun W, Song LG, Zhu Y, Yang Z, Chen Z, Miao Y, Song J, Lan X, Qiu Y, Zhu X, Fan S. Y, A web based dynamic MANA Nomogram for predicting the malignant cerebral edema in patients with large hemispheric infarction. *BMC Neurol.* 2020;20(1):360.
38. Augustine D, Anbu RR, Chidambara Murthy J. KN, *In vitro Antiproliferative Effect of Earthworm Coelomic Fluid of Eudrilus Eugeniae, Eisenia Foetida, and Perionyx Excavatus on Squamous Cell Carcinoma-9 Cell Line: A Pilot Study.* *Pharmacognosy Res*, 2017.
39. Zhang C, Li ZY, Hu X, Qi X, Luo F. J, Genome-wide mutation profiling and related risk signature for prognosis of papillary renal cell carcinoma. *Ann Transl Med.* 2019;7(18):427.
40. Yu G, Han WL, He Y. QY, clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS.* 2012;16(5):284–7.
41. Sun X. W.K., Cook D, *PKgraph: an R package for graphically diagnosing population pharmacokinetic models.* *Comput Methods Programs Biomed*; 2011.
42. Chhabra R. N.M., Lysophosphatidic acid reverses Temsirolimus-induced changes in lipid droplets and mitochondrial networks in renal cancer cells. *PLoS One.* 2020;15(6):e0233887.
43. Wang H, Dai QY, Zhu B, Shi Y, Zhu G, Shen Y, Zhang Y, Ye H. D, PBRM1 regulates proliferation and the cell cycle in renal cell carcinoma through a chemokine/chemokine receptor interaction pathway. *PLoS One.* 2017;12(8):e0180862.
44. Li X. Y.Q., PON1 hypermethylation is associated with progression of renal cell carcinoma. *J Cell Mol Med.* 2019;23(10):6646–57.
45. Yang Y, Zhang WF, Sun J, Li R, Li F, Chang Y, Wang S, Wang L, Liu X, Huang L. C., EGR1 interacts with DNMT3L to inhibit the transcription of miR-195 and plays an anti-apoptotic role in the development

- of gastric cancer. *J Cell Mol Med.* 2019;23(11):7372–81.
46. Xu Y, Choi HB, Gopalan AJ, Lee B, Kalady BH, Church MF, Ting JM. AH, Unique DNA methylome profiles in CpG island methylator phenotype colon cancers. *Genome Res.* 2012;22(2):283–91.
47. Xing X, Leng JS, Wang Y, Li Q, Dong Z, Guo B, Cheng T, Du X, Hu H, Feng Y, Lian Q, Luan S, Ma F, Li X, Ni Z, Li M, Ji Z. J, An integrated classifier improves prognostic accuracy in non-metastatic gastric cancer. *Oncoimmunology.* 2020;9(1):1792038.
48. Deng X, Zhang LD, Shen X, Yang X, Yang Z, Lu L, Yu X, Zhang L, Lin N. J, Profiles of immune-related genes and immune cell infiltration in the tumor microenvironment of diffuse lower-grade gliomas. *J Cell Physiol.* 2020;235(10):7321–31.
49. Bu F, Liu ZX, Lin S, Zhu K, Huang J. J, Comprehensive analysis of Syk gene methylation in colorectal cancer. *Immun Inflamm Dis.* 2021;118:109228.
50. Oshi M, Le HM, Tokumaru L, Yan Y, Matsuyama L, Endo R, Takabe I. K, Abundance of Microvascular Endothelial Cells Is Associated with Response to Chemotherapy and Prognosis in Colorectal Cancer. *Cancers (Basel).* 2021;13(6):1477.
51. Finotello F, Plattner MC, Laschober C, Rieder G, Hackl D, Krogsdam H, Loncova A, Posch Z, Wilflingseder W, Sopper D, Ijsselsteijn S, Brouwer M, Johnson TP, Xu D, Wang Y, Sanders Y, Estrada ME, Ericsson-Gonzalez MV, Charoentong P, Balko P, de Miranda J, Trajanoski NFDCC. Z, Molecular and pharmacological modulators of the tumor immune contexture revealed by deconvolution of RNA-seq data. *Genome Med.* 2019;11(1):34.
52. Li Y, Xiao ZX, Yang H, Liu B, Rao J, Dai W, Li X, Dai M, Yang N, Wang Y. D, APE1 may influence CD4 + naïve T cells on recurrence free survival in early stage NSCLC. *BMC Cancer.* 2021;21(1):233.
53. Zeng F, Liu LG, Zhang X, Huang K, Jiang H, Zhang T Y, *Plasminogen Activator Urokinase Receptor Implies Immunosuppressive Features and Acts as an Unfavorable Prognostic Biomarker in Glioma.* *Oncologist,* 2021.
54. Tamminga M, Schuurin HT, Timens E, Fehrmann W, Groen RS. HJ, Immune microenvironment composition in non-small cell lung cancer and its association with survival. *Clin Transl Immunology.* 2020;9(6):e1142.

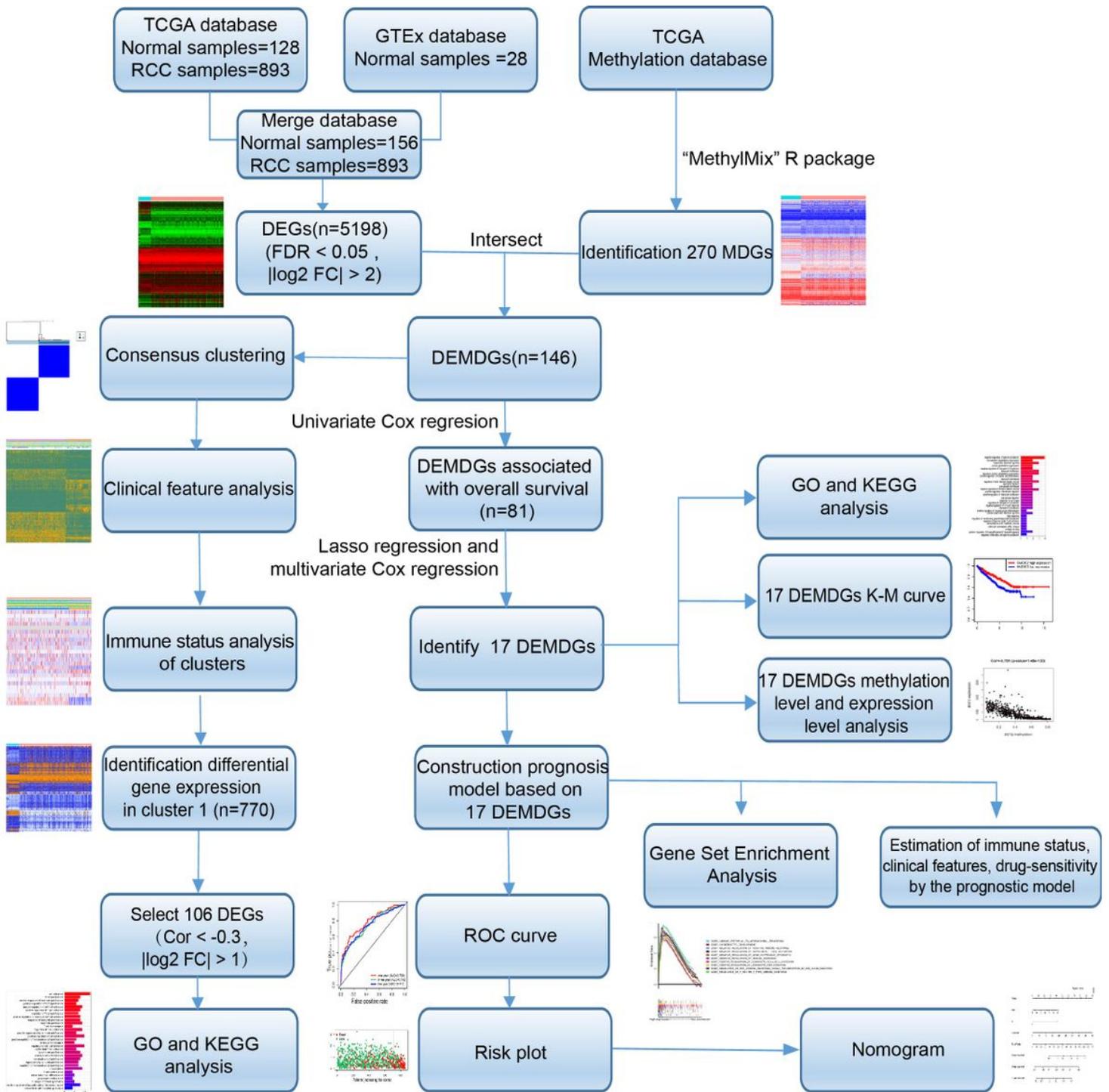
## Tables

**Table 1.** Multivariate Cox regression analysis of 17DEMDGs

gene	coef	HR	HR.95L	HR.95H	P value
SH3BGR	-0.089591804	0.914304325	0.835447896	1.00060387	0.05155323
BDH1	-0.066725946	0.93545153	0.87129257	1.004334934	0.065673923
MOB3A	-0.024448556	0.975847889	0.951130436	1.001207686	0.061795815
CX3CL1	-0.034306301	0.966275488	0.956947105	0.975694805	4.17E-12
NMI	0.051625706	1.052981544	0.9865486	1.123887999	0.120503784
NAPSA	-0.006786017	0.993236956	0.984133671	1.002424448	0.148596204
PPP1R18	0.037539918	1.038253441	1.015357262	1.061665927	0.000968562
TCF19	0.023148109	1.023418106	1.006399079	1.040724938	0.006820406
FMNL1	0.094605219	1.099224816	1.032025414	1.170799846	0.003288516
C1orf54	-0.034766087	0.96583131	0.937848268	0.994649296	0.020470177
FAXDC2	-0.035496821	0.965125802	0.921525969	1.010788459	0.132323501
BST2	0.003667575	1.003674309	1.00153526	1.005817926	0.000753672
MYH14	-0.040967121	0.959860689	0.922629315	0.998594481	0.042392429
ESRP2	-0.083628303	0.919773069	0.84971504	0.995607303	0.038557916
NCKAP1L	-0.190480134	0.826562178	0.772928142	0.883917917	2.63E-08
TRIM4	-0.084023915	0.919409268	0.872812021	0.968494226	0.001543821
EMP3	-0.013155528	0.986930627	0.979068285	0.994856108	0.001265433

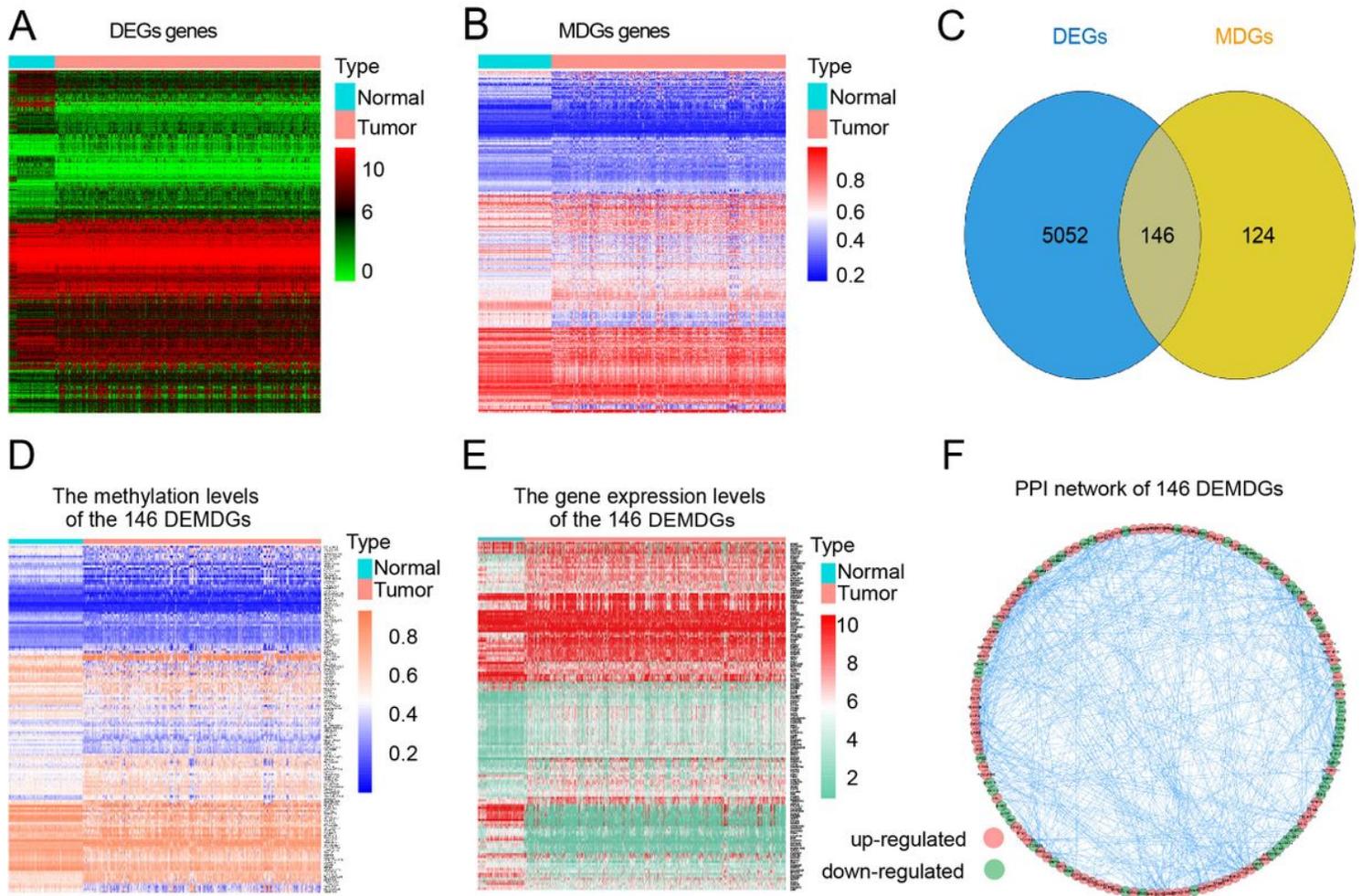
Coef, coefficient; HR, hazard ratio; CI, confidence interval.

## Figures



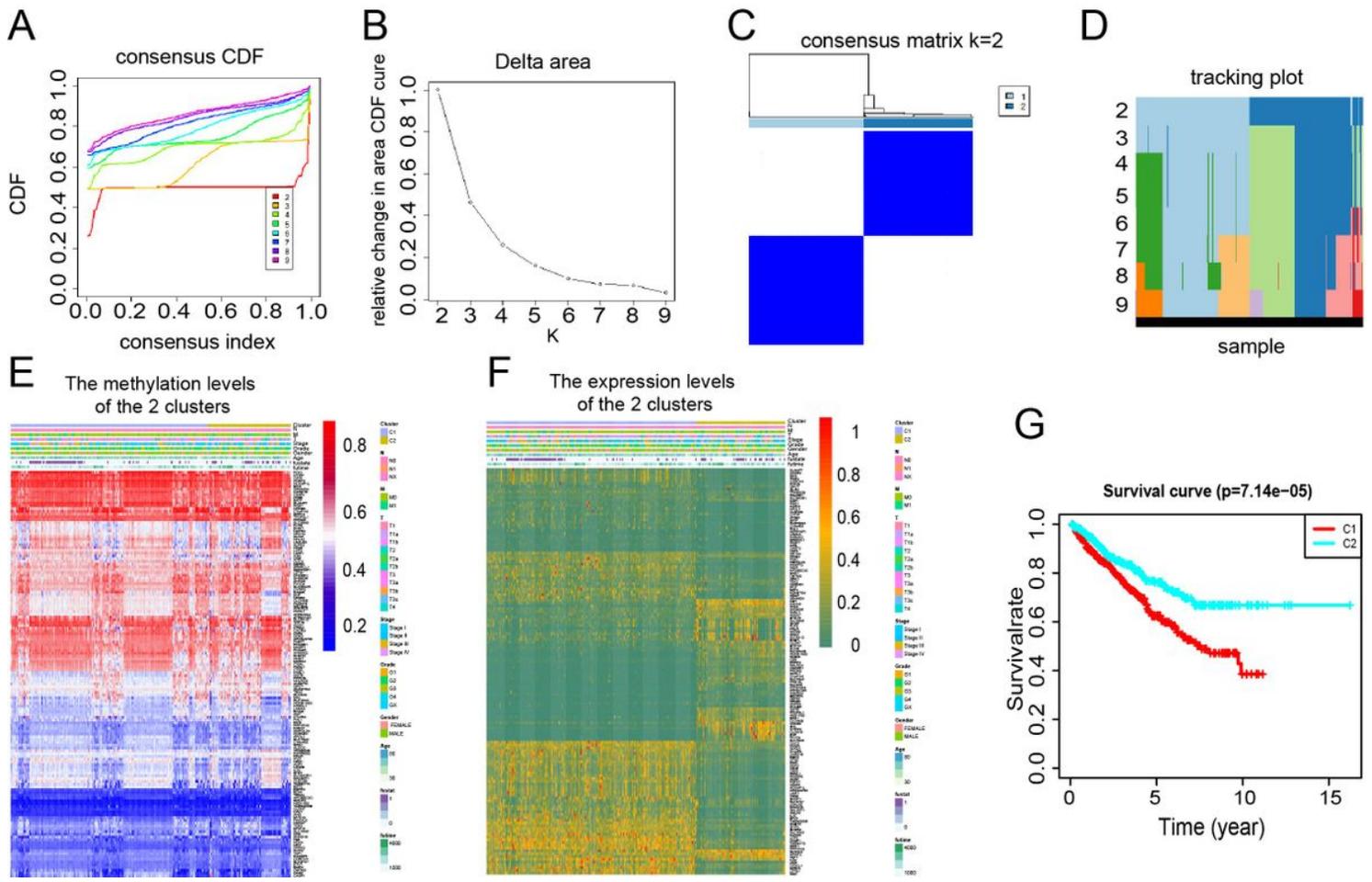
**Figure 1**

The flow diagram. TCGA, The Cancer Genome Atlas; GTEx, Genotype-Tissue Expression; DEGs, differentially expressed genes; MDGs, methylation-driven genes; DEMDGs, differentially expressed methylation-driven genes; KEGG, Kyoto Encyclopedia of Genes and Genomes; GO, Gene Ontology; K-M, Kaplan-Meier.



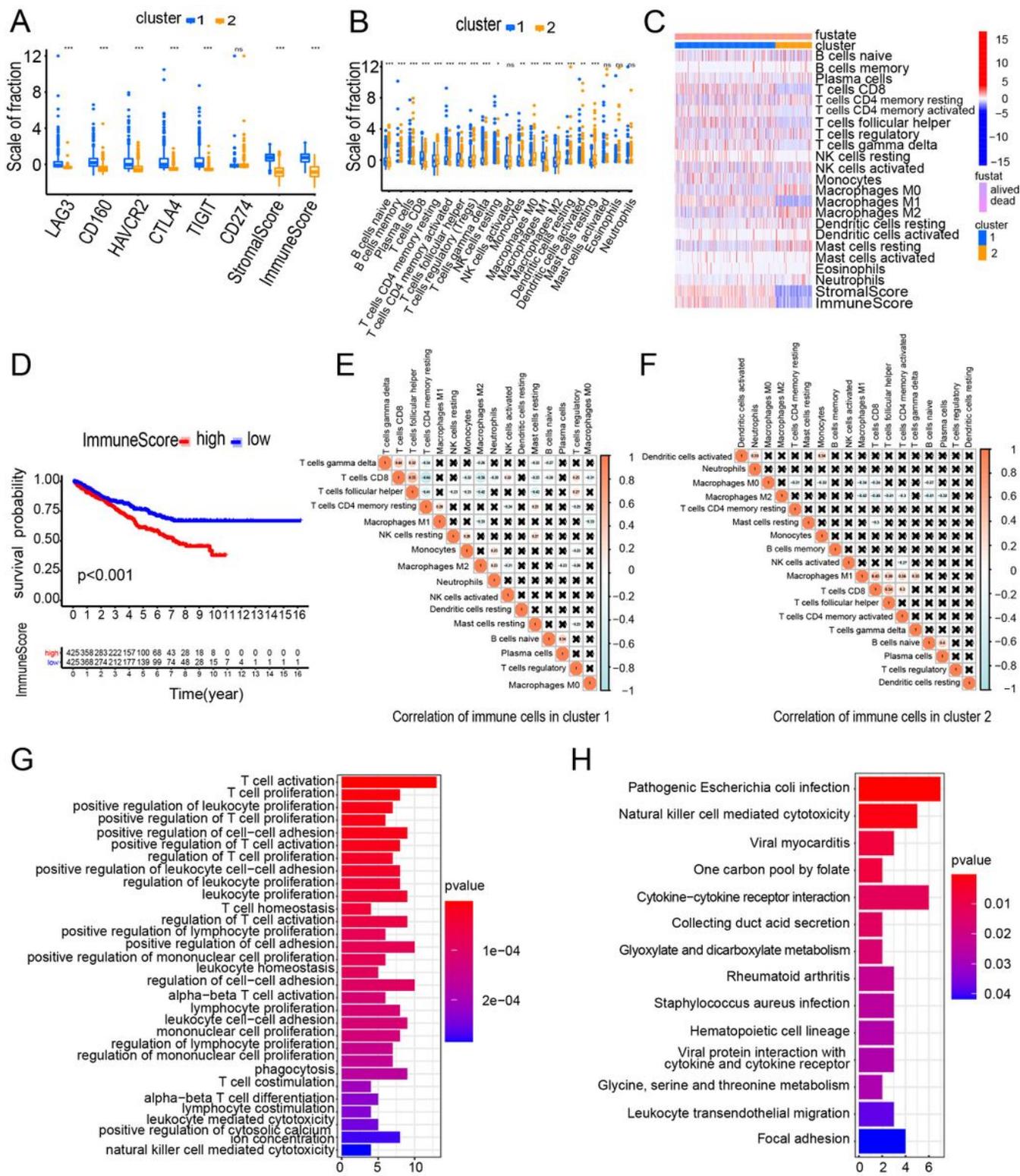
**Figure 2**

Screening of 146 DEMDGs. (A) Heatmap of DEGs in normal samples and RCC samples. (B) Heatmap of MDGs in normal samples and RCC samples. (C) Venn diagram for 146 DEMDGs in normal samples and RCC samples. The blue circle represents 5198 DEGs and the yellow circle represents 270 MDGs. (D) Heatmap of methylation levels of 146 DEMDGs. (E) Heatmap of gene expression levels of 146 DEMDGs. (F) PPI network of 146 DEMDGs.



**Figure 3**

Consensus clustering based on 146 DEMDGs. (A) CDF for RCC. (B) The area under the CDF curve in RCC. (C) Consensus clustering matrix for RCC at  $k = 2$ . (D) Tracking plot for  $k$  from 2 to 9. (E) Heatmap of methylation levels of 146 DEMDGs in the two clusters, and the distribution of clinical features were compared in the two clusters. (F) Heatmap of gene expression levels of 146 DEMDGs in the two clusters, and the distribution of clinical features were compared in the two clusters. (G) The survival curves showed significant prognostic differences in two clusters.



**Figure 4**

Immune status analysis of two clusters and functional enrichment analysis of 106 DEGs. (A) Expression of immune checkpoints in two clusters of RCC. (B) The abundance of immune cells in two clusters of RCC. (C) The heatmap of abundance of immune cells in two clusters. (D) The survival curves of high and low immune score group showed significant prognostic differences. (E) Correlation matrix of infiltrating immune cells in cluster 1. (F) Correlation matrix of infiltrating immune cells in cluster 2. The numbers in

the two matrices represent the Pearson correlation coefficient. Coral circle represents a positive correlation, blue circle represents a negative correlation, and white circle represent no correlation between two kinds of cells. (G) GO term enrichment in biological process of 106 DEGs. (H) KEGG enrichment analysis of 106 DEGs. \* $p < 0.05$ ; \*\* $p < 0.01$ ; \*\*\* $p < 0.001$ .

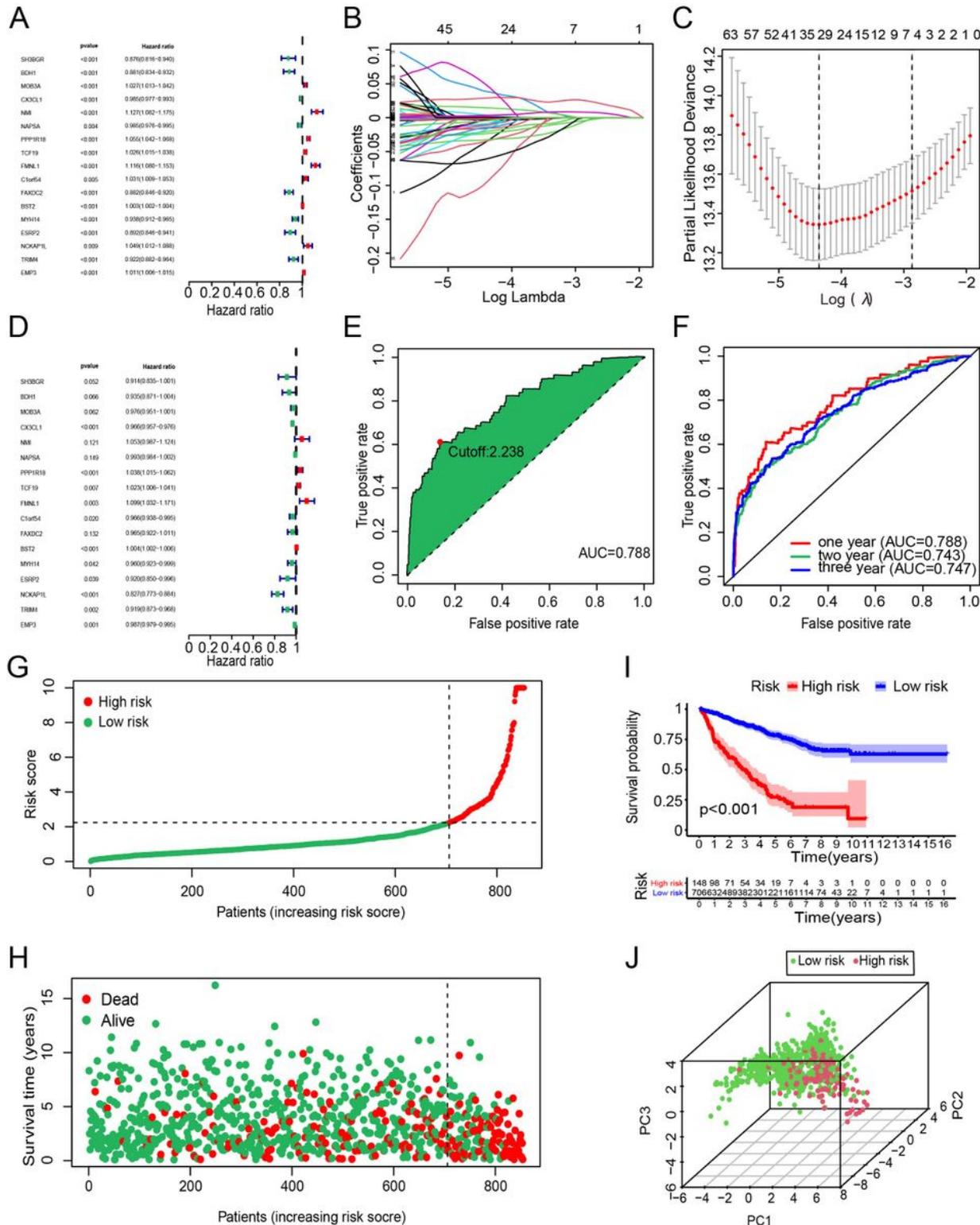
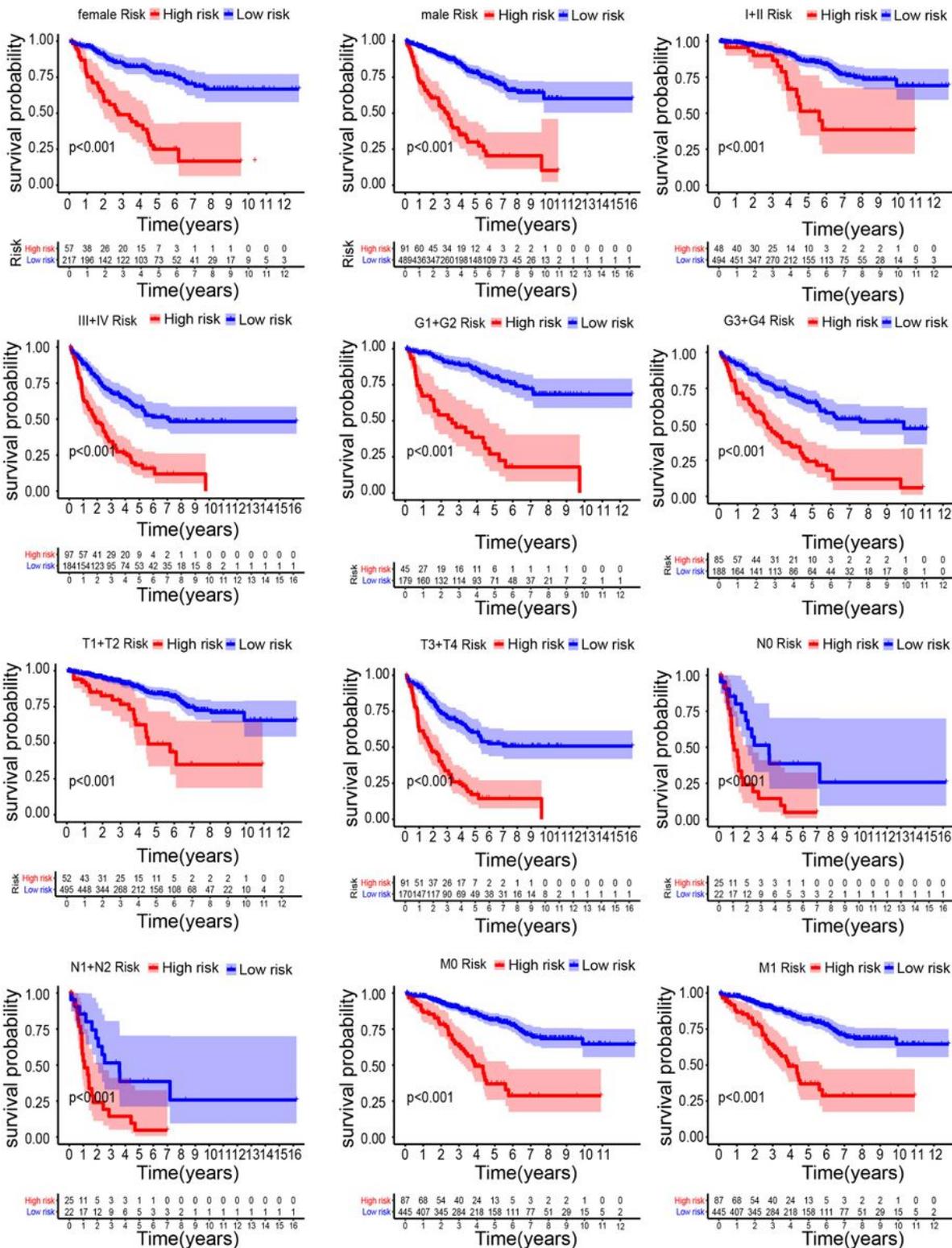


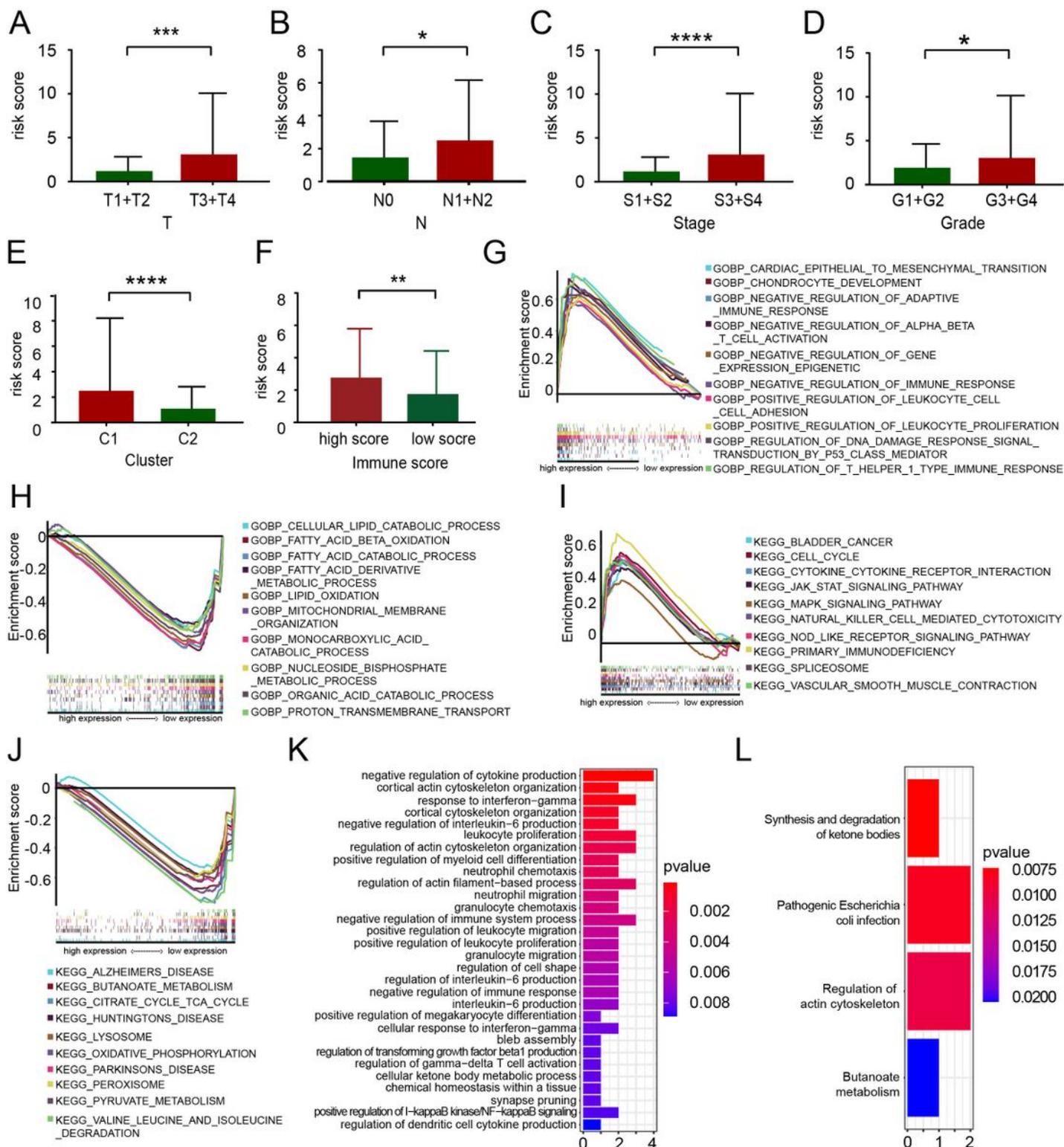
Figure 5

Establishment and validation of the prognostic model based on 146 DEMDGs. (A) The hazard ratios (HR) and 95% confidence intervals (CI) of 17 DEMDGs in RCC were computed by univariate Cox regression analysis. (B) The changing trajectory of each independent variable. (C) Confidence intervals for each optimal lambda. 10-fold cross-validation for tuning parameter selection in the LASSO model. (D) The HR and 95% CI of 17 DEMDGs in RCC were computed by multivariate Cox regression analysis. (E) Risk score for 893 RCC samples; the maximum inflection point is the cut-off point accessed by the AIC. (F) The 1-, 2-, and 3-year ROC of the optimal model suggested that all AUC values were over 0.7. (G-H) Distribution of the risk scores of RCC patients. (I) The survival curves were plotted to show the survival difference based on the risk score. (J) Principal component analysis performed on RCC samples based on risk scores.



**Figure 6**

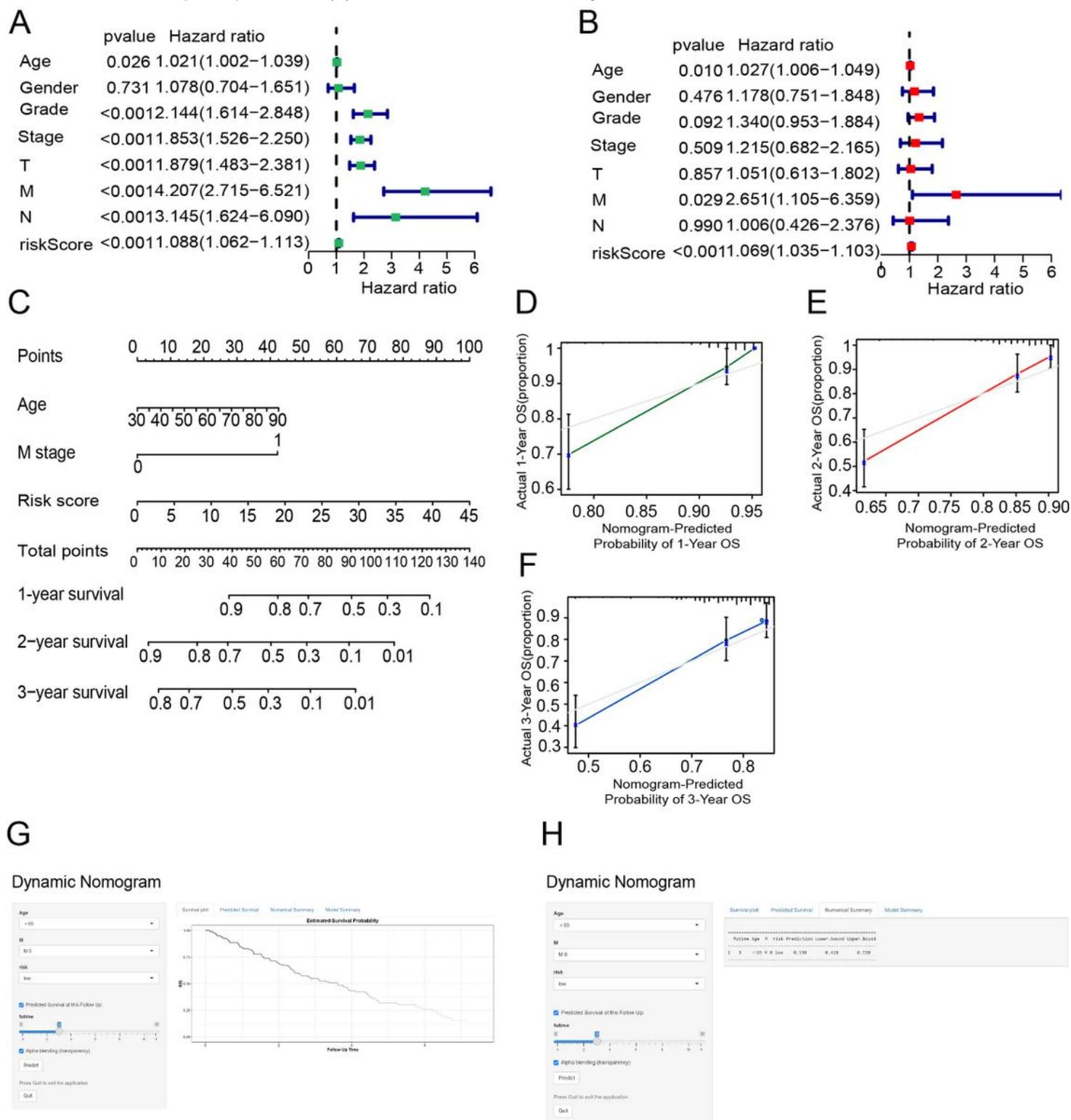
Survival curves for the high- and low-risk groups stratified by clinical features. Kaplan-Meier survival curve analysis showed the OS rates of high- and low-risk RCC patients stratified by female and male; clinical stages I-II and clinical stages III-IV; grades I-II and grades III-IV; T1-2 and T3-4; N0 and N1-2; M0 and M1-2. (T, tumor size; N, lymph node metastasis; M, metastasis).



**Figure 7**

Relationship between risk score and clinical features of RCC and functional enrichment analysis of prognostic model. (A-F) The distribution of risk score in high- and low-risk patients stratified according to (A) T stage (T1-2 vs. T3-4); (B) N stage (N0 vs. N1-2). (C) clinical stage (stages I-II vs. stages III-IV). (D) grade (grades I-II vs. grades III-IV). (E) cluster (cluster 1 vs. cluster 2). (F) immune score group (high score vs. low score). T, tumor size; N, lymph node metastasis. (G) Biological processes enriched in the high-risk

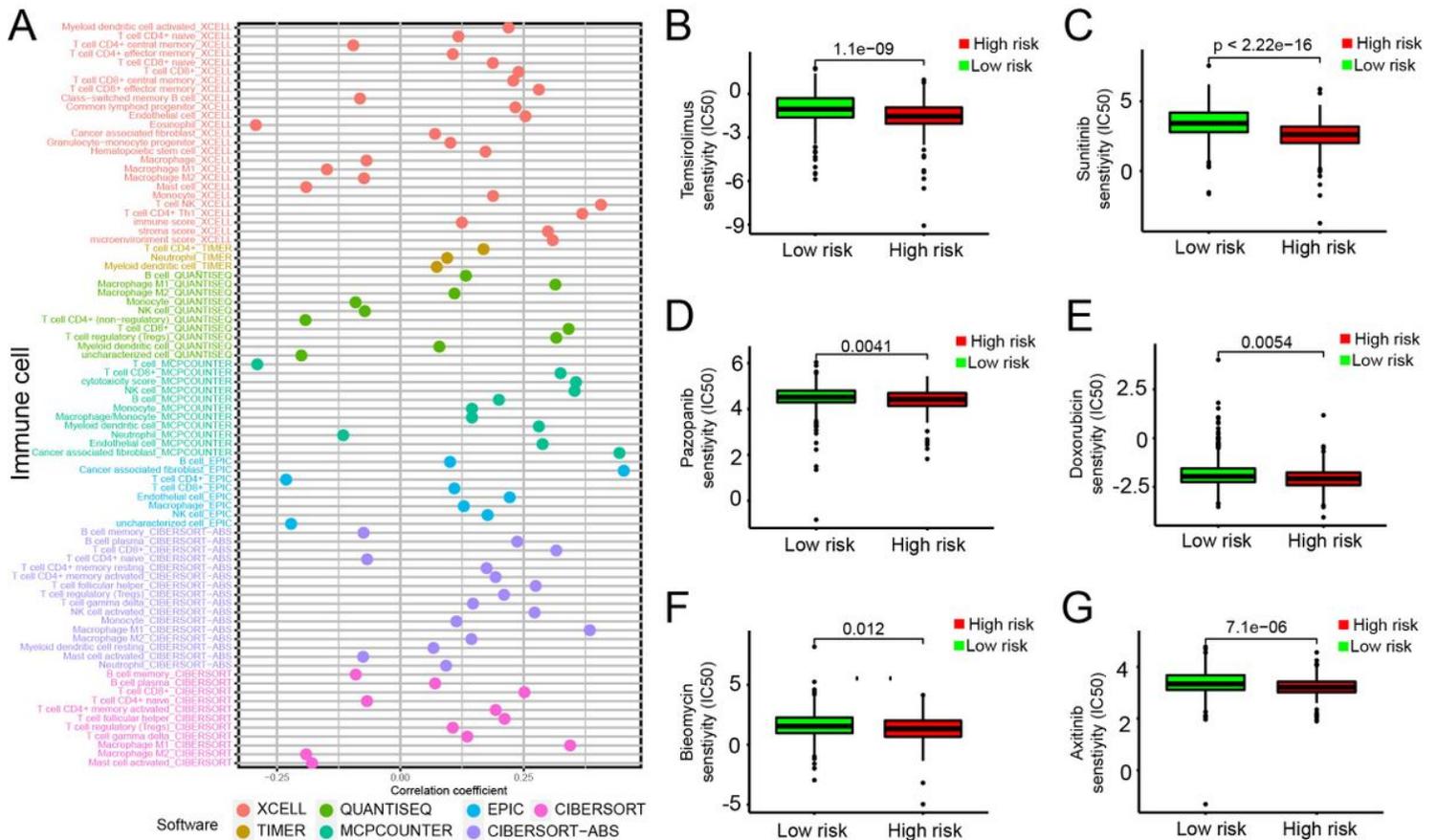
group. (H) Biological processes enriched in the low-risk group. (I) Signaling pathways enriched in the high-risk group. (J) Signaling pathways enriched in the low-risk group. (K) GO enrichment analysis of 17 DEMDGs in biological process. (L) KEGG enrichment analysis of 17 DEMDGs.



**Figure 8**

Assessment of the accuracy of the prognostic model. (A-B) Univariate (A) and multivariate (B) Cox regression analysis for the relation between risk score and clinical features. (C) Construction of the

nomogram model. (D-F) The calibration curves of 1-, 2-, 3-year in nomogram. (G) Patients with low-risk, M0 stage and aged < 65 years according to the web survival rate calculator (95% CI 42–72%). (H) 95% confidence interval according to the web survival rate calculator.



**Figure 9**

Estimation of tumor-infiltrating cells and drug susceptibility by the prognostic model. (A) Spearman correlation analysis between risk score and tumor-infiltrating immune cells. (B-G) The model had the potential to predict chemosensitivity because high-risk score was associated with a lower half maximal inhibitory concentration (IC50) of chemotherapeutics such as Axitinib ( $p = 7.1 \times 10^{-6}$ ), Bleomycin ( $p = 0.012$ ), Doxorubicin ( $p = 0.0054$ ), Pazopanib ( $p = 0.0041$ ), Sunitinib ( $p < 2.22 \times 10^{-16}$ ), Temsirrolimus ( $p = 1.1 \times 10^{-9}$ ).

## Supplementary Files

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

- [Fig.S1.jpg](#)
- [Fig.S2.jpg](#)
- [Fig.S3.jpg](#)
- [Fig.S4.jpg](#)

- [Fig.S5.jpg](#)
- [TableS1.xlsx](#)
- [TableS2.xlsx](#)
- [TableS3.xlsx](#)