

# Comprehensive analysis of DNA methylation and gene expression reveals specific predictive biomarkers for renal cell carcinoma

**Yuan Zhang**

Jilin University First Hospital

**Meng Xu**

Jilin University First Hospital

**Jianmei Yin**

Jilin University First Hospital

**Chunlv Li**

Jilin University First Hospital

**Zhuang Ye**

Jilin University First Hospital

**jing shang** (✉ [jyshangjing1987@163.com](mailto:jyshangjing1987@163.com))

the first of jilin university <https://orcid.org/0000-0001-6713-292X>

---

## Primary research

**Keywords:** methylation-driven genes (MDGs); prognosis; molecular classification; clear cell renal cell carcinoma (ccRCC).

**Posted Date:** April 7th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-20244/v1>

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

[Read Full License](#)

---

# Abstract

**Background** We aimed to identify methylation-driven genes (MDGs) and predict the prognosis of clear cell renal cell carcinoma (ccRCC) based on several cohorts with high-throughput data.

**Methods** The transcriptome profiles, 450K methylated data and corresponding clinical information were extracted from the cancer genome atlas (TCGA) database. MethylMix package was used to screen aberrant methylation events. Next, the Cox regression models were applied via survival package. Functional analysis was mainly performed based on ConsensusPathDB database. Besides, “mimifi” R package were adopted to analyze methylated alterations from three Gene Expression Omnibus (GEO) datasets. The predictive efficiency of constructed MDGs signature was assessed and validated in TCGA-KIRC and ICGC-RCC cohort, respectively. Unsupervised clustering analysis was conducted via ConsensusClusterPlus package. Moreover, MDGs-nomogram for OS prediction was conducted via glm and survival packages.

**Results** Totally, we collected We finally identified 761 samples from TCGA-KIRC, ICGC-RCC, GSE61441, GSE105260 and GSE105261. We combined the expression data and 450K methylated data to find 5 hub prognostic MDGs (TAGLN2, PDK2, HHLA2, HOXA2, XAF1). We used the TCGA-KIRC cohort as the training dataset to verify the superior predictive significance of the 5 MDGs signature (AUC = 0.713), and validated it in ICGC-RCC cohort with AUC = 0.769. Kaplan-Meier analysis suggested that patients with high MDGs levels suffered from worse survival outcomes. Besides, we further conducted the unsupervised clustering analysis in the whole ccRCC patients and identified the sub-cluster with the worst prognosis, indicating the MDGs could be used as efficient molecular classifier for ccRCC. We also identified 32 prognostic risk loci associated with hub MDGs in KIRC. Superior predictive efficiency was found in the MDGs-nomogram [Area Under Curve (AUC) of 3-year: 0.842, AUC of 5-year: 0.862], compared with traditional independent feature such as TNM stage (AUC of 3-year: 0.759, AUC of 5-year: 0.717). The 5 MDGs signature were mainly associated with oxygen metabolism, glycometabolism, even the HIF-1 signaling pathway. C

**Conclusions** Collectively, this study indicated several hub-MDGs and explored the prognostic value in ccRCC, which would provide new insights on the exploration of epigenetic pathogenesis and therapeutic targets for ccRCC.

## Background

Clear cell renal cell carcinoma (ccRCC) is the most common histologic subtype of renal cell carcinoma (RCC) defined by pathologic classification(1). Accounting for approximately 2-3% among all human malignant neoplasms, RCC is a serious challenge to urologists and oncologists globally(2). Unlike surgical intervention, the clinical efficacy of radiotherapy and chemotherapy are unsatisfactory(3). Despite the progression in immunotherapy in recent years, the prognosis of advanced RCC is far from

optimistic(4,5). Therefore, identification of biomarkers for the early diagnosis and treatment will be substantially beneficial for early surgical intervention and improvement of prognosis.

Studies have shown that the tumorigenesis and development of ccRCC is a complex process driven by multiple genes and follows multiple steps(6-8). However, the molecular regulation mechanism of ccRCC has not been elucidated. As one of the most important epigenetic factors, DNA methylation is extensively participated in the regulation progress of a variety of malignancies, including lung cancer, hepatic carcinoma, breast cancer and ccRCC(9-12). Aberrant methylation genes in CpG islands, which located in or near promoter regions of the genome, are often manifest as hypermethylation and eventually lead to silencing of tumor suppressor genes(13). In ccRCC, Li et al. revealed that CLDN7 which appears as a major constituent of tight junctions in cell-to-cell adhesion, is frequently downregulated via hypermethylation of its promoter(14). Besides, Schelleckes et al. identified two separated methylation-sensitive CpG islands in independent regions in promoter of tumor suppressor KIBRA, which was inhibited caused by promoter methylation(15). However, due to the heterogeneity of the disease itself and the large number of genetic and molecular regulatory networks involved, the regulation of DNA methylation in ccRCC has not been clarified.

In recent years, the interaction between methylation and gene expression has attracted more and more attention. The wide DNA methylation arrays, advent of deep RNA-Seq approach and emergence of new statistical algorithms promoted the rapid development of bioinformatics in malignant tumor research. An original computational algorithm called MethyImix invented by Olivier Gevaert et al. was applied in correlation analysis between abnormality of DNA methylation and predict transcription(16). In addition, public databases contain cancer genomic profiles are also facilitating the process of bioinformatics study. The Cancer Genome Atlas (TCGA) is a database with a view to catalogue and explore major cancer-causing genomic alterations(17). With a wide variety of online resources for biological information and data provided by the National Center for Biotechnology Information (NCBI), the utilization of Gene Expression Omnibus (GEO) database provides a reliable data foundation for bioinformatics analysis(18).

In this study, we analyzed and screened methylation-driven genes (MDGs) extracted from gene methylation profiling combined with gene expression data. The analyzed results from TCGA and GEO database were integrated and hub MDGs were filtrated by using bioinformatics tools. Finally, we explored the prognosis value of selected MDGs in ccRCC.

## **2. Materials And Methods**

### **2.1 Data acquisition and processing**

We obtained mRNA expression profiles and corresponding clinical information from TCGA (<https://cancergenome.nih.gov/>) database, including 541 tumors and 70 normal samples. Normalized fragments per kilobase of exon per million reads mapped (FPKM) values of the mRNA-seq data were

selected for subsequent analysis. Methylation profiles were assessed based on the Illumina Infinium HumanMethylation450 platform, which could cover nearly 482,421 CpG sites throughout the human genome. We utilized the TCGA-Assembler to download the beta values, which represent the average methylation level for all CpG sites related with one gene. Besides, clinical characteristics including age, gender, radiation or pharmaceutical therapy, tumor grade, TNM stage, pathologic status and survival information of each ccRCC patient were available from the TCGA data portal.

## 2.2 Methylmix Package

Since we have prepared the files of methylation and normalized expression data for tumors and normal tissues, we conducted the integrative analysis with the bioconductor MethylMix package to investigate potential correlations between methylation events and gene expression. MethylMix has been shown to produce unique findings apart from results consistent with other methods of integrative analysis<sup>16</sup>. The concordance of these complementary “omics” may facilitate revealing relevant information(19). Firstly, we considered the methylation events that correlated with alterations of gene expression only when genes reached the correlation filter cutoff. Secondly, a methylation state based on the large number of patients was defined using a beta mixture model by which the need for arbitrary cutoff could be precluded. Thirdly, the DNA methylation states were compared in tumor versus normal samples using the Wilcoxon rank sum test.

## 2.3 Pathway Analysis Of The Methylation-driven Genes

ConsensusPathDB database consists of 17 public resources for interactions, including genetic, metabolic, signaling, gene regulatory and drug-target interactions, as well as biochemical pathways(20). In our study, we selected the following sub-databases for analysis: Kegg, Inoh, Wikipathways, Signalink, Biocarta, Netpath, Humancyc, Smpdb, Pharmgkb, Ehm, and Smpdb. We used the default settings: minimum overlap and p value cutoff of 0.01. A list of 175 methylation-driven genes found to be statistically significant by MethylMix were conducted the pathway analysis and we set the cutoff with *P* value of 0.01.

## 2.4 Identification Of Hub Prognostic Mdgs

Integrating the corresponding clinical information of 319 ccRCC patients, we utilized univariate Cox regression model to screen survival associated MDGs from a list of 175 genes. Then, a total of 52 prognostic MDGs were chosen with  $P < 0.01$ . Besides, we continued stepwise genes selection procedure and got several related models. We calculated the Akaike information criterions (AICs) of all models and an optimized model including 38 genes with the smallest AIC was selected finally. To further obtain and validate the hub MDGs, we download methylation data and transcriptome profiles from the GEO database. GSE61441 and GSE105260 were used for analysis of methylation profiles based on the

GPL13534 platform (Illumina HumanMethylation450 BeadChip), while gene expression data of GSE105261 was performed using GPL10558 platform (Illumina HumanHT-12 V4.0 expression beadchip). There were 46 paired ccRCC tumor and adjacent normal tissues in GSE61441 and GSE105260 consisted of 35 ccRCC and 9 normal tissues. The GSE105161 series included 35 ccRCC and 9 normal tissues. For gene methylation profiling of GSE61441 and GSE105260, we utilized the “mimifi” R package to identify genes differentially methylated between tumor and normal tissues with cutoff of  $q < 0.05$  and  $|\text{Fold Change}| \geq 1$ . The differential genes analysis of GSE105261 were conducted using “limma” package and  $P \leq 0.05$  with  $|\text{Fold Change}| \geq 1$  was set as the thresholds. Afterwards, we obtained hypomethylation and upregulated genes via overlapping hypomethylated and high-expressed genes from GSE61441, GSE105260, GSE105261 and the 38 prognostic MDGs from TCGA cohort. Similarly, the hypermethylation with down expression genes were searched through the same process across four gene set. At last, we successfully identified 5 hub MDGs from TCGA and GEO databases that related with survival outcomes.

## 2.5 Construction Of Mdgs Signature And Nomogram For Os Prediction

Since we have already identified five prognostic MDGs from the above procedures, we conducted the multivariate Cox analysis for the five MDGs and obtained the respective coefficients ( $\beta_i$ ). Then, we constructed the methylation-related risk score with the following formula:  $\text{risk score} = (\beta_i * \text{Exp}_i)$ . According to the risk results, all 319 ccRCC patients were divided into high- and low-risk group with the median risk value as the threshold. We evaluated the predictive value of risk score via receiver operating characteristic (ROC) curve using “timeROC” package(21). and Kaplan-Meier analysis were drawn to compare the difference between high- and low-groups. Moreover, we integrated MDGs signature with clinical data to establish a comprehensive nomogram, in which only the variables that reaching statistical significance from univariate and multivariate Cox analysis can be selected into the final model. Meanwhile, the calibration graphs of 3-, 5-year OS prediction was drawn to assess the performance of nomogram. We used “rms” package to exhibited the plots. What is more, ROC curves for 3-, 5-year OS prediction of ccRCC patients were illustrated to compare the predictive accuracy between MDGs signature-based nomogram and traditional clinical factors of TNM stage.

## 2.6 Analysis of genes and methylated loci associated with OS, joint survival analysis

To further detect the specific methylated sites in the 5 hub MGDs, we extracted the all methylation data in 440 files from TCGA using the Perl scripts. We merged the value of methylated sites into one matrix and conducted univariate Cox analysis to select potential risk methylated loci. Survival analysis was then performed to assess the difference between hypo- and hypermethylation state of specific sites. Additionally, a joint survival analysis was conducted to further evaluate the vital MDGs signature

associated with prognosis in KIRC patients, where we combined the the methylation levels and expression data of one gene. The conjoint analysis was also performed by survival package.

## 2.7 Meta-analysis Based On Oncomine Database

The expression levels of five MDGs signature were further validated in Oncomine database (<https://www.oncomine.org/resource/main.html>). We compared the median rank of a gene across 6 analyses with the cutoff of  $P$  value  $\leq 0.05$  (FigureS2).

## 2.8 Gene Set Enrichment Analysis

The gene set enrichment analysis (GSEA)(22) was conducted to further uncover the potential pathways that enriched statistically significant between low- and high-risk phenotypes. We chose the “c2.cp.kegg.v6.2.symbols.gmt gene sets” for enrichment analysis downloaded from the MSigDB. Besides, we considered the significant gene sets with false discovery rate (FDR)  $\leq 0.05$ . All the procedures were based on the JAVA8 platform (<http://software.broadinstitute.org/gsea/index.jsp>). Moreover, Co-expression analysis for the 5 MDGs signature was conducted and we selected top 6 pathways to show by dotplot via “clusterProfiler” package.

## 2.9 Statistical Analysis

Identification of hub MDGs signature was achieved mainly by “MethylMix”, “minifi” and “limma” packages(23). The correlation of gene methylation state with expression level was evaluated by Spearman’s correlation. Cox regression analysis or Kaplan-Merier curves were performed based on “survival” package. Establishment of nomogram or calibration graphs were conducted via “rms” package. The heatmaps were generated by “pheatmap” package(24). and ROC curves were drawn by “timeROC” package. The meta-analysis results from Oncomine database were exhibited with  $P$  values and median ranks. All statistical analysis was conducted based on R studio (version 3.5.3).

## 3. Results

### 3.1 Identification of methylation-driven genes (MDGs) in RCC

We collected 485 samples with 450K methylation data from TCGA cohort, including 319 ccRCC samples and 166 precancerous samples. Transcriptome files were obtained from 611 samples, including 541 tumors and 70 normal samples. We then extracted and normalized the methylation and gene expression information of ccRCC using the “limma” package. MethyMix was utilized to conduct correlation analysis, where a mixed model was constructed and Wilcoxon rank test was calculated for identifying differential

methylation with  $|\logFC| > 0$ ,  $P \leq 0.05$ ,  $|\text{Cor}| > 0.3$ . A total of 175 driven genes were selected via the screening procedure (Additional Table S1). We illustrated the top hypermethylated and top hypomethylated genes in Fig. 1.

## 3.2 Functional Pathway Analysis Of 175 Mdgs

Firstly, we exhibited the clustering heatmap of 175 MDGs in tumors and normal tissues using “pheatmap” package in Fig. 2A. To further investigate the molecular mechanisms and underlying biological processes that the drivers may participate in, we conducted the functional pathway analysis based on ConsensusPathDB database. The results revealed that these 175 driven genes were mainly involved in oxygen metabolism process, one carbon metabolism, inflammatory response pathway, cell communication, as well as glucose metabolism pathway (Fig. 2B, Table 2).

Table 1  
Clinical characteristics of all 618 ccRCC patients included in  
this study.

<b>Variables</b>	<b>TCGA</b>	<b>ICGC</b>
	<b>(N = 537)</b>	<b>(N = 91)</b>
<b>Age (Mean ± SD)</b>	<b>60.59 ± 12.14</b>	<b>60.47 ± 9.97</b>
<b>Follow-up (y)</b>	<b>3.12 ± 2.23</b>	<b>4.14 ± 1.73</b>
<b>Status</b>		
<b>Alive</b>	<b>367 (68.34)</b>	<b>61 (67.03)</b>
<b>Dead</b>	<b>170 (31.66)</b>	<b>30 (32.97)</b>
<b>Gender</b>		
<b>Male</b>	<b>346 (64.43)</b>	<b>52 (57.14)</b>
<b>Female</b>	<b>191 (35.57)</b>	<b>39 (42.86)</b>
<b>AJCC-T</b>		
<b>T1</b>	<b>275 (51.21)</b>	<b>54 (59.34)</b>
<b>T2</b>	<b>69 (12.85)</b>	<b>13 (14.28)</b>
<b>T3</b>	<b>182 (33.89)</b>	<b>22 (24.18)</b>
<b>T4</b>	<b>11 (2.05)</b>	<b>2 (2.20)</b>
<b>AJCC-N</b>		
<b>N0</b>	<b>240 (44.69)</b>	<b>79 (86.81)</b>
<b>N1</b>	<b>17 (3.17)</b>	<b>2 (2.20)</b>
<b>Unknow</b>	<b>280 (52.14)</b>	<b>10 (10.99)</b>
<b>AJCC-M</b>		
<b>M0</b>	<b>426 (79.33)</b>	<b>81 (89.01)</b>

<b>M1</b>	<b>79 (14.71)</b>	<b>9 (9.89)</b>
<b>Unknow</b>	<b>32 (5.96)</b>	<b>1 (1.10)</b>
<b>Pathological stage</b>		
<b>I</b>	<b>269 (50.09)</b>	<b>-</b>
<b>II</b>	<b>57 (10.61)</b>	<b>-</b>
<b>III</b>	<b>125 (23.28)</b>	<b>-</b>
<b>IV</b>	<b>83 (15.46)</b>	<b>-</b>
<b>Unknow</b>	<b>3 (0.56)</b>	<b>-</b>
<b>Grade</b>		
<b>G1</b>	<b>14 (2.61)</b>	<b>-</b>
<b>G2</b>	<b>230 (42.83)</b>	<b>-</b>
<b>G3</b>	<b>207 (38.54)</b>	<b>-</b>
<b>G4</b>	<b>78 (14.53)</b>	<b>-</b>
<b>Unknow</b>	<b>8 (1.49)</b>	<b>-</b>

Data are shown as n (%).

Table 2  
Functional analysis from ConsensusPathDB database of methylated-driven genes

Enriched pathway	p-value	q-value	source	external_id
Biological oxidations	0.006187155	0.0783234	Reactome	R-HSA-211859
One Carbon Metabolism	0.00238886	0.0783234	Wikipathways	WP241
Inflammatory Response Pathway	0.003841911	0.0783234	Wikipathways	WP453
Tight junction	0.005985322	0.0783234	KEGG	path:hsa04530
glucose Metabolism	0.000273727	0.02965375	Wikipathways	WP1495
Cell-Cell communication	0.006663174	0.0783234	Reactome	R-HSA-1500931
Cell-cell junction organization	0.003138643	0.0783234	Reactome	R-HSA-421270
Beta3 integrin cell surface interactions	0.008638311	0.0783234	PID	integrin3_pathway
Generation of second messenger molecules	0.007099411	0.0783234	Reactome	R-HSA-202433
Tight junction interactions	0.000184328	0.02965375	Reactome	R-HSA-420029
Trans-sulfuration and one carbon metabolism	0.003209727	0.0783234	Wikipathways	WP2525
Glutathione synthesis and recycling	0.007746806	0.0783234	Reactome	R-HSA-174403
Nicotinate and Nicotinamide Metabolism	0.007746806	0.0783234	SMPDB	SMP00048
&gamma;-glutamyl cycle	0.005689606	0.0783234	HumanCyc	PWY-4041
O <sub>2</sub> /CO <sub>2</sub> exchange in erythrocytes	0.006681952	0.0783234	Reactome	R-HSA-1480926
sarcosine oncometabolite pathway	0.006681952	0.0783234	SMPDB	SMP02313
IRF3-mediated induction of type I IFN	0.006681952	0.0783234	Reactome	R-HSA-3270619
Nicotinamide salvaging	0.005689606	0.0783234	Reactome	R-HSA-197264
folate transformations	0.004771254	0.0783234	HumanCyc	PWY-2201
superpathway of choline degradation to L-serine	0.001868254	0.0783234	HumanCyc	PWY66-414
glycine betaine degradation	0.000900954	0.073202476	HumanCyc	PWY-3661-1

### 3.3 Construction of a prognostic MDGs model and molecular-classification of patients based on the signature

Since we have extracted the 175 drivers in 485 samples, we utilized univariate Cox regression method to select 52 prognostic epi-drivers that were significantly associated with OS ( $P \leq 0.01$ ). The stepwise selection methods in multivariate Cox model revealed that 38 MDGs that were screened out with the most robust combinations. To further validate the MDGs and simplify variables, we integrated 450K methylation data and transcriptome profiles to conduct correlation analysis in GSE61441, GSE105260 and GSE105261. With thresholds of  $FDR \leq 0.05$  and  $|\text{Fold Change}| \geq 1$  to deal with methylation data, we detected a list of 59 hypomethylated and upregulated genes by overlapping 145 hypomethylated genes (1049 in GSE61441 and 1102 in GSE105260) and 619 upregulated genes in GSE105261. Meanwhile, a total of 155 hypermethylated and downregulated genes were screened out by overlapping 156 hypermethylated genes (734 in GSE61441 and 897 in GSE105260) and 639 downregulated genes in GSE105261. Therefore, we finally identified 5 hub prognostic MDGs (TAGLN2, PDK2, HHLA2, HOXA2, XAF1) by integrating across three GSE series and TCGA cohort (Fig. 4). Besides, we constructed a representative volcano plot for GSE105261 in Fig. 4C. We conducted multivariate Cox regression model to obtain respective coefficients ( $\beta_i$ ), where  $\beta_i$  represented the weight of each variables. Then, five genes can be calculated as a prognostic index with the following formula:  $(0.64916 * \text{expression value of TAGLN2} - 0.17096 * \text{expression value of PDK2} - 0.18242 * \text{expression value of HHLA2} + 0.64882 * \text{expression value of HOXA2} + 0.43195 * \text{expression value of XAF1})$ . With the median risk value as the cutoff, 319 ccRCC in the methylated data samples that matched the clinical information were divided into low- ( $n = 159$ ) and high-risk group ( $n = 158$ ). The distribution of vital status of ccRCC patients in two groups were shown in Fig. 4A-B, and the heatmap of the 5 MDGs was exhibited in Fig. 4C. To assess the predictive value of the risk signature, ROC curve was drawn and the AUC was 0.713 for 3-year OS prediction (Fig. 4D). Kaplan-Meier survival curve analysis of patients revealed that the prognosis was poor in high-risk group, and the difference across the two groups showed statistically significant with log-rank test of  $P \leq 0.001$  (Fig. 4E). Furthermore, we utilized another ICGC-RCC cohort as a validation dataset and we found the AUC was 0.769 (Fig. 4F). Patients with high risk suffered from worse clinical outcomes compared with that in low-risk group (Fig. 4J). We thus summarized that the risk model based on 5 MDGs signature has certain accuracy in assessing the prognosis of ccRCC patients.

Given that the 5 MDGs signature was proved to be well associated with prognosis, we considered whether the 5 genes could be utilized as the molecular markers for guiding classifications of patients. We totally collected 618 patients from the TCGA-KIRC and ICGC-RCC cohort and normalized the sequencing data for following analysis. Besides, we used the ConsensusClusterPlus package and iterated 1000 times ( $K = 1: 10$ ) to stabilize the classification categories, obtaining the optimized classification of the samples (Fig. 5A and 5B). We observed that the classifications reached the best stable status with  $k = 5$  in Fig. 5C. Kaplan-Meier analysis suggested that MDGcluster classifications correlated significantly with survival outcomes when  $k = 5$ , among which the MDGcluster4 ( $N = 35$ ) has the best prognosis, and MDGcluster5 ( $N = 105$ ) has the worst prognosis (Fig. 5D). Sankey diagram also illustrated the underlying associations

among MDGcluster, risk score levels and vital outcomes (Fig. 5E). Given the differential survival outcomes between MDGcluster4 and MDGcluster5, we intended to figure out the underlying mechanisms between two groups. We performed the differential analysis between MDGcluster4 and MDGcluster5 to identify totally 162 DEGs with FDR < 0.001 (Table S2). Meanwhile, GO functional analysis also revealed that the top 400 genes were mainly enriched in mitochondrion organization, notch developmental processes and cell cycle biological items (Fig. 5F). The GSEA analysis also indicated that the notch signaling pathway, wnt signaling pathway and EMT pathway were significantly up-regulated in MDGcluster5 than MDGcluster4, partially contributing to the poor prognosis of MDGcluster5 (Fig. 5J).

### 3.4 Establishment Of A Comprehensive Nomogram

To figure out that whether the MDGs risk score could be utilized as an independent predictor, we combined with other clinical risk variables associated with prognosis based on the univariate Cox regression results, including age, tumor grade, pharmaceutical therapy, as well as pathological stages. Besides, we conducted the multivariate Cox regression method to integrate clinical features with the MDGs signature. Based on the multi-variate analysis consisting of the above variables, the risk score maintained a significant and independent factor for OS prediction ( $P=0.001122$ ) in the TCGA cohort (Table 3). Hence, a MDGs-nomogram was accordingly constructed to predict the 1-, 3-, 5-year OS and the calibration graphs for the 3-, 5-year OS rate revealed the well curve fitting between observed outcomes and predicted survival (Fig. 6A-B). In addition, our integrated MDGs-nomogram showed the superior predictive efficiency (AUC of 3-year: 0.842, AUC of 5-year: 0.862), compared with traditional independent variables such as TNM stage (AUC of 3-year: 0.759, AUC of 5-year: 0.717) in Fig. 6D.

Table 3

Univariate and multivariate Cox regression analysis of clinical characteristics for 319 ccRCC patients.

<b>Variables</b>	<b>Univariate analysis</b>	<b>Multivariate analysis</b>		
	<b>Hazard ratio (95%CI)</b>	<b>P value</b>	<b>Hazard ratio (95%CI)</b>	<b>P value</b>
<b>Age, years</b>				
< 50	1.086 (0.5079–2.321)	0.832	1.2448 (0.578–2.682)	0.576
50–59	1.849 (0.911–3.753)	0.089	1.793 (0.873–3.683)	0.112
60–69	2.170 (1.054–4.469)	0.036	2.380 (1.142–4.959)	0.021
70–79	6.372 (2.743–14.8)	1.65e-05	4.961 (2.095–11.746)	0.000271
≥ 80				
<b>Gender</b>	1.073 (0.711–1.618)	0.738		
Female				
Male				
<b>Race</b>	4.020 (0.507–31.853)	0.188		
Black or Africa	1.373 (0.691–21.727)	0.366		
Asian				
White				
<b>Radiation</b>	1.266 (0.311–5.137)	0.742		
Yes				
No				
<b>Pharmaceutical</b>	0.1894 (0.128–0.2803)	<2e-16	0.361 (0.232–0.563)	7.03e-06
No				
Yes				
<b>Tumor grade</b>	3.3335 (2.063–5.387)	8.78e-07	1.686 (1.004–2.829)	0.048
Low				
High				
<b>Pathological stage</b>	4.417 (2.873–6.789)	1.27e-11	2.718 (1.686–4.380)	4.04e-05
Stage I&II				
Stage III&IV				
<b>AJCC-T</b>	3.08 (2.072–4.58)	2.72e-08		
T1&2				
T3&4				
<b>AJCC-N</b>	1.043 (0.705–1.543)	0.835		
N0				
N1&Nx				
<b>AJCC-M</b>	3.312 (2.229–4.921)	3.11e-09		
M0				
M1&Mx				
<b>Risk score</b>	3.285 (2.137–5.049)	5.85e-08	2.1927 (1.367–3.517)	0.001122
Low				
High				
Abbreviations: HR, hazard ratio; 95%CI, 95% confidence interval; AJCC, American Joint Committee on Cancer.				

## 3.5 Mapping of Kaplan-Meier curves of driver genes and joint survival analysis

We performed the prognostic survival analysis of 5 drivers based on the “survival” R package, where log-rank test of  $P \leq 0.05$  was used as a meaningful statistical value. We found methylation levels of 3 MDGs were associated with the prognosis of OS (TAGLN2, HHLA2, XAF1), while no significant difference was observed in PDK2 and HOXA2 (Fig. 7A-E). What is more, the joint survival analysis showed that combination of methylation and expression levels had a highly significant correlation with survival outcomes in ccRCC. The hypomethylation with high expression level of TAGLN2, HOXA2 and XAF1 correlated with lower survival rate, while the hypomethylation with high expression level of PDK2 and HHLA2 showed the higher survival rate (Fig. 7F-J).

Given that we have already identified the 5 hub MDGs associated highly with survival prognosis of KIRC, we intended to further detect the significant methylation loci harboring these hazard aberrant methylated drivers. We extracted the  $\beta$  value of methylation data according the 5 gene symbols from the 874 files and merged into a matrix. A list of 187 methylation sites were screened out and the univariate Cox analysis revealed that 32 key methylation loci were associated with survival outcomes in Table 4 ( $P \leq 0.05$ ).

## 3.6 Enriched pathways and co-expression analysis of the 5 MDGs signature

The GSEA was conducted to investigate potential pathways that associated with the 5 MDGs signature. The result revealed that there were six MDGs related pathways enriched significantly in high-risk group with  $P \leq 0.05$  and the top four were mainly involved in glycolysis, sucrose metabolism, P53 signaling pathways and cytochrome p450 crosstalk (Fig. 8A-D). Besides, we further extracted the five-MDGs signature related genes with filter cutoff value of correlation coefficient = 0.5,  $P = 0.001$ . A total of 70 genes were then selected and put into “DAVID 6.8” (<https://david-d.ncifcrf.gov/>)<sup>20</sup> to perform the functional enriched pathway analysis, including gluconeogenesis, fructose metabolism, HIF-1 signaling pathway, carbon metabolism, as well as central carbon metabolism in cancer (Fig. 8E). In accordance with GSEA results, the functional pathway analysis also indicated that the 5 MDGs signature were mainly associated with oxygen metabolism, glycometabolism, even the HIF-1 signaling pathway.

## 4. Discussion

Recent studies have revealed that the oncogenesis and progression of malignant tumors are inseparable from the extensive participation of epigenetic regulation mechanisms(26, 27). As one of the most common epigenetic regulatory mechanisms, DNA methylation may lead to genomic rearrangements and chromosomal instability, which in turn leads to malignant progression(28). In ccRCC, the methylation status of specific genes was associated with worse prognosis(29). Since microarrays and high-

throughput sequencing can simultaneously provide expression levels or modifications of thousands of genes and mRNAs in the human genome, they have been widely used for screening differentially expressed genes or abnormal RNA modifications in various malignancies. However, the risk relationship and specific regulatory mechanisms of DNA methylation and tumor progression have not yet been elucidated in ccRCC.

In this study, we revealed critical methylation events in the landscape of ccRCC by characterizing the genome-wide methylated-differentially expressed genes. Firstly, mRNA expression profiles and corresponding clinical information were extracted from the TCGA database. We conducted the integrative analysis with the bioconductor MethylMix package to investigate potential correlations between methylation events and gene expression. Next, the ConsensusPathDB database and the univariate Cox regression model were applied to analyze and screen out the prognosis-related MDGs. To further validate the above results, we used the 'mimif' R package to analyze chip data from ccRCC tumors and adjacent normal tissues from the GEO database and to search for differential genes. Finally, we integrated the results above and identified 5 hub MDGs (TAGLN2, PDK2, HHLA2, HOXA2 and XAF1) from TCGA and GEO databases and constructed MDGs signatures and nomograms for OS prediction. Meanwhile, the screened MDGs signature could be utilized as the well classifier for molecular discrimination and we successfully identified the prognostic MDGs cluster sub-group with worse clinical outcomes based on the signature. Survival analysis was performed based on Oncomine database and gene set enrichment analysis combined with meta-analysis of Co-expression analysis. Our results indicated that the 5 MDGs signature were mainly associated with oxygen metabolism, glycometabolism and HIF-1 signaling pathway.

TAGLN2 is a subtype of the TAGLN (Transgelin) family, which is class of actin actin-binding protein. The current literature reports that the expression level of TAGLN2 in tumors remains controversial. Several studies have found that TAGLN2 is up-regulated in malignant diseases by proteomic analysis(30–33). However, some studies demonstrated the opposite results in breast cancer and colon cancer through proteomic analysis(34). In our study, the hypomethylation with high expression level of TAGLN2 correlated with lower survival rate, consistent with the description of oncogene.

PDK2 is a member of the pyruvate dehydrogenase kinase (PDK) family. As a phosphatase inhibitor of pyruvate dehydrogenase complex (PDC), PDK enhances cell glycolysis and promotes tumor cell proliferation. Cui et al. revealed that overexpression of PDK2 reflects poor prognosis in acute myeloid leukemia(35). Besides, PDK2 may induces cisplatin-resistance in lung adenocarcinoma by transcriptional regulating cation transport mediator 3 (CNNM3)(36). However, we observed no significant difference between methylation levels of PDK2 and the prognosis of OS. This may be caused by heterogeneity of ccRCC.

HERV-HLTR-associating 2 (HHLA2) is an immune checkpoint belonging to the B7 family PMC(37). Study has shown that HHLA2 may inhibit T cell receptors mediated proliferation and cytokine production of CD4 + and CD8 + T cells(38). High expression level of HHLA2 were observed in non-small-cell lung cancer

(NSCLC) and osteosarcoma, and has association with advanced and metastatic diseases(39, 40). In PD-L1 negative NSCLC patients, HHLA2 were identified as a high-expression level(40). On the contrary, our results revealed that hypomethylation with high expression level HHLA2 showed the higher survival rate in ccRCC. The role of HHLA2 in ccRCC requires further investigation to explore its potential value as a novel immunotherapeutic target(39).

HOXA2 is a part of HOXA, which is a cluster of the human class I homeobox genes (HOX)(41). Recently, HOXA2 was firstly defined as an androgen-responsive gene and an oncogenic transcription factor in prostate cancer (PCa). In this study, Gao et.al discovered that the binding site of HOXA2 may altered by rs11672691, which was identified at 19q13 associated with aggressive PCa by Genome-wide association studies (GWAS)(42). In ccRCC, we found that the hypomethylation with high expression level of HOXA2 correlated with lower survival rate, while no significant difference was observed in methylation levels of HOXA2 associated with the prognosis of OS. The regulatory mechanism of HOXA2 in ccRCC progression needs further study.

X-linked inhibitor of apoptosis protein (XIAP)-associated factor 1 (XAF1) is identified as a pro-apoptotic protein and an antagonist of XIAP, counteracting anti-caspase activity of XIAP(43). It was demonstrated that XAF1 may promote cell apoptosis through in a multi-mechanism environment(44–47). Down-regulated or deleted of XAF1 was observed in many malignancies, which suggesting that low expression of XAF1 may promote malignant tumor progression by relieving tumor cell apoptosis inhibition (48–50). However, XAF1 was up-regulated in testicular germ cell tumors exceptionally, which may lead to high sensitivity to chemotherapy(51). Also, we found that hypomethylation with high expression level of XAF1 correlated with lower survival rate. It is suggested that XAF1 may participate in the ccRCC regulation process through some unconventional approaches.

Our functional pathway analysis indicated that such 5 MDGs signature were mainly associated with oxygen metabolism, glycometabolism, even the HIF-1 signaling pathway, which is unique in our study. In head and neck cancer (HNC), it was revealed that mitochondrial oxidation may be activated by PDK2 inhibition, and reverse cisplatin resistance(52). Another research focused on HNC demonstrated that, mitochondrial mutations contribute to HIF1- $\alpha$  accumulation via increased reactive oxygen species and up-regulated PDK2, which contribute to HNSCC development(53). Combined with our results, it is suggested that in ccRCC, PDK2 may regulate mitochondrial oxidation through methylation, and thus leading to tumor progression.

In addition, we summarized that the risk model based on 5 MDGs signature has certain accuracy in assessing the prognosis of ccRCC patients. Nomogram was integrated and showed that the superior predictive efficiency, compared with traditional independent feature such as TNM stage. The results suggested that our nomogram might serve as an accurate prognosis prediction for ccRCC.

Apart from the discovery of 5 MDGs and prognostic value of our nomogram, some limitations were also observed in ours research. Firstly, screening and verification were performed through bioinformatics approaches, while no other external clinical samples were used to validate the data. Secondly, the

underlying biological signaling pathways that the 5 hub MGDs may participate in warranted further functional experiments to validate. Moreover, selection bias might occur, which could reduce the credibility of our results. Finally, as mentioned above, these 5 selected MDGs should be further explored to illuminate the specific regulatory mechanism in ccRCC.

## Conclusion

In the current study, we explored and identified several hub-MDGs that could be an effective classifier and predictive factor in ccRCC using the multi-omics data across several datasets, providing new insights on epigenetic pathogenesis and therapeutic targets for ccRCC.

## Declarations

### Acknowledgements

Not applicable.

### Author Contributions

Zhuang Ye and Jing Shang designed the study and analyzed the data; Yuan Zhang and Meng Xu drafted the article; Jianmei Yin and Chunnv Li were responsible for language correction. All authors finally approved the paper.

### Funding

None

### *Availability of data and materials*

*Additional data not presented in the manuscript can be obtained by contacting the authors.*

### Ethics approval and consent to participant

Not applicable.

### Consent for publication

Not applicable.

### Competing interests

The authors declare that they have no competing interests.

## References

1. Frew IJ, Moch H. A clearer view of the molecular complexity of clear cell renal cell carcinoma. *Annu Rev Pathol* 2015;10:263-89.
2. Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2018;68:394-424.
3. Jonasch E, Gao J, Rathmell WK. Renal cell carcinoma. *BMJ* 2014;349:g4797.
4. Rijnders M, de Wit R, Boormans JL, et al. Systematic Review of Immune Checkpoint Inhibition in Urological Cancers. *Eur Urol* 2017;72:411-23.
5. Ljungberg B, Campbell SC, Choi HY, et al. The epidemiology of renal cell carcinoma. *Eur Urol* 2011;60:615-21.
6. Schodel J, Grampp S, Maher ER, et al. Hypoxia, Hypoxia-inducible Transcription Factors, and Renal Cancer. *Eur Urol* 2016;69:646-57.
7. Shenoy N, Pagliaro L. Sequential pathogenesis of metastatic VHL mutant clear cell renal cell carcinoma: putting it together with a translational perspective. *Ann Oncol* 2016;27:1685-95.
8. Rydzanicz M, Wrzesinski T, Bluysen HA, et al. Genomics and epigenomics of clear cell renal cell carcinoma: recent developments and potential applications. *Cancer Lett* 2013;341:111-26.
9. Koch A, Joosten SC, Feng Z, et al. Analysis of DNA methylation in cancer: location revisited. *Nat Rev Clin Oncol* 2018;15:459-66.
10. Kettunen E, Hernandez-Vargas H, Cros MP, et al. Asbestos-associated genome-wide DNA methylation changes in lung cancer. *Int J Cancer* 2017;141:2014-29.
11. Yuan SX, Zhang J, Xu QG, et al. Long noncoding RNA, the methylation of genomic elements and their emerging crosstalk in hepatocellular carcinoma. *Cancer Lett* 2016;379:239-44.
12. Deckers IA, van Engeland M, van den Brandt PA, et al. Promoter CpG island methylation in ion transport mechanisms and associated dietary intakes jointly influence the risk of clear-cell renal cell cancer. *Int J Epidemiol* 2017;46:622-31.
13. Wang Q, Wang G, Liu C, et al. Prognostic value of CpG island methylator phenotype among hepatocellular carcinoma patients: A systematic review and meta-analysis. *Int J Surg* 2018;54:92-9.
14. Li Y, Gong Y, Ning X, et al. Downregulation of CLDN7 due to promoter hypermethylation is associated with human clear cell renal cell carcinoma progression and poor prognosis. *J Exp Clin Cancer Res* 2018;37:276.
15. Schelleckes K, Schmitz B, Ciarimboli G, et al. Promoter methylation inhibits expression of tumor suppressor KIBRA in human clear cell renal cell carcinoma. *Clin Epigenetics* 2017;9:109.
16. Gevaert O, Tibshirani R, Plevritis SK. Pancancer analysis of DNA methylation-driven genes using MethylMix. *Genome Biol* 2015;16:17.
17. Tomczak K, Czerwinska P, Wiznerowicz M. The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Contemp Oncol (Pozn)* 2015;19:A68-77.

18. Coordinators NR. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* 2016;44:D7-19.
19. Rhee JK, Kim K, Chae H, et al. Integrated analysis of genome-wide DNA methylation and gene expression profiles in molecular subtypes of breast cancer. *Nucleic Acids Res* 2013;41:8464-74.
20. Kamburov A, Stelzl U, Lehrach H, et al. The ConsensusPathDB interaction database: 2013 update. *Nucleic Acids Res* 2013;41:D793-800.
21. Blanche P . timeROC: Time-Dependent ROC Curve and AUC for Censored Survival Data. 2015.
22. Tilford CA, Siemers NO. Gene set enrichment analysis. *Methods Mol Biol* 2009;563:99-121.
23. Smyth G K . limma: Linear Models for Microarray Data. *Bioinformatics & Computational Biology Solutions Using R & Bioconductor* 2011;397-420.
24. Kolde R. pheatmap: Pretty Heatmaps. 2015.
25. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 2009;4:44-57.
26. Ge Y, Chen S, Mu W, et al. Epigenetic regulation of UDP-Glucuronosyltransferase by microRNA-200a/-183: implications for responses to sorafenib treatment in patients with hepatocellular carcinoma. *Cancer Lett* 2019;454:14-25.
27. Farooqi AA, Fuentes-Mattei E, Fayyaz S, et al. Interplay between epigenetic abnormalities and deregulated expression of microRNAs in cancer. *Semin Cancer Biol* 2019.
28. Jeziorska DM, Murray RJS, De Gobbi M, et al. DNA methylation of intragenic CpG islands depends on their transcriptional activity during differentiation and disease. *Proc Natl Acad Sci U S A* 2017;114:E7526-E35.
29. Shenoy N, Vallumsetla N, Zou Y, et al. Role of DNA methylation in renal cell carcinoma. *J Hematol Oncol* 2015;8:88.
30. Zhang Y, Ye Y, Shen D, et al. Identification of transgelin-2 as a biomarker of colorectal cancer by laser capture microdissection and quantitative proteome analysis. *Cancer Sci* 2010;101:523-9.
31. Yoshino H, Chiyomaru T, Enokida H, et al. The tumour-suppressive function of miR-1 and miR-133a targeting TAGLN2 in bladder cancer. *Br J Cancer* 2011;104:808-18.
32. Jin H, Cheng X, Pei Y, et al. Identification and verification of transgelin-2 as a potential biomarker of tumor-derived lung-cancer endothelial cells by comparative proteomics. *J Proteomics* 2016;136:77-88.
33. Yakabe K, Murakami A, Kajimura T, et al. Functional significance of transgelin-2 in uterine cervical squamous cell carcinoma. *J Obstet Gynaecol Res* 2016;42:566-72.
34. Shields JM, Rogers-Graham K, Der CJ. Loss of transgelin in breast and colon tumors and in RIE-1 cells by Ras deregulation of gene expression through Raf-independent pathways. *J Biol Chem* 2002;277:9790-9.
35. Cui L, Cheng Z, Liu Y, et al. Overexpression of PDK2 and PDK3 reflects poor prognosis in acute myeloid leukemia. *Cancer Gene Ther* 2018.

36. Hu T, Yu S, Li Y, et al. PDK2 induces cisplatin-resistance in lung adenocarcinoma via transcriptional regulation of CNNM3. *J Drug Target* 2019;27:460-5.
37. Yan H, Qiu W, Koehne de Gonzalez AK, et al. HHLA2 is a novel immune checkpoint protein in pancreatic ductal adenocarcinoma and predicts post-surgical survival. *Cancer Lett* 2019;442:333-40.
38. Zhao R, Chinai JM, Buhl S, et al. HHLA2 is a member of the B7 family and inhibits human CD4 and CD8 T-cell function. *Proc Natl Acad Sci U S A* 2013;110:9879-84.
39. Koirala P, Roth ME, Gill J, et al. HHLA2, a member of the B7 family, is expressed in human osteosarcoma and is associated with metastases and worse survival. *Sci Rep* 2016;6:31154.
40. Cheng H, Janakiram M, Borczuk A, et al. HHLA2, a New Immune Checkpoint Member of the B7 Family, Is Widely Expressed in Human Lung Cancer and Associated with EGFR Mutational Status. *Clin Cancer Res* 2017;23:825-32.
41. Boncinelli E, Acampora D, Pannese M, et al. Organization of human class I homeobox genes. *Genome* 1989;31:745-56.
42. Gao P, Xia JH, Sipeky C, et al. Biology and Clinical Implications of the 19q13 Aggressive Prostate Cancer Susceptibility Locus. *Cell* 2018;174:576-89 e18.
43. Holcik M, Gibson H, Korneluk RG. XIAP: apoptotic brake and promising therapeutic target. *Apoptosis* 2001;6:253-61.
44. Steigerwald C, Rasenberger B, Christmann M, et al. Sensitization of colorectal cancer cells to irinotecan by the Survivin inhibitor LLP3 depends on XAF1 proficiency in the context of mutated p53. *Arch Toxicol* 2018;92:2645-8.
45. Wang TL, Ouyang CS, Lin LZ. beta-Asarone suppresses Wnt/beta-catenin signaling to reduce viability, inhibit migration/invasion/adhesion and induce mitochondria-related apoptosis in lung cancer cells. *Biomed Pharmacother* 2018;106:821-30.
46. Lee MG, Han J, Jeong SI, et al. XAF1 directs apoptotic switch of p53 signaling through activation of HIPK2 and ZNF313. *Proc Natl Acad Sci U S A* 2014;111:15532-7.
47. Jeong SI, Kim JW, Ko KP, et al. XAF1 forms a positive feedback loop with IRF-1 to drive apoptotic stress response and suppress tumorigenesis. *Cell Death Dis* 2018;9:806.
48. Law CT, Wei L, Tsang FH, et al. HELLS Regulates Chromatin Remodeling and Epigenetic Silencing of Multiple Tumor Suppressor Genes in Human Hepatocellular Carcinoma. *Hepatology* 2019;69:2013-30.
49. Schluckebier L, Aran V, De Moraes J, et al. XAF1 expression levels in a non-small cell lung cancer cohort and its potential association with carcinogenesis. *Oncol Rep* 2017;38:402-10.
50. Irvani O, Bay BH, Yip GW. Silencing HS6ST3 inhibits growth and progression of breast cancer cells through suppressing IGF1R and inducing XAF1. *Exp Cell Res* 2017;350:380-9.
51. Kempkensteffen C, Hinz S, Jager T, et al. [Expression levels of the IAP antagonists XAF1, Smac/DIABLO and HtrA2 in testicular germ cell tumours]. *Aktuelle Urol* 2008;39:436-41.

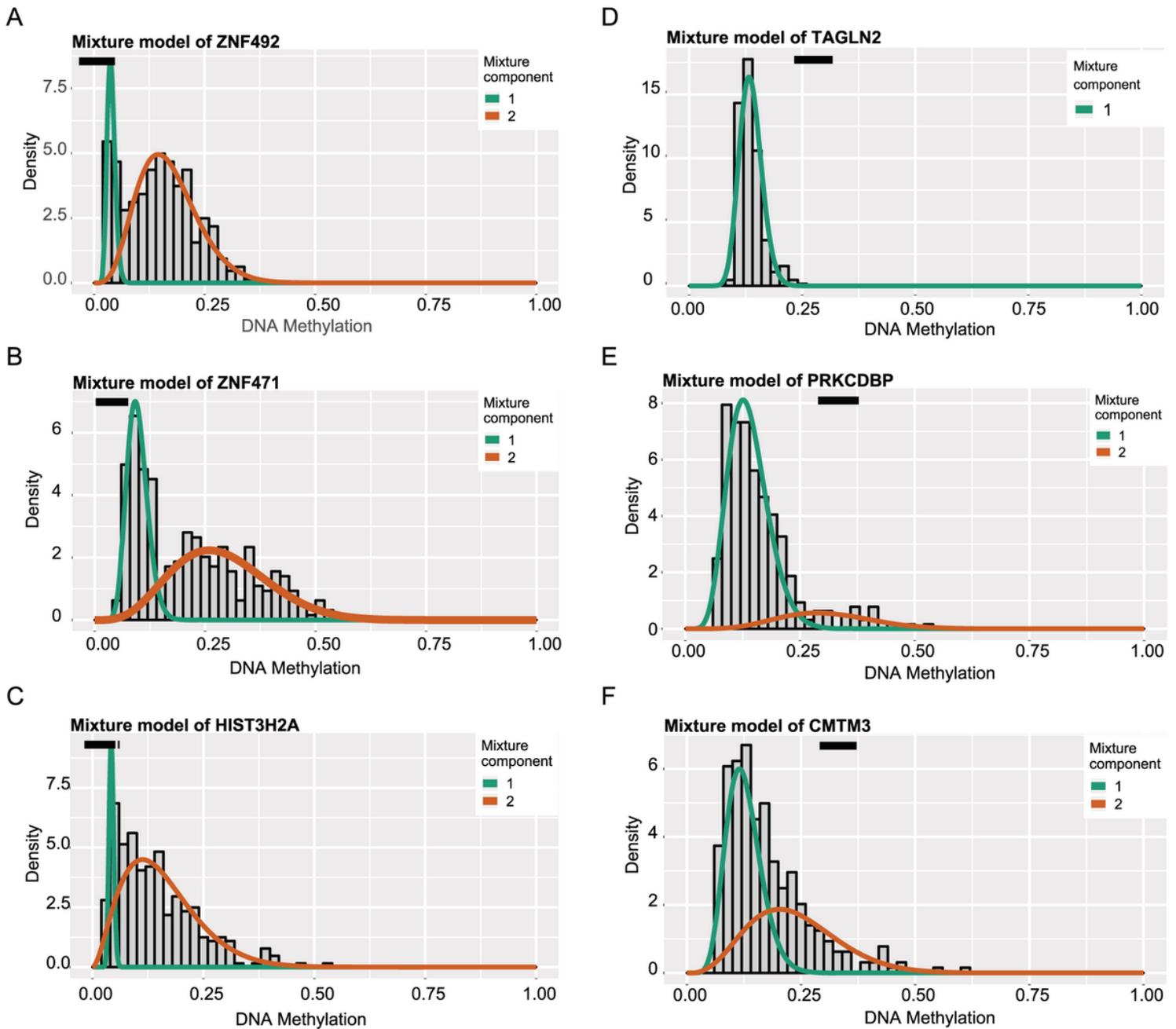
52. Roh JL, Park JY, Kim EH, et al. Activation of mitochondrial oxidation by PDK2 inhibition reverses cisplatin resistance in head and neck cancer. *Cancer Lett* 2016;371:20-9.
53. Sun W, Zhou S, Chang SS, et al. Mitochondrial mutations contribute to HIF1alpha accumulation via increased reactive oxygen species and up-regulated pyruvate dehydrogenase kinase

## Supplemental Figure Legends

**Figure S1.** Correlation analysis of methylation with expression levels for 5 hub genes. (A) TAGLN2, (B) PDK2, (C) HHLA2, (D) HOXA2, (E) XAF1.

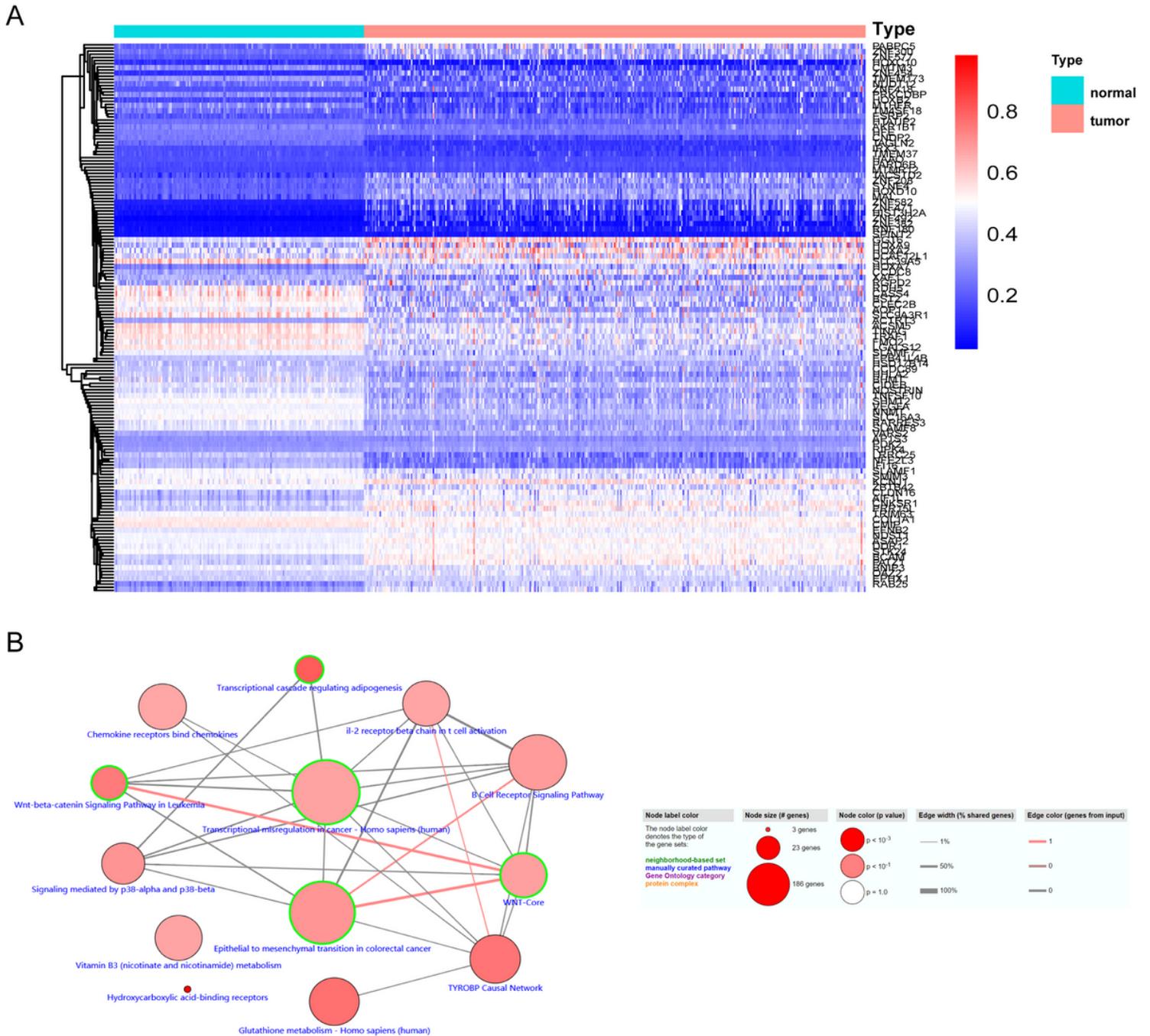
**Figure S2.** Meta-analysis of 5 hub MDGs in Oncomine database across 6 analysis. The rank for a gene is the median rank for that gene across each of the analyses.

## Figures



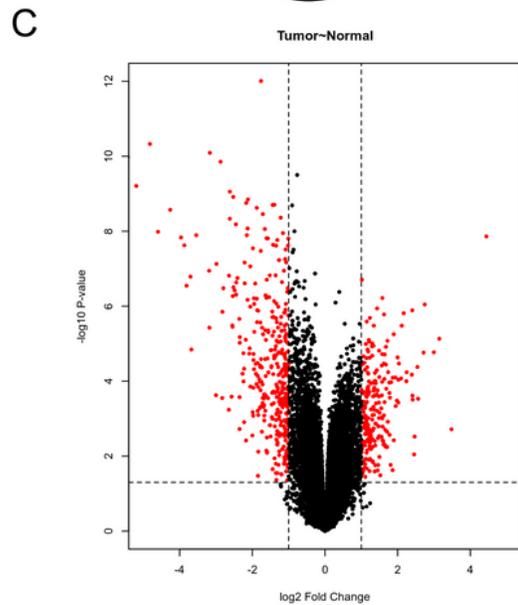
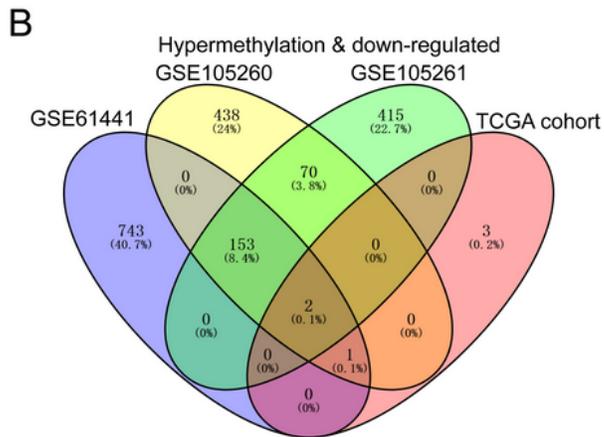
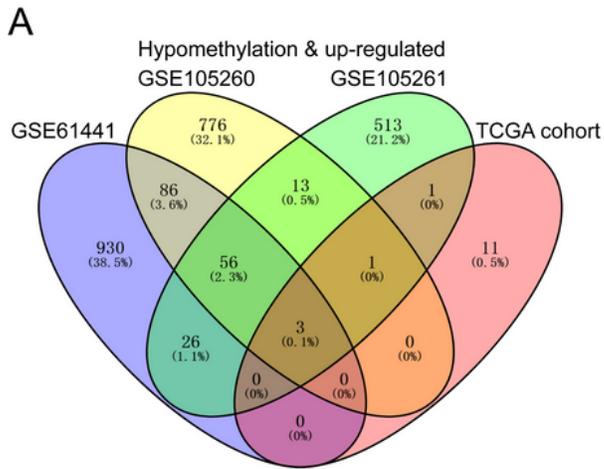
**Figure 1**

Top hypermethylated and top hypomethylated genes based on the “MethylMix” package. The histogram demonstrates the distribution of methylation in tumor samples, where two fitted lines with green or yellow represent the two components. The horizontal black bar above the histogram represents the distribution of methylation values ( $\beta$ ) in the normal samples. (A-C). Top 3 hypermethylated genes. (D-F). Top 3 hypomethylated genes.



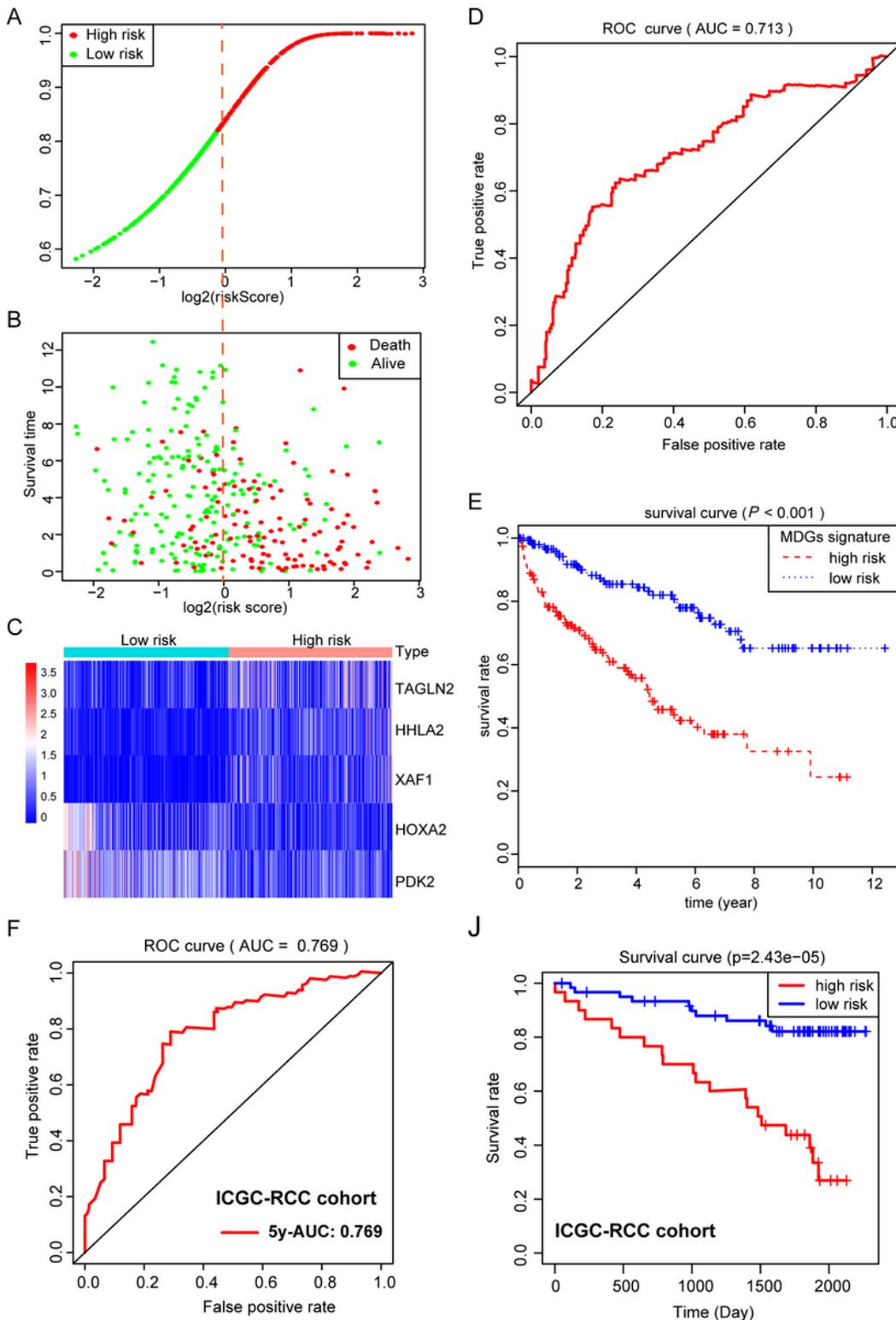
**Figure 2**

(A) Heatmap of 175 methylation-driven genes in tumor and normal samples. A trend from hypomethylation to hypermethylation was represented by the color from blue to red. (B) The significantly enriched pathways of 175 MDGs by MethylMix criteria. Only the pathways which  $P < 0.01$  are selected to show here, and a list of top pathways is included in Table 2.



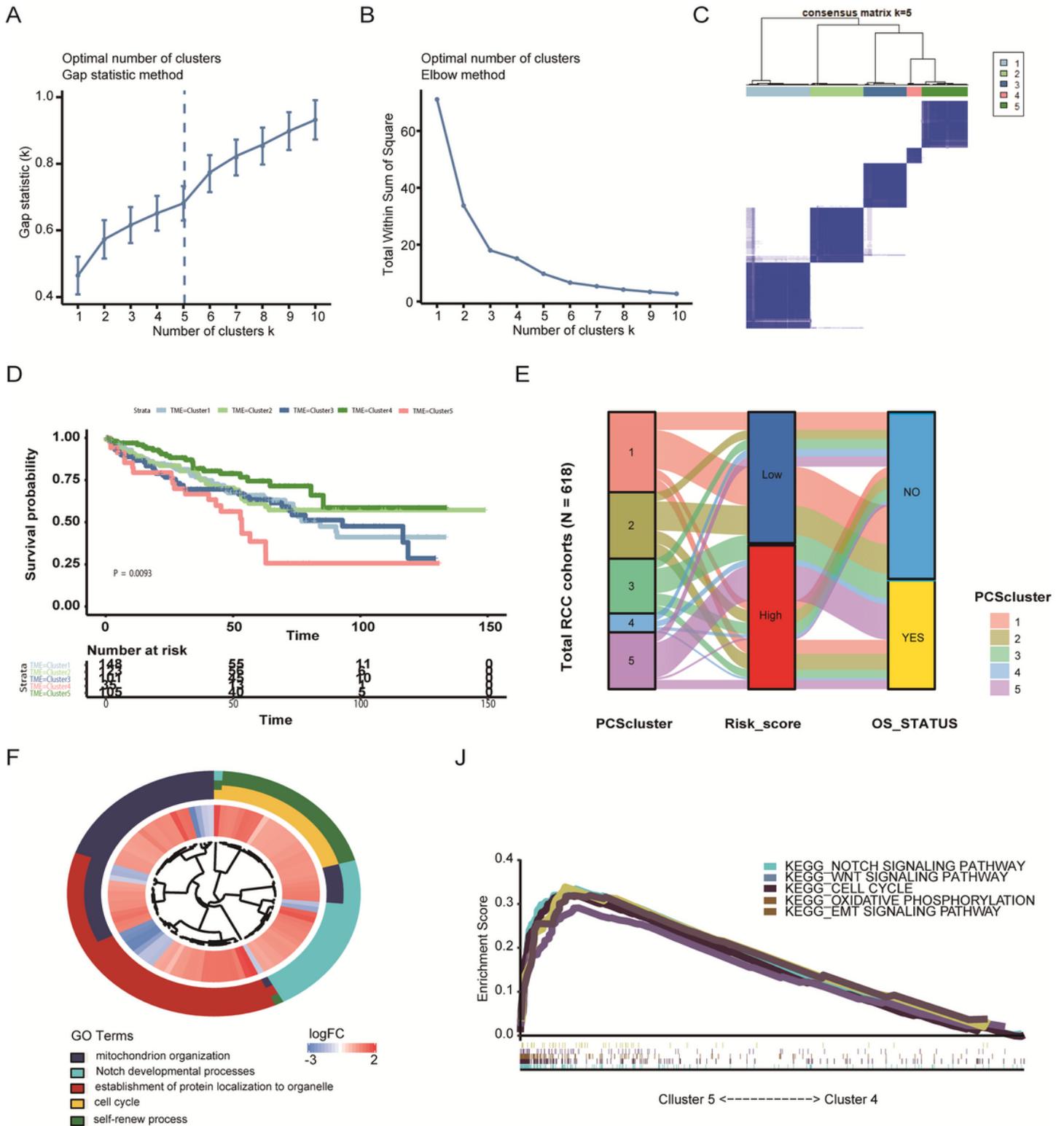
**Figure 3**

Identification of 5 hub MDGs signature. (A-B) Venn graph of MDGs in gene methylation series (GSE61441, GSE105260, TCGA cohort) and differentially expressed genes in GSE105261. (C) Volcano plot of GSE105261 is drawn, where red dot represents significantly differential mRNA ( $|\text{Fold Change}| > 2$ ,  $P < 0.05$ ).



**Figure 4**

Prognostic risk model was constructed based on the MDGs signature and validation in ICGC cohort. (A-B) The 5 MDGs-based risk score distribution and vital status of 319 ccRCC patients in high- and low-risk groups. (C) Heatmap of the 5-MDGs expression profiles in low- and high-risk levels. (D) ROC plot of 5 MDGs signature for OS prediction. (E) Kaplan-Meier curve of two risk groups based on MDGs signature.



**Figure 5**

Molecular classification assessment based on TCGA-KIRC and ICGC-RCC cohort. (A-C) We totally collected 618 patients from the TCGA-KIRC and ICGC-RCC cohort and used the ConsensusClusterPlus package to conduct the unsupervised clustering. We iterated 1000 times (K = 1: 10) to stabilize the classification categories, obtaining the optimized classification of the samples. (D) Kaplan-Meier analysis suggested that the MDGcluster4 (N = 35) has the best prognosis, whereas MDGcluster5 (N = 105) has the

worst prognosis. (E) Sankey diagram also illustrated the underlying associations among MDGcluster, risk score levels and vital outcomes. (F-J) Differential analysis and functional analysis revealed that the notch signaling pathway, wnt signaling pathway and EMT pathway were significantly up-regulated in MDGcluster5 than MDGcluster4, partially contributing to the poor prognosis of MDGcluster5.

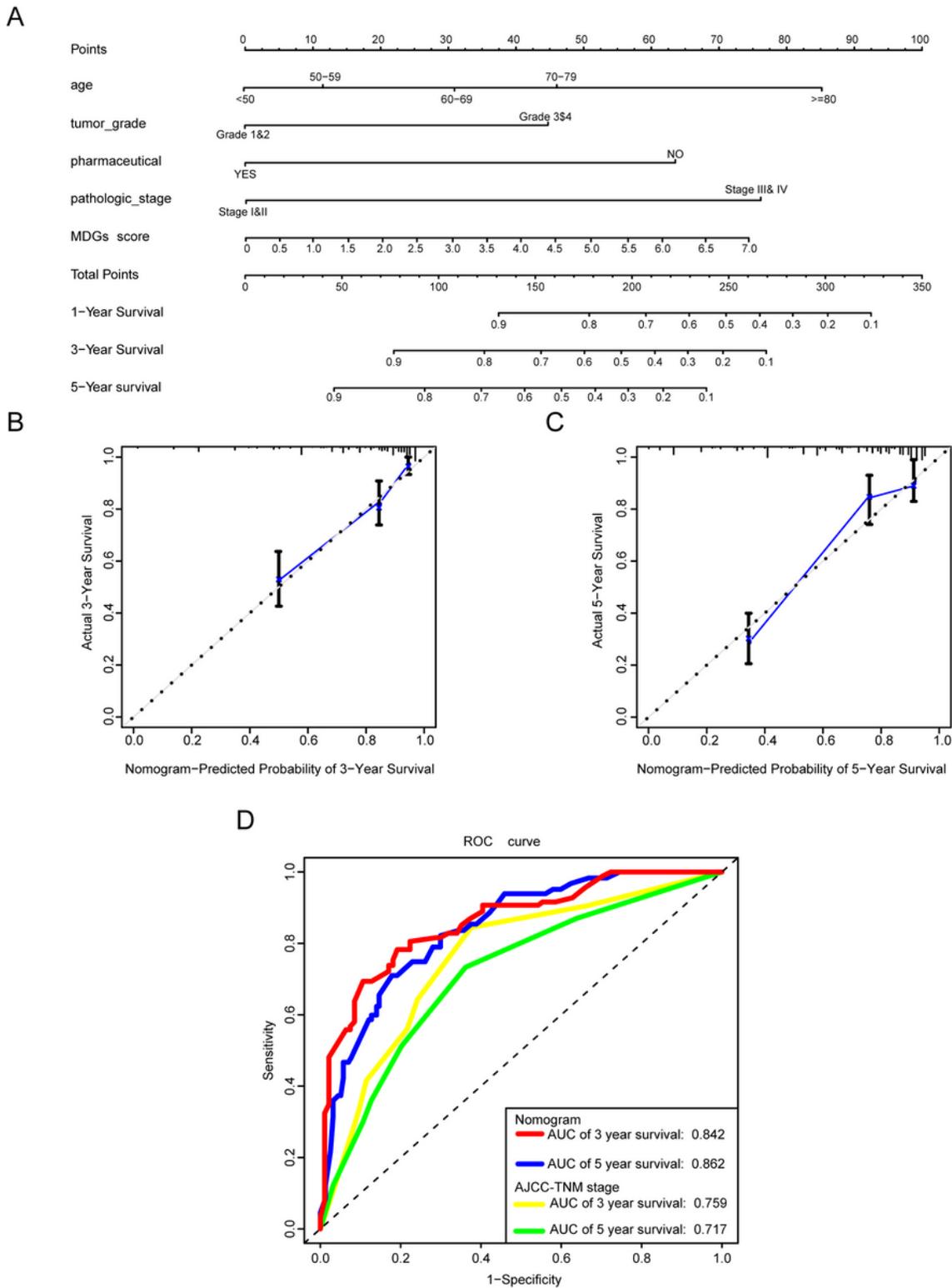


Figure 6

Construction of the OS prediction for ccRCC patients. (A) The integrated nomogram for predicting OS of ccRCC patients. (B-C) Calibration plots represent the OS prediction at 3 year (left) and 5 year (right), where the predicted and actual probabilities were shown on the x and y axis, respectively. (D) Comparison of predictive efficiency between nomogram and traditional TNM stage.

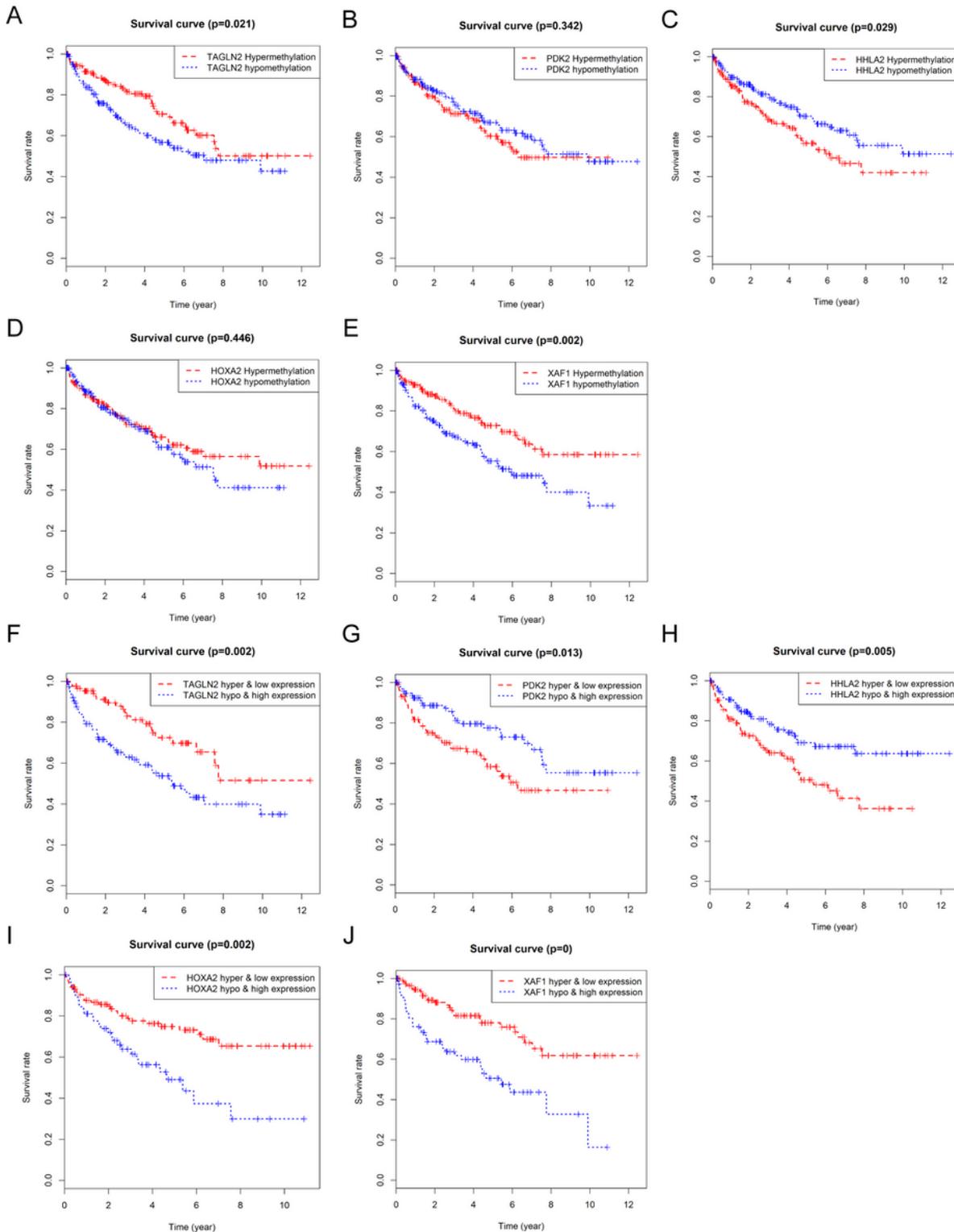
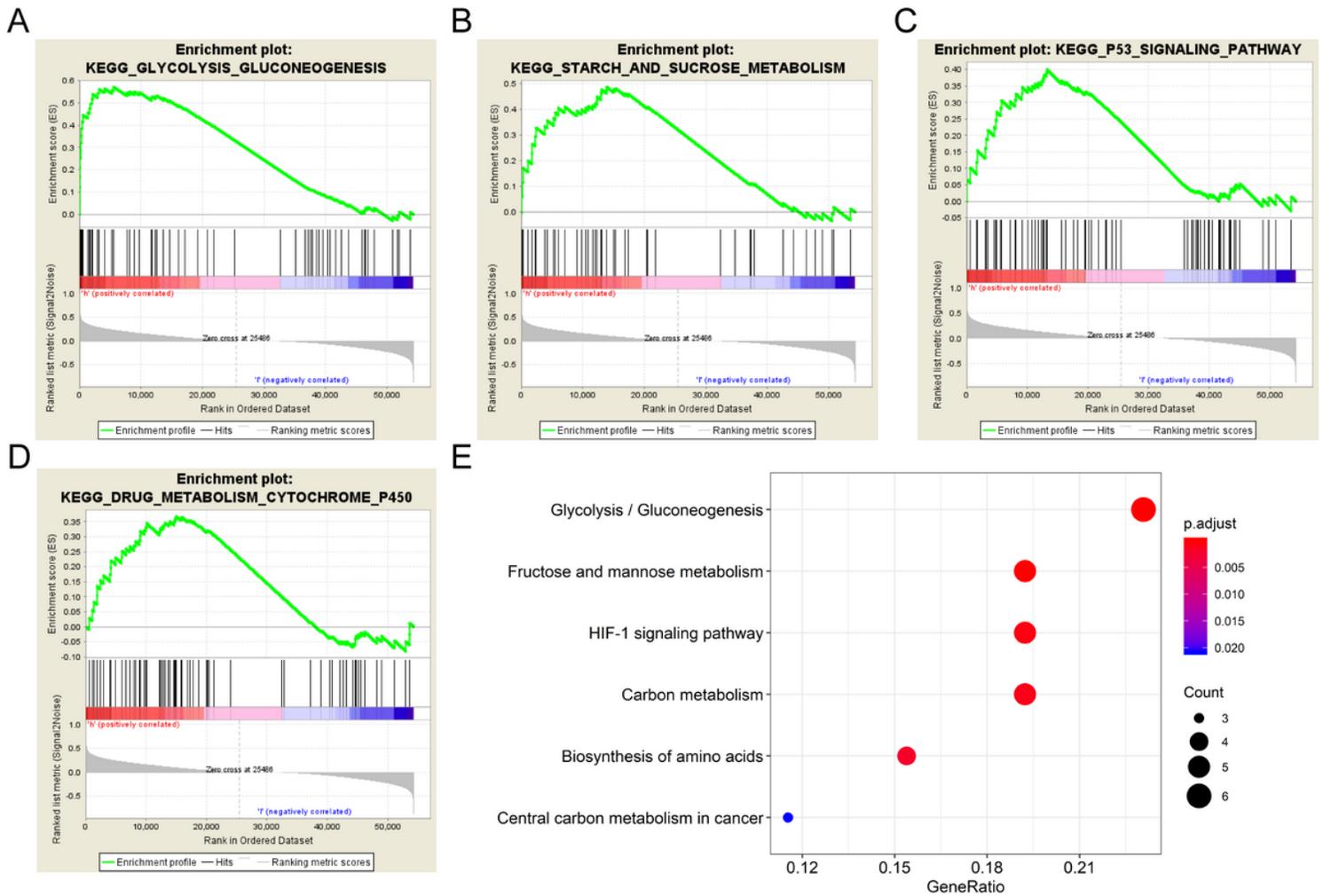


Figure 7

Kaplan-Meier survival curves of 5 hub MDGs. (A-E) Methylation levels of 5 genes (TAGLN2, PDK2, HHLA2, HOXA2, XAF1) with OS. (F-J) A joint survival analysis of methylation status with expression levels of 5 genes for ccRCC patients.



**Figure 8**

(A-D) Significant enrichment of top 4 MDGs-related pathways in high-risk group. (E) Dotplot represents KEGG functional pathway correlated analysis of the MDGs- associated genes obtained from co-expression analysis.

## Supplementary Files

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

- [TableS1.xls](#)
- [FigureS1.tif](#)
- [TableS2.xlsx](#)
- [FigureS2.tif](#)