Received: 7 July 2021 | Revised: 14 September 2021 | Accepted: 19 September 2021

DOI: 10.1002/jcla.24031

RESEARCH ARTICLE

WILEY

A five-gene methylation signature predicts overall survival of patients with clear cell renal cell carcinoma

Xiao Jing[1] | Gang Xu[2] | Yu Gong[1] | Junlong Li[2] | LingfengWu[3] | Wei Zhu[3] | Yi He[3] | Zhongyi Li[1] | Shouhua Pan[2]

[1]Department of Urology, The Second Affiliated Hospital of Zhejiang University School of Medicine, Hangzhou, China

[2]Department of Urology, Shaoxing People's Hospital, Shaoxing, China

[3]Department of Urology, The Affiliated Hospital of Jiaxing University, Jiaxing, China

Correspondence

Shouhua Pan, Department of Urology, Shaoxing People's Hospital, No. 568 Zhongxing North Road, Yuecheng District, Shaoxing, Zhejiang 312000, China.
Email: 13606550587@163.com(SP)

Zhongyi Li, Department of Urology, The Second Affiliated Hospital of Zhejiang University School of Medicine, No. 88 Jiefang Road, Shangcheng District, Hangzhou, Zhejiang 310009, China.
Email: 1178700168@zju.edu.cn(ZL)

Yi He, Department of Urology, The Affiliated Hospital of Jiaxing University, No. 1882 Zhonghuan South Road, Jiaxing, Zhejiang 314001, China.
Email: heyi@zjxu.edu.cn(YH)

Funding information

This work was supported by the Zhejiang Provincial Medicine, Health, and Science and Technology Project (grant number 2017KY152), Innovative Talents Project of Zhejiang Medicine and Health Science and Technology Plan (grant number 2021RC129), and Zhejiang Medical and Health Research Fund Project (grant number 2019KY710)

Abstract

Background: In this study, we aimed to screen methylation signatures associated with the prognosis of patients with clear cell renal cell carcinoma (ccRCC).

Methods: Gene expression and methylation profiles of ccRCC patients were downloaded from publicly available databases, and differentially expressed genes (DEGs)-differentially methylated genes (DMGs) were obtained. Subsequently, gene set enrichment and transcription factor (TF) regulatory network analyses were performed. In addition, a prognostic model was constructed and the relationship between disease progression and immunity was analyzed.

Results: A total of 23 common DEGs-DMGs were analyzed, among which 14 DEGs-DMGs were obtained with a cutoff value of PCC < 0 and $p < 0.05$. The enrichment analysis showed that the 14 DEGs-DMGs were enriched in three GO terms and three KEGG pathways. In addition, a total of six TFs were shown to be associated with the 14 DEGs-DMGs, including RP58, SOX9, NF-κB65, ATF6, OCT, and IK2. A prognostic model using five optimized DEGs-DMGs which efficiently predicted survival was constructed and validated using the GSE105288 dataset. Additionally, four types of immune cells (NK cells, macrophages, neutrophils, and cancer-associated fibroblasts), as well as ESTIMATE, immune, and stromal scores were found to be significantly correlated with ccRCC progression (normal, primary, and metastasis) in addition to the five optimized DEGs-DMGs.

Conclusion: A five-gene methylation signature with the predictive ability for ccRCC prognosis was investigated in this study, consisting of *CCNB2*, *CDKN1C*, *CTSH*, *E2F2*, and *ERMP1*. In addition, potential targets for methylation-mediated immunotherapy were highlighted.

KEYWORDS

clear cell renal cell carcinoma, immune, methylation, pathway, transcription factor

Xiao Jing, Gang Xu and Yu Gong authors contributed equally to this work.

1 | INTRODUCTION

Renal cell carcinoma (RCC) is a malignant tumor of the urinary system, with an annual incidence rate accounting for 2–3% of all malignant tumors.[1] The incidence rate in males is higher than that in females, with the ratio of males to females being approximately 2:1.[2] Clear cell RCC (ccRCC) is the most common histological subtype of RCC, accounting for more than 70% of all RCC cases. Moreover, its patho-physiological behavior is extremely complex.[3] Approximately 25% of ccRCC patients are first diagnosed in advanced stages and approximately 33% of ccRCC patients show recurrence or metastasis after surgery.[4] Thus, there is an urgent need to develop more effective therapeutic strategies against ccRCC.

Methylation is a first-line biochemical process playing an important role in the transmission of life and essentially involved in DNA and histone modification.[5] Numerous studies have shown that dysregulation of methylation processes (histones and DNA) can result in cancer development. For instance, Botezatu et al.[6] have reported that changes in DNA methylation can lead to the activation of typically silent genes or silencing of generally active genes. Koch et al.[7] have also suggested that cancer-related changes in DNA methylation are promising targets for the development of powerful diagnostic, prognostic, and predictive biomarkers. Moreover, McCabe et al.[8] have validated numerous potential therapeutic targets for cancer, including many that affect histone methylation. As compared to DNA mutations, gene methylation changes are often events that happen early in the process of cellular carcinogenesis. Thus, they can be used as important risk factors for tumor occurrence and as molecular marker for early diagnosis.[9] As such, screening of high-risk groups via methylation detection could improve the accuracy of early diagnoses and provide valuable therapeutic intervention time for patients with tumors[10,11]. However, few studies have reported methylation signatures that are associated with the prognosis of patients with different stages of ccRCC progression.

In this study, gene expression and methylation profiles were downloaded from Gene Expression Omnibus (GEO) and The Cancer Genome Atlas (TCGA) databases. Differentially expressed genes (DEGs) and differentially methylated genes (DMGs) were obtained between primary *vs*. normal, metastasis *vs*. normal, and metastasis *vs*. primary ccRCC samples. Additionally, a prognostic model was constructed and the relationship between disease progression and immunity was analyzed. Consequently, our study highlights the potential implication of novel methylation-related biomarkers for ccRCC progression offering a theoretical basis for efficacious drug development for the same. The systematic workflow of this study is illustrated in Figure 1.

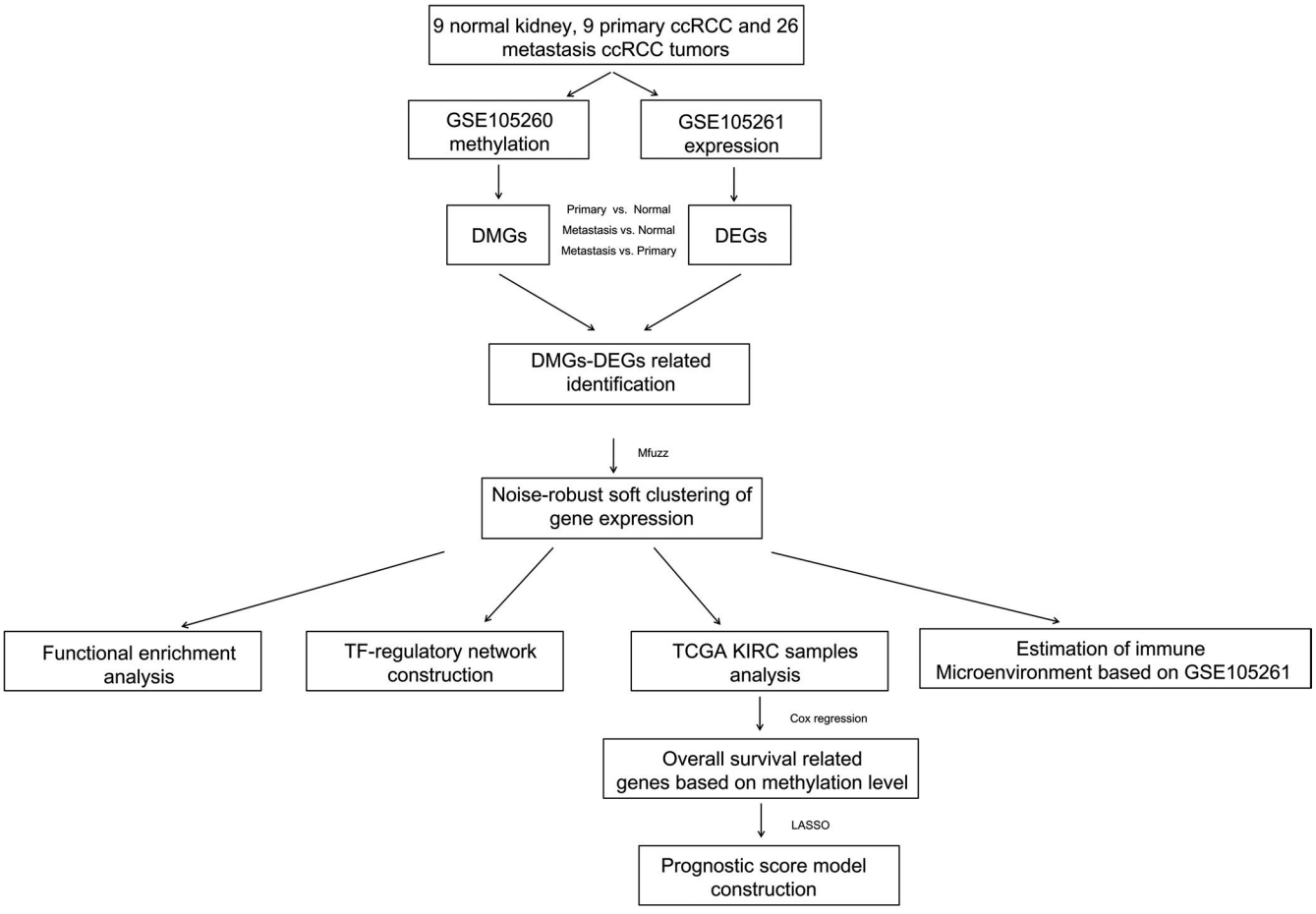


FIGURE 1 The schematic workflow of the study

2 | MATERIALS AND METHODS

2.1 | Data acquisition and preprocessing

The GSE105288 dataset[12] was downloaded from the GEO database,[13] which included two sub-datasets, namely GSE105260 (methylation profiling, 44 samples) and GSE105261 (gene expression profiling, 44 samples), obtained using the Illumina HumanMethylation450 BeadChip and Illumina HumanHT-12 V4.0 expression beadchip detection platforms, respectively. Preprocessed data with methylated beta detection values and standardized gene expression levels were downloaded, and the methylation and gene expression profile of the probe corresponding to the gene information were obtained according to the annotation information for each detection platform.

The data of ccRCC patients with methylated beta detection values and level 3 normalized data of log(FPKM+1,2) expression level based on the Illumina Infinium Human Methylation 450 BeadChip and Illumina HiSeq 2000 RNA Sequencing detection platforms, respectively, were downloaded from TCGA. A total of 333 samples with information regarding methylation and expression levels were obtained, among which 306 samples that had survival prognostic information were used in this study.

2.2 | Identification of DEGs-DMGs

Based on the GSE105261 and GSE105260 datasets, the DEGs and DMGs were screened between primary *vs*. normal, metastasis *vs*. normal, metastasis *vs*. primary using limma package[14] with the threshold of FDR <0.05, $|\log_2 FC| > 0.263$. The overlapping DEGs and DMGs were then obtained, and a heatmap was constructed using the "pheatmap" package[15] in R software. The overlapping DEGs were intersected with overlapping DMGs, and the overlapping co-expressed genes were considered to be the common DEGs-DMGs. The Pearson correlation coefficient (PCC) and Spearman's coefficient (Rho) of methylation and expression levels of the common DEGs-DMGs were calculated using the cor.test function in R software. The DEGs-DMGs with cutoff values of PCC <0 and $p < 0.05$ were used for further analysis. In addition, the DAVID tool[16] was used to perform gene set enrichment analysis with a cutoff value of $p < 0.05$.

2.3 | Gene expression trends based on the Mfuzz clustering algorithm

Based on the expression level of DEGs-DMGs with cutoff values of PCC < 0 and $p < 0.05$, the "Mfuzz" package[17] in R software was used to conduct the change trend analysis of expression patterns for these DEGs-DMGs and the expression trend module gene clustering was obtained. Here, special attention was paid to related genes whose expression levels continued to increase or decrease along with the progression (normal, primary, and metastasis) of ccRCC.

2.4 | Construction of the transcription factor (TF) regulatory network

Based on the expression level of the DEGs-DMGs with a cutoff value of PCC < 0 and $p < 0.05$, the DAVID tool was then used to screen TFs significantly associated with DEGs-DMGs. Subsequently, the interactions between TFs and DEGs-DMGs were obtained and the TF-DEGs-DMGs regulatory network was built using Cytoscape.[18]

2.5 | Construction of a prognostic model

The data retrieved from TCGA were used to construct a prognostic model. The 306 ccRCC tumor samples from the database were randomly divided into two groups, including the training and validation datasets. Based on the methylation and expression level of the DEGs-DMGs with a cutoff value of PCC < 0 and $p < 0.05$, the "survival" R package[19] was used to conduct a univariate Cox regression analysis on the training dataset to screen DEGs-DMGs significantly associated with survival prognosis at both methylation and expression levels. $p < 0.05$ was the criteria for determining prognosis-related DEGs-DMGs. The LASSO regression[20] from the "lars" R package was used to conduct the survival regression analysis to screen the optimized DEGs-DMGs based on the gene methylation levels in the training dataset samples. Moreover, the optimized DEGs-DMGs were verified in the GSE105288 dataset, which included 9 normal, 9 primary, and 26 metastasis samples. Subsequently, a prognostic score (PS) model was established. The formula $PS = \sum Coef_{genes} \times Methylation_{genes}$ (where $Coef_{genes}$ represents the LASSO prognosis coefficient of genes and $Methylation_{genes}$ represents the gene methylation levels in the training dataset) was used to derive the risk score. The PS was then calculated for the training dataset, validation dataset, and whole groups. The samples were later divided into the low- and high-PS groups using the median PS value. Subsequently, the Kaplan-Meier curve method in the "survival" package was used to compare differences in survival between the low- and high-PS groups. To further understand the correlation between the methylation and expression levels of the optimized DEGs-DMGs screened and survival prognosis, the Kaplan-Meier curve analyses were also performed on the whole dataset. Based on the clinical information of the samples in TCGA dataset, the aov function was used to conduct variance analysis to compare the differences in the distribution of methylation levels of optimized DEGs-DMGs in different clinical groups.

2.6 | Analysis of disease progression and immunity

Based on the genome-wide expression levels obtained in the GSE105261 dataset, the "MCPcounter"[21] R package was utilized to evaluate the level of immune cell infiltration, including eight immune cell types [$CD3^+$ T cells, $CD8^+$ T cells, cytotoxic lymphocytes, NK cells, B lymphocytes, cells originating from monocytes

(monocytic lineage), myeloid dendritic cells, and neutrophils], and two stromal cell types (endothelial cells and fibroblasts). The aov function was used to conduct variance analysis to compare the differences between the proportions of different immune cells in the progression groups (normal, primary, and metastatic). In addition, the "estimate" R package[22] was used to calculate the ESTIMATE, immune, and stromal scores. The aov function was also used to conduct variance analysis to compare the differences between the distribution of the three scores in the progression groups (normal, primary, and metastatic). Subsequently, the correlation between DEGs-DMG methylation levels and the proportion of various immune cells along with the three scores were calculated.

2.7 | Statistics analysis

The DEGs and DMGs were screened with the threshold of FDR < 0.05, $|\log_2 FC| > 0.263$, and the DEGs-DMGs were identified using the "limma" R package with cutoff values of PCC < 0 and $p < 0.05$. Correlation analysis was performed using the Pearson and Spearman correlation test. A univariate Cox regression analysis was conducted using the "survival" package. Survival regression analysis was carried out utilizing LASSO algorithm. Kaplan-Meier curves were used to evaluate survival time in patients with ccRCC. The aov function was used to conduct variance analysis to compare the differences in the distribution of methylation levels of optimized DEGs-DMGs in different clinical groups. $p < 0.05$ was considered statistically significant.

3 | RESULTS

3.1 | Identification of DEGs-DMGs

As shown in Figure 2A, total of 862, 2431, and 1333 DEGs were screened between primary *vs*. normal, metastasis *vs*. normal, metastasis *vs*. primary groups, respectively. Likewise, a total of 1460, 2811, and 1044 DMGs were screened between primary *vs*. normal, metastasis *vs*. normal, metastasis *vs*. primary groups, respectively. Our analyses thus revealed a total of 350 and 367 overlapping DEGs and DMGs, respectively (Figure 2B and C); the heatmap is shown in Figure 2D and E. Subsequent intersections of overlapping DEGs with overlapping DMGs led to the identification of 23 common DEGs-DMGs (Figure 2F and Supplementary File S1). The overall correlation between the methylation and expression levels of the 23 common DEGs-DMGs was analyzed, and the results showed that 14 DEGs-DMGs were above the cutoff threshold of PCC < 0 and $p < 0.05$ (Figure 2G and Supplementary File S2). Further, gene set enrichment analysis showed that the 14 DEGs-DMG were enriched in three gene ontology (GO) terms and three Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways (Table 1).

3.2 | Identification of the DEGs-DMGs with the same expression pattern trends

The "Mfuzz" package in R software was used to conduct the change trend analysis of expression patterns for the DEGs-DMGs, and the expression trend module gene clustering was obtained. The results revealed that the 14 DEGs-DMGs could be clustered into two expression trends categories, with the expression of Cluster 1 being significantly downregulated, while the expression of Cluster 2 continued to be significantly upregulated (Figure 3A). A total of 11 DEGs-DMGs (*CAPS*, *CDKN1C*, *CKMT1B*, *CTSH*, *ERMP1*, *ERP27*, *GGT6*, *GSTM3*, *KCNJ1*, *SCNN1A*, and *STK33*) were found in Cluster 1, while Cluster 2 included three DEGs-DMGs (*CCNB2*, *E2F2*, and *SLFN11*).

3.3 | Construction of the TF regulatory network associated with the 14 DEGs-DMGs

The 14 DEGs-DMGs were subjected to functional annotation using Database for Annotation, Visualization, and Integrated Discovery (DAVID) tools, and a total of six TFs were found to be associated with the 14 identified DEGs-DMGs, including RP58, SOX9, NF-κB65, ATF6, OCT, and IK2 (Table 2). Accordingly, a TF-DEGs-DMGs regulatory network was built using Cytoscape (Figure 3B).

3.4 | Construction of a ccRCC prognostic model based on five-gene DEG-DMG signature

Based on the univariate Cox regression analysis, a total of seven DEGs and eight DMGs were found to be significantly associated

TABLE 1 Enrichment analysis of the 14 DEG-DMRs

Category	Term	Count	*p*-Value	Genes
Biology Process	GO:0006749~glutathione metabolic process	2	1.43E−02	GSTM3, GGT6
	GO:0006508 ~ proteolysis	3	2.56E−02	ERMP1, GGT6, CTSH
	GO:0051726 ~ regulation of cell cycle	2	3.92E−02	CCNB2, E2F2
KEGG Pathway	hsa04110:Cell cycle	3	8.41E−03	CDKN1C, CCNB2, E2F2
	hsa04960:Aldosterone-regulated sodium reabsorption	2	2.44E−02	SCNN1A, KCNJ1
	hsa00480:Glutathione metabolism	2	4.58E−02	GSTM3, GGT6

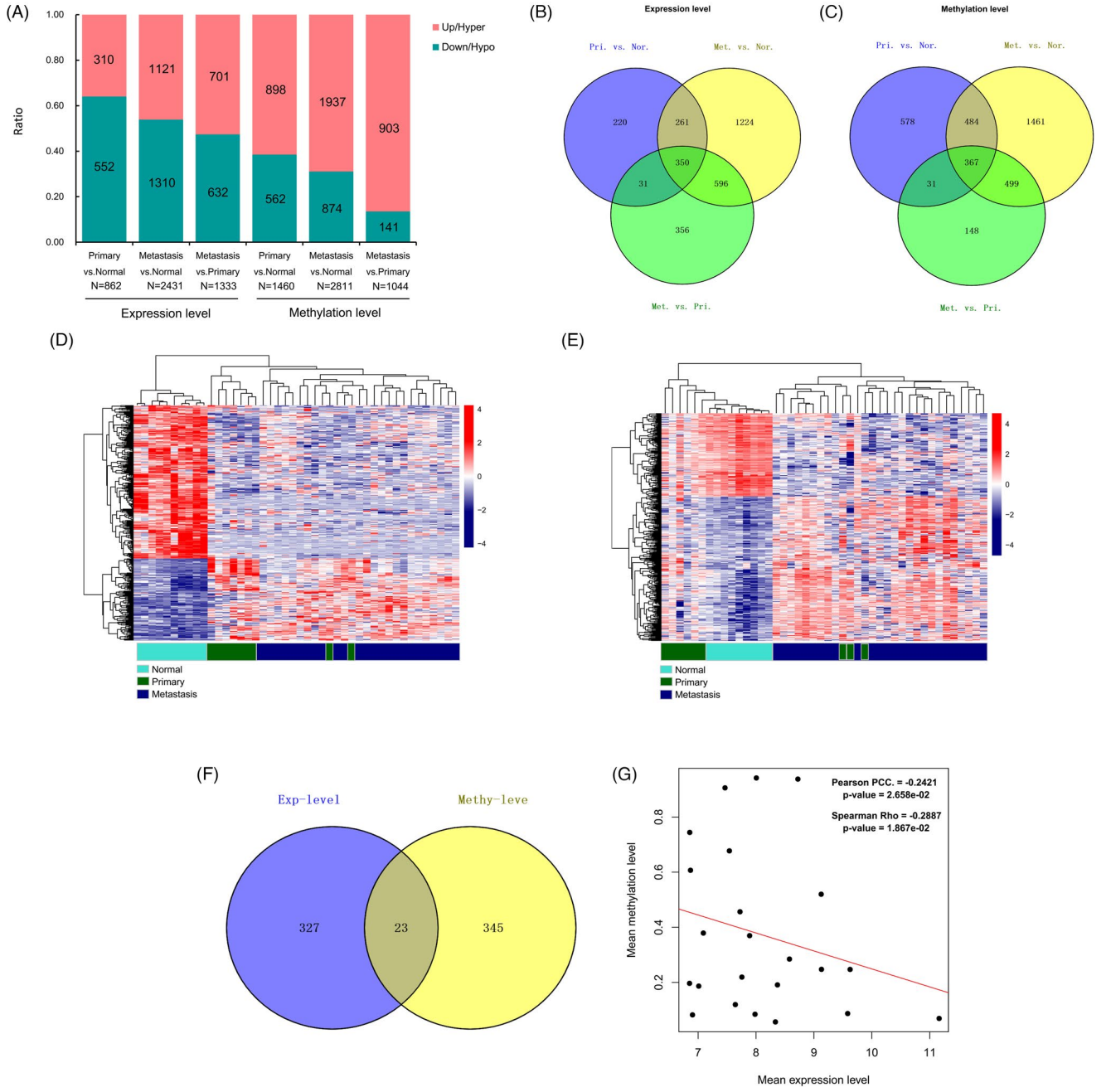


FIGURE 2 Identification of differentially expressed genes (DEGs)-differentially methylated genes (DMGs). (A) DEGs and DMGs screened between the primary *vs.* normal, metastasis *vs.* normal, and metastasis *vs.* primary groups, respectively. The overlapping DEGs (B) and DMGs (C) obtained between the primary *vs.* normal, metastasis *vs.* normal, and metastasis *vs.* primary groups. The heatmap of overlapping DEGs (D) and DMGs (E). (F) The common DEGs-DMGs screened from overlapping DEGs and DMGs. G: Scatter plot of the correlation between the expression level of 23 common DEGs-DMGs and the overall methylation level. The red line represents the correlation trend line

with prognosis. Moreover, both the methylation and expression levels of seven genes were shown to be significantly associated with prognosis. Thus, these seven DEGs-DMGs were used for further analyses. Based on the seven DEGs-DMGs, five optimized DEGs-DMGs were obtained using the LASSO regression, including *CCNB2*, *CDKN1C*, *CTSH*, *E2F2*, and *ERMP1* (Table 3). Moreover, the five optimized DEGs-DMGs were verified in GSE105288 dataset with the results showing that the gene expression levels of *CCNB2* and *E2F2* were upregulated with tumor progression, whereas those of *CDKN1C*, *CTSH*, and *ERMP1* were downregulated. Furthermore, the change trends of these DEGs-DMGs were consistent with the above results (Figure 3C). Subsequently, the five optimized DEGs-DMGs were used to construct a prognostic model. Based on the calculated PS values, the samples were divided into low- or high-PS groups based on the median PS value. In the training, validation, and entire datasets, patients in the low-PS group exhibited significantly longer OS than those in the high-PS group (Figure 4A, D, and G). The predictive ability of the five DEGs-DMGs-based

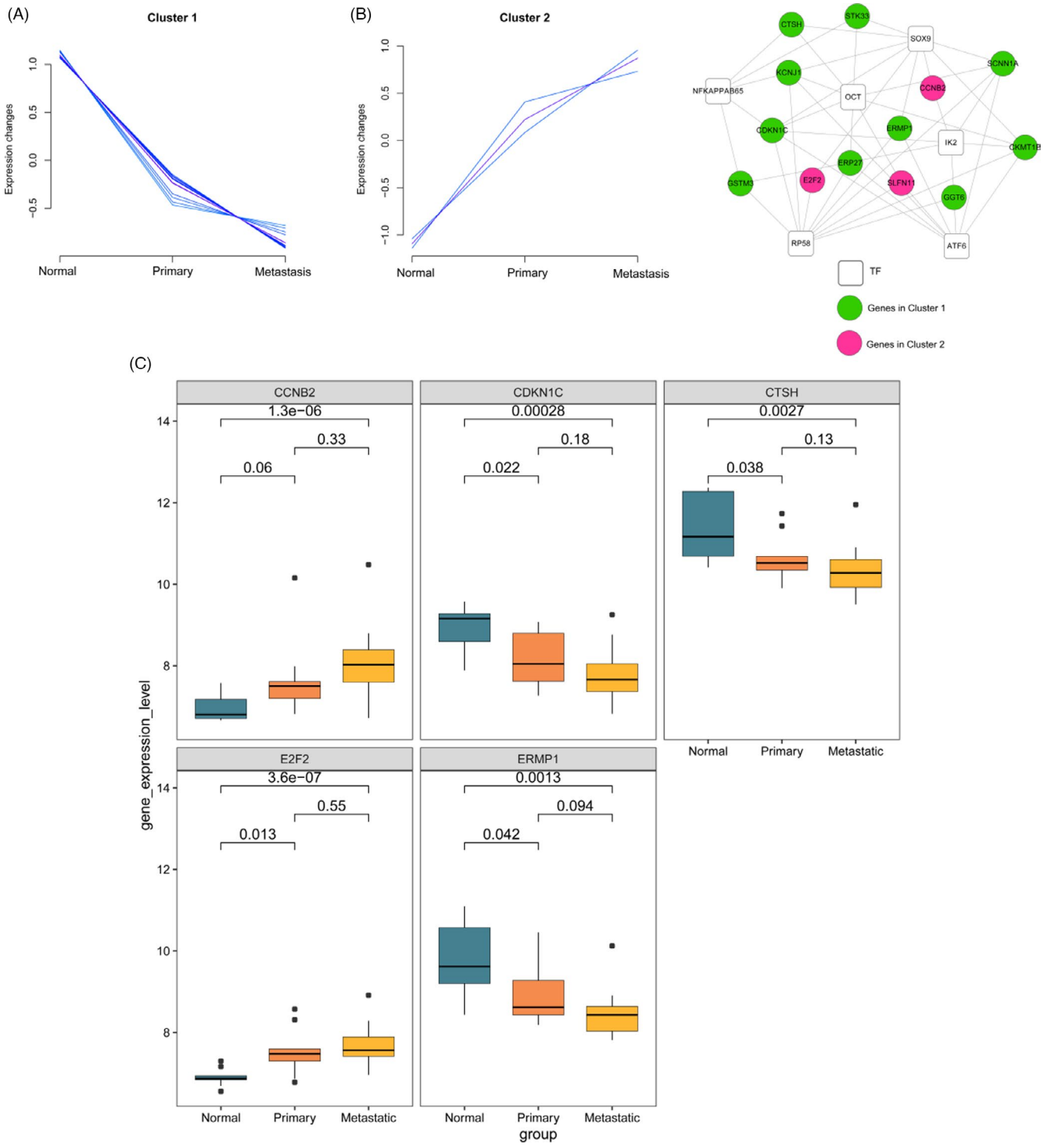


FIGURE 3 Identification of differentially expressed genes (DEGs)-differentially methylated genes (DMGs) with the same expression pattern trends, construction of the transcript factor (TF) regulatory network, and verification of the optimized DEGs-DMGs. (A) The expression trend module gene clustering based on 14 DEGs-DMGs. (B) The TF regulatory network. The red nodes had high *p* values. C: Five optimized DEGs-DMGs verified in GSE105288 dataset

prognostic model was evaluated subsequently by calculating the area under the curve (AUC) value of the receiver operating characteristic (ROC) curve. The AUC values for the 1-, 3-, and 5-year survival curves in the training, validation, and entire datasets are shown in Figure 4B, E, and H, respectively. The distribution of PS and survival status for the samples was displayed for the training, validation, and entire datasets (Figure 4C, F, and I). Based on the expression and methylation levels, the samples were divided into low- and high-expression or low- and high-methylation based on median expression or methylation levels, respectively. Survival

TABLE 2 Total 6 TFs obtained related to the 14 DEG-DMRs

Term	Count	*p*-Value	Genes
RP58	10	1.33E−02	CDKN1C, GSTM3, ERMP1, SLFN11, GGT6, SCNN1A, CKMT1B, E2F2, ERP27, KCNJ1
SOX9	9	1.48E−02	CDKN1C, CCNB2, STK33, ERMP1, SCNN1A, CKMT1B, CTSH, E2F2, KCNJ1
NFKAPPAB65	5	4.61E−02	CDKN1C, GSTM3, STK33, CTSH, KCNJ1
ATF6	8	4.64E−02	CDKN1C, ERMP1, SLFN11, GGT6, SCNN1A, CKMT1B, ERP27, KCNJ1
OCT	8	4.70E−02	CDKN1C, STK33, SLFN11, SCNN1A, CKMT1B, CTSH, ERP27, KCNJ1
IK2	6	4.74E−02	CDKN1C, GSTM3, CCNB2, GGT6, SCNN1A, CKMT1B

analysis showed that the expression or methylation level of the five optimized DEGs-DMGs was significantly associated with survival of ccRCC patients (Figure 5). In addition, the differences between the distributions of methylation levels for the five optimized DEGs-DMGs in different clinical groups were evaluated and the results showed that the methylation level of *CCNB2* was significantly correlated with pathological M, T, and stage of ccRCC samples (Figure 6A). The methylation level of *CDKN1C* was also shown to be significantly correlated with age, neoplasm histological grade, and pathological M (Figure 6B). Lastly, the methylation level of *ERMP1* was significantly correlated with neoplasm histological grade, pathological T, and tumor stage (Figure 6C).

3.5 | Correlation analysis of ccRCC progression with immunity and the five-gene DEGs-DMGs signature

Based on the genome-wide expression levels obtained in the GSE105261 dataset, the "MCPcounter" R package was utilized to evaluate the level of immune cell infiltration. The results of this analysis showed that four immune cell types were significantly correlated with ccRCC progression (normal, primary, and metastatic), which included NK cells, macrophages, neutrophils, and cancer-associated fibroblasts (Figure 7A–D). Moreover, the ESTIMATE, immune, and stromal scores were also found to be correlated with ccRCC progression (normal, primary, and metastasis) (Figure 7E–G). A significant correlation was also found among the four immune cell types, the three scores (ESTIMATE, immune, and stromal score), and the five optimized DEGs-DMGs (Figure 7H).

4 | DISCUSSION

Methylation is an important epigenetic modification of proteins and nucleic acids, which plays a crucial role in the regulation of gene expression; however, it is also closely associated with the development of several diseases, such as cancer.[23] In this study, methylation signatures associated with the prognosis of ccRCC patients were evaluated. A total of 23 common DEGs-DMGs were analyzed, among which 14 DEGs-DMGs were above the cutoff threshold of PCC < 0 and $p < 0.05$. Gene set enrichment analysis showed that the 14 DEGs-DMGs were enriched in three GO terms and three KEGG pathways. In addition, a total of six TFs were shown to be significantly associated with the 14 identified DEGs-DMGs, including RP58, SOX9, NF-κB65, ATF6, OCT, and IK2. Further, a prognostic model which effectively predicted survival of ccRCC patients was constructed using five optimized DEGs-DMGs and then validated using the GSE105288 dataset. Moreover, infiltration of four immune cells types (NK cells, macrophages, neutrophils, and cancer-associated fibroblasts) and the ESTIMATE, immune, and stromal scores were found to be significantly correlated with ccRCC progression (normal, primary, and metastasis). Additionally, the five optimized DEGs-DMGs were also shown to be correlated with the

TABLE 3 Five optimized DEG-DMRs obtained using LASSO arithmetic

		Methylation level				Expression level			
ID	LASSO Coefficient	Hazard Ratio (95%CI)	Standard error	Z score	*p*-value	Hazard Ratio (95%CI)	Standard error	Z score	*p*-value
CCNB2	−0.8117	0.080(0.011–0.842)	0.0761	−2.032	4.21E−02	2.239(1.695–2.958)	0.447	5.676	2.26E−09
CDKN1C	0.9393	2.611(1.025–6.650)	0.00383	1.391	1.64E−02	0.784(0.566–0.985)	1.276	−1.469	1.41E−02
CTSH	6.4132	2.482(1.393–25.73)	0.00239	1.457	1.48E−02	0.774(0.503–0.911)	1.293	−1.168	2.37E−02
E2F2	−0.2661	0.475(0.268–0.798)	0.0211	−1.240	1.98E−02	4.091(1.779–9.404)	0.245	3.317	7.54E−04
ERMP1	1.0382	8.876(4.123–70.13)	0.0113	4.254	1.31E−05	0.750(0.496–0.913)	1.334	−1.368	1.72E−02

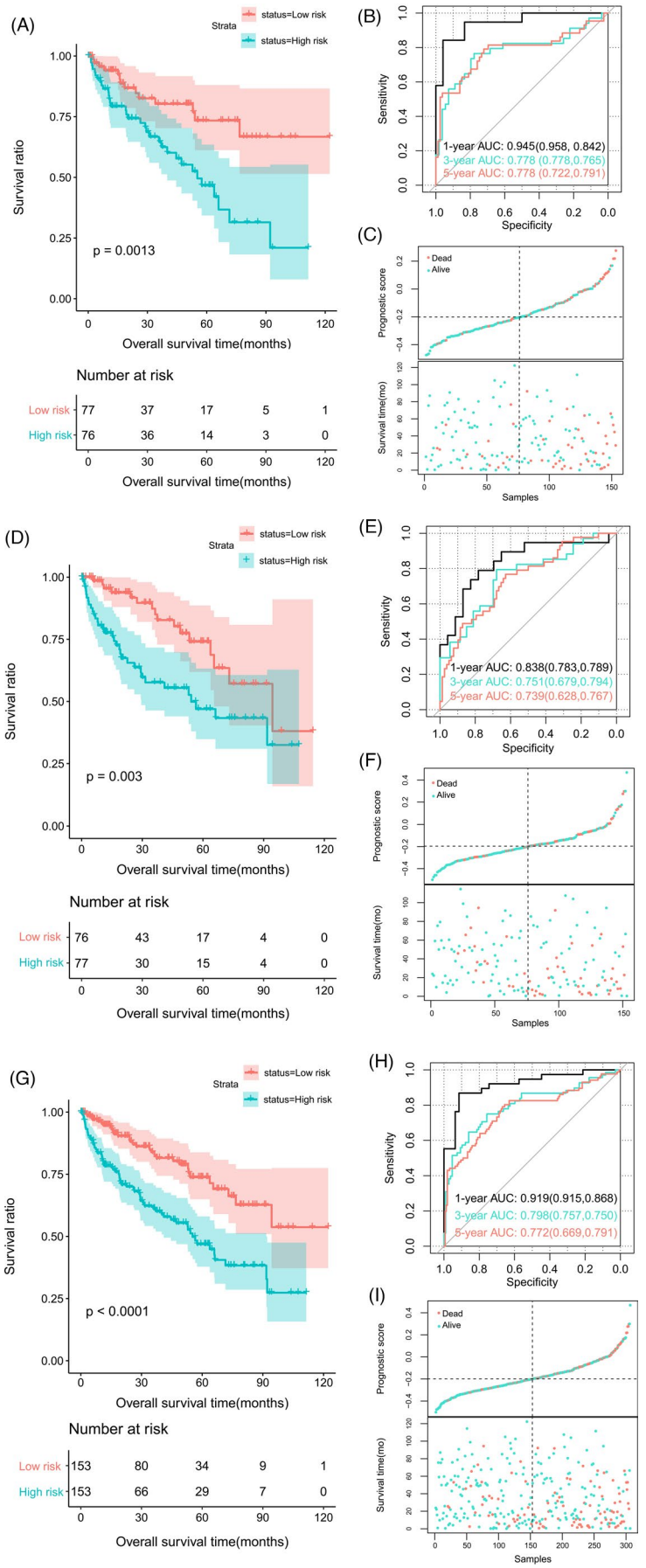


FIGURE 4 Construction of ccRCC prognostic model with the five-gene DEGs-DMGs signature. The Kaplan-Meier analysis, time-dependent ROC analysis, and risk score analysis for the five optimized DEGs-DMGs for the overall survival of ccRCC patients. (A, B, C) Training dataset. (D, E, F) Validation dataset. (G, H, I) Entire dataset

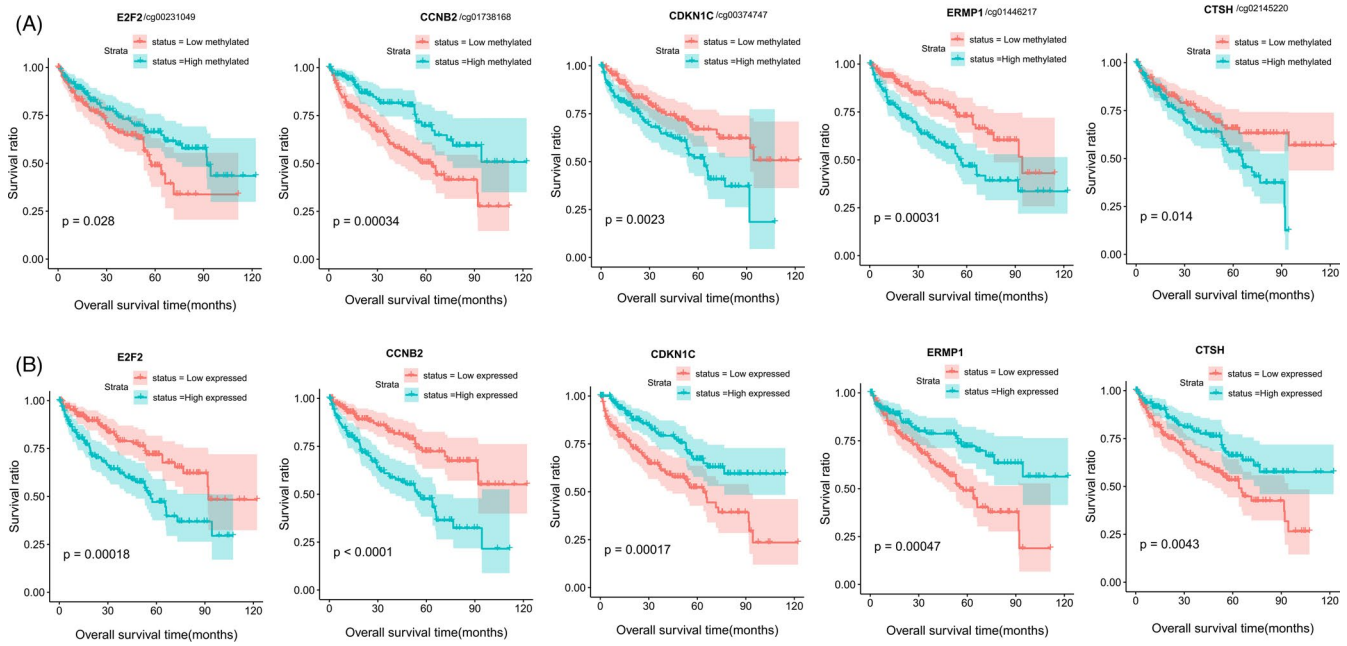


FIGURE 5 Kaplan-Meier analysis of the expression or methylation level of the five optimized DEGs-DMGs. (A) Methylation level. (B) Expression level

infiltration of these four immune cell types and the ESTIMATE, immune, and stromal score.

In this study, we found that the gene expression levels of 14 DEGs-DMGs were significantly correlated with methylation level. Their enrichment analysis showed that the 14 DEGs-DMGs were enriched in three KEGG pathways, including cell cycle, aldosterone-regulated sodium reabsorption, and glutathione metabolism. Cell cycle plays an important role in cancer development. Uncontrolled cell proliferation is a hallmark of carcinogenesis, and tumor cells usually exhibit impairment in genes that directly regulate the cell cycle.[24] Cell cycle dysregulation via the cyclin D/CDK/pRb pathway is frequently observed in breast cancer, which supports the rationale underlying the development of drugs targeting the cell cycle control machinery.[25] Hao et al.[26] have also identified potential biomarkers for ccRCC as being enriched in the aldosterone-regulated sodium reabsorption pathway. Glutathione is a major low-molecular-weight antioxidant, and targeting of the glutathione metabolism is important mechanism associated with anticancer therapies.[27] Notably, the augmented oxidative stress typically exhibited by cancer cells is also accompanied by an increase in glutathione levels, which confers growth advantages and resistance to a number of chemotherapeutic agents.[28] A total of six TFs were shown to be associated with the 14 identified DEGs-DMGs, including RP58, SOX9, NF-κB65, ATF6, OCT, and IK2. Some SOX proteins are potential molecular markers of cancer prognosis and are recognized as potential therapeutic targets, including SOX9, SOX2, and SOX4.[29] NF-κB proteins are key regulators of innate and adaptive immune responses and can accelerate cell proliferation. Notably, increased NF-κB activity is most often observed in solid tumors.[30] Moreover, higher but non-homogeneous expression of vascular endothelial growth factor is associated with NF-κB 65 activity in ccRCC cells.[31] ATF6 is one of the three major endoplasmic reticulum stress transducers and has been shown to promote chemotherapy resistance by regulating cancer cell survival.[32] Few studies have investigated the role of OCT, IK2, and RP58 in ccRCC.

As an important part of this study, we have constructed a prognostic model using five optimized DEGs-DMGs, including *CCNB2*, *CDKN1C*, *CTSH*, *E2F2*, and *ERMP1*, which effectively predict survival of ccRCC patients and are verified in GSE105288 dataset. Moreover, the model was verified using an external dataset and was found to have statistical significance. The five DEGs-DMGs used have been reported to be associated with ccRCC progression. For instance, there is evidence that *CCNB2* is associated with the prognosis of human cancers, such as RCC,[33] bladder cancer,[34] and lung cancer.[35] Qiu et al.[36] have shown that chromodomain Y-like promotes chemoresistance in small cell lung cancer by silencing its downstream mediator *CDKN1C*. Moreover, Na et al.[37] have reported that lncRNA STEAP3-AS1 modulates cell cycle progression by affecting *CDKN1C* expression via STEAP3 in colon cancer. LBX2-AS1 has also been shown to promote ovarian cancer progression by facilitating *E2F2* gene expression via miR-455-5p and miR-491-5p sponging.[38] Abnormal expression or activation of E2Fs is a common phenomenon in malignant tumors, and E2Fs are significantly correlated with the occurrence or progression of various types of tumors.[39] Notably, a study by Zhou et al.[40] has suggested that E2F2/5/8 could serve as a potential prognostic biomarker and target for human ovarian cancer. *ERMP1*, a novel potential oncogene involved in the unfolded protein response and oxidative stress defense, is highly expressed in human cancers.[41] In addition, *CCNB2* methylation levels were shown to be significantly correlated with pathological M, T, and stage, whereas

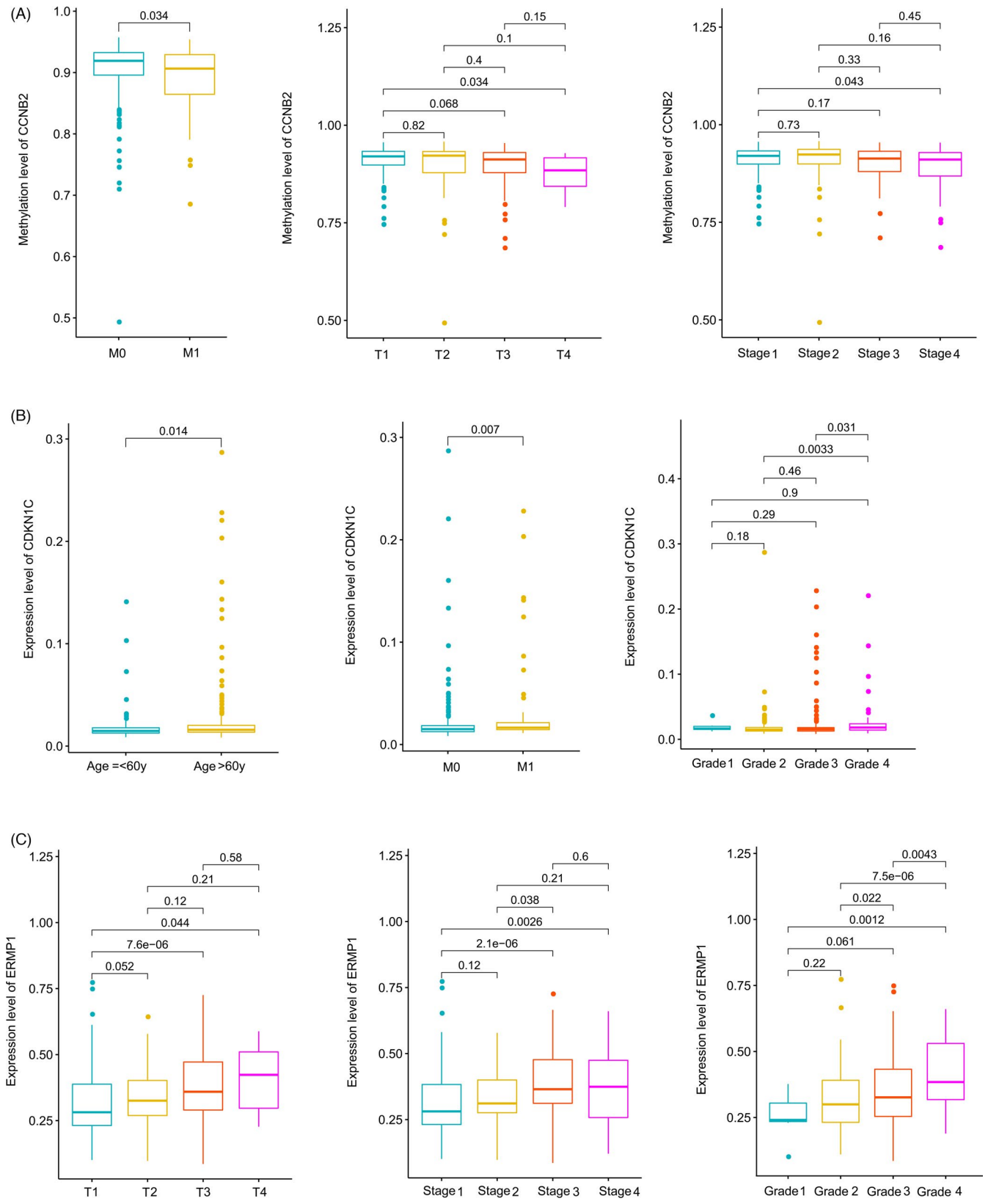
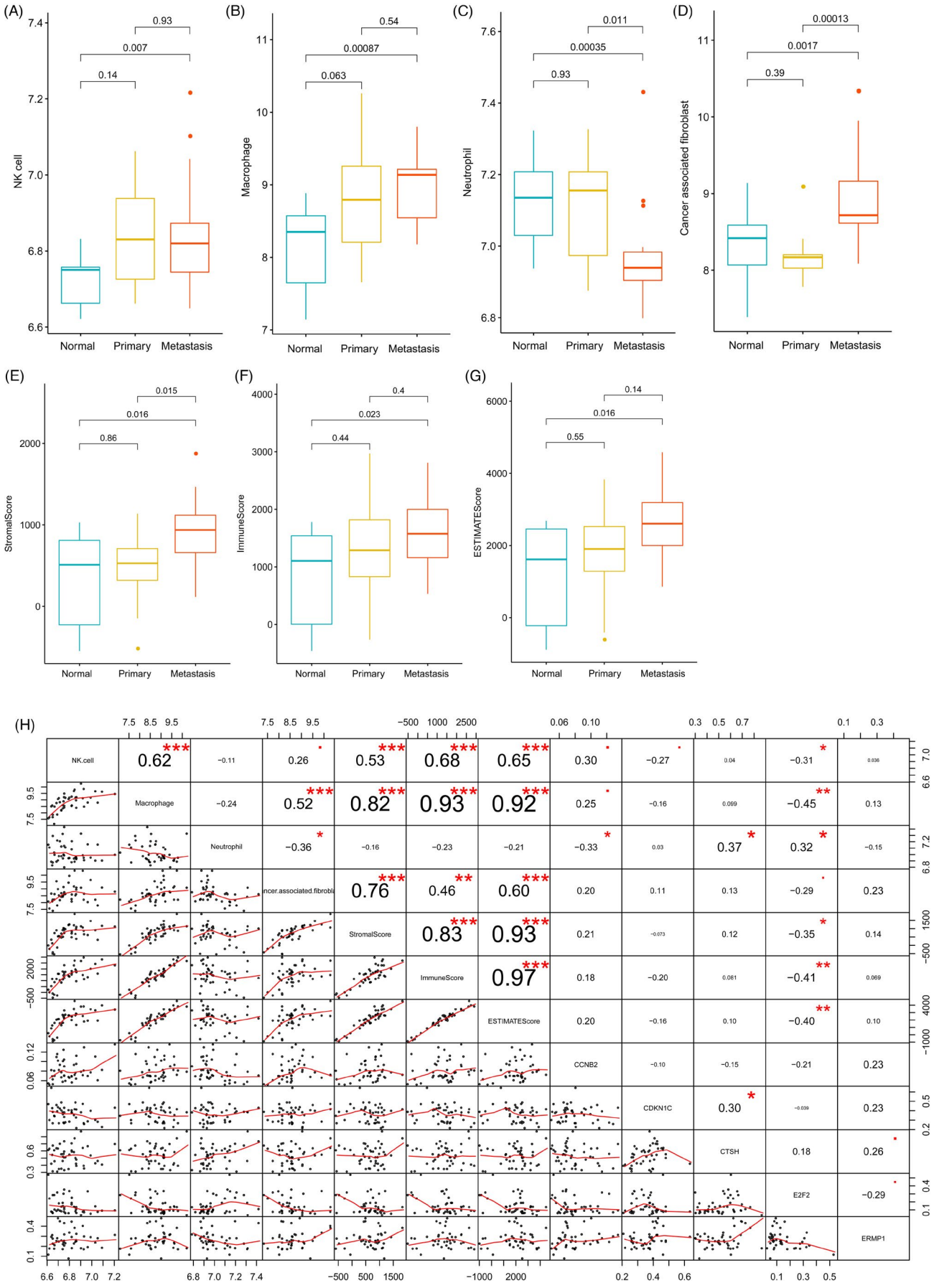


FIGURE 6 The differences between the distributions of methylation levels of the five optimized DEGs-DMGs in different clinical groups. (A) *CCNB2*. (B) *CDKN1C*. (C) *ERMP1*

FIGURE 7 Analysis of ccRCC progression and immunity. (A–G) Four immune cell types were found to be significantly correlated with ccRCC progression (normal, primary, and metastasis), including NK cells (A), macrophages (B), neutrophils (C), and cancer-associated fibroblasts (D). The stromal score (E), immune score (F), and ESTIMATE score (G) were shown to be correlated with ccRCC progression. (H) The correlation of the five optimized DEGs-DMGs and the four immune cell types, ESTIMATE score, immune score, and stromal score


(A)
NK cell
0.93
0.007
0.14
Normal
Primary
Metastasis
(B)
Macrophage
0.54
0.00087
0.063
(C)
Neutrophil
0.011
0.00035
0.93
(D)
Cancer associated fibroblast
0.00013
0.0017
0.39
(E)
StromalScore
0.015
0.016
0.86
(F)
ImmuneScore
0.4
0.023
0.44
(G)
ESTIMATEScore
0.14
0.016
0.55
(H)
NK.cell
Macrophage
Neutrophil
StromalScore
ImmuneScore
ESTIMATEScore
CCNB2
CDKN1C
CTSH
E2F2
ERMP1

the methylation levels of *CDKN1C* were significantly correlated with age, neoplasm histological grade, and pathological M. Finally, *ERMP1* methylation levels were found to be significantly correlated with neoplasm histological grade, pathological T, and stage. These findings further suggest that these five DEGs-DMGs are associated with the development and prognosis of ccRCC and could be used as potential therapeutic targets.

It has been reported that ccRCC is characterized by mutation of the von Hippel-Lindau (*VHL*) gene in factors governing the hypoxia signaling pathway[42,43], resulting in metabolic dysregulation, heightened angiogenesis, intratumoral heterogeneity, and tumor microenvironmental (TME) crosstalk. Thus, we speculated whether crosstalk also occurs between metabolic dysregulation, intratumoral heterogeneity, and TME in ccRCC owing to methylation. The GO functional analysis and KEGG enrichment analysis revealed that these 14 DEGs-DMGs were crucial in functions related to metabolic dysregulation. Moreover, TME is a complex integrated system categorized into the immune microenvironment and non-immune microenvironment.[44] Stromal components dominate the non-immune microenvironment, particularly cancer-associated fibroblasts.[45] Tumor-infiltrating immune cells in the TME have also been shown to critically relate cancer outcomes and play a vital role in tumor immunotherapy[46,47]. The present study also reports that four immune cell types (NK cells, macrophages, neutrophils, and cancer-associated fibroblasts) and the ESTIMATE, immune, and stromal scores are significantly correlated with ccRCC progression (normal, primary, and metastatic). NK cells are cytotoxic lymphocytes that are part of the innate immune system and are capable of killing virus-infected and/or cancerous cells.[48] Macrophages act as scavengers, modulating the immune response against pathogens and maintaining tissue homeostasis,[49] while tumor-associated macrophages exhibit both anti-inflammatory and pro-tumoral effects.[50] Neutrophils are the first responders to inflammation and infection and respond to diverse inflammatory cues, including cancer development.[51] Finally, cancer-associated fibroblasts, a type of perpetually activated fibroblasts, have been shown to have a strong tumor-modulating effect.[52] In addition, we also found that the five optimized DEGs-DMGs used in our prognostic model significantly correlate with these four immune cell types in addition to the ESTIMATE, immune, and stromal scores. For instance, *E2F2* was negatively correlated with NK cells, macrophages, cancer-associated fibroblasts, ESTIMATE score, immune score, and stromal score, while being positively correlated with neutrophils. Thus, the results of this study might reveal crosstalk between metabolic dysregulation and TME in ccRCC owing to methylation. While these findings essentially require further validation, these results also suggested that immunotherapy might serve as an effective treatment strategy for the treatment of ccRCC.

However, our study had some limitations. First, the prognostic model constructed from five optimized DEGs-DMGs was only validated using the GSE105288 dataset; thus, further verification using different datasets is vital. In addition, *in vitro* and *in vivo* studies should be performed in order to verify the methylation signatures and to help understand their functional role in ccRCC progression.

5 | CONCLUSION

This study has identified a five-gene methylation signature with significant predictive ability for ccRCC prognosis. In addition, it has highlighted potential targets for methylation-mediated immunotherapy for the treatment of ccRCC. It is imperative for future research to focus on the validation of these findings of this study through pre-clinical studies and clinical trials.

ACKNOWLEDGEMENTS

None.

CONFLICT OF INTEREST

The authors declare that they have no competing interests.

AUTHOR CONTRIBUTIONS

XJ involved in conception and design of the research; GX involved in acquisition of data; YG involved in analysis and interpretation of data; JL and LW involved in statistical analysis; SP involved in drafting the manuscript; YH, ZL, and WZ involved in revision of manuscript for important intellectual content. All authors have read and approved the final manuscript.

CONSENT FOR PUBLICATION

Not applicable.

DATA AVAILABILITY STATEMENT

All data generated or analyzed during this study are included in this published article.

ORCID

Shouhua Pan https://orcid.org/0000-0001-6960-8418

REFERENCES

1. Díaz-Montero CM, Rini BI, Finke JH. The immunology of renal cell carcinoma. *Nat Rev Nephrol*. 2020;16(12):721-735.
2. Gray RE, Harris GT. Renal cell carcinoma: diagnosis and management. *Am Fam Physician*. 2019;99(3):179-184.
3. Wolf MM, Kimryn Rathmell W, Beckermann KE. Modeling clear cell renal cell carcinoma and therapeutic implications. *Oncogene*. 2020;39(17):3413-3426.
4. Song E, Song W, Ren M, et al. Identification of potential crucial genes associated with carcinogenesis of clear cell renal cell carcinoma. *J Cell Biochem*. 2018;119(7):5163-5174.
5. Menezo Y, Clement P, Clement A, Elder K. Methylation: an ineluctable biochemical and physiological process essential to the transmission of life. *Int J Mol Sci*. 2020;21(23):9311.
6. Botezatu A, Iancu IV, Plesa A, et al. Methylation of tumour suppressor genes associated with thyroid cancer. *Cancer Biomark*. 2019;25(1):53-65.
7. Koch A, Joosten SC, Feng Z, et al. Analysis of DNA methylation in cancer: location revisited. *Nat Rev Clin Oncold*. 2018;15(7):459-466.

8. McCabe MT, Mohammad HP, Barbash O, Kruger RG. Targeting histone methylation in cancer. *Cancer J*. 2017;23(5):292-301.
9. Chen X, Zhang J, Ruan W, et al. Urine DNA methylation assay enables early detection and recurrence monitoring for bladder cancer. *J Clin Investig*. 2020;130(12):6278-6289.
10. Pan Y, Liu G, Zhou F, Su B, Li Y. DNA methylation profiles in cancer diagnosis and therapeutics. *Clin Exp Med*. 2018;18(1):1-14.
11. Delpu Y, Cordelier P, Cho WC, Torrisani J. DNA methylation and cancer diagnosis. *Int J Mol Sci*. 2013;14(7):15029-15058.
12. Nam HY, Chandrashekar DS, Kundu A, et al. Integrative epigenetic and gene expression analysis of renal tumor progression to metastasis. *Mol Cancer Res*. 2019;17(1):84-96.
13. Barrett T, Troup DB, Wilhite SE, et al. NCBI GEO: mining tens of millions of expression profiles–database and tools update. *Nucleic Acids Res*. 2007;35(Database):D760-D765.
14. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43(7):e47.
15. Wang L, Cao C, Ma Q, et al. RNA-seq analyses of multiple meristems of soybean: novel and alternative transcripts, evolutionary and functional implications. *BMC Plant Biol*. 2014;14:169.
16. da Huang W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 2009;4(1):44-57.
17. Kumar L, Futschik ME. Mfuzz: a software package for soft clustering of microarray data. *Bioinformation*. 2007;2(1):5-7.
18. Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003;13(11):2498-2504.
19. Wang P, Wang Y, Hang B, Zou X, Mao JH. A novel gene expression-based prognostic scoring system to predict survival in gastric cancer. *Oncotarget*. 2016;7(34):55343-55351.
20. Tibshirani R. The lasso method for variable selection in the Cox model. *Stat Med*. 1997;16(4):385-395.
21. Shi J, Jiang D, Yang S, et al. LPAR1, correlated With immune infiltrates, is a potential prognostic biomarker in prostate cancer. *Front Oncol*. 2020;10:846.
22. Hu D, Zhou M, Zhu X. Deciphering immune-associated genes to predict survival in clear cell renal cell cancer. *Biomed Res Int*. 2019;2019:2506843.
23. Wang J, Yang J, Li D, Li J. Technologies for targeting DNA methylation modifications: basic mechanism and potential application in cancer. *Biochim Biophys Acta*. 2021;1875(1):188454.
24. Sherr CJ. Cancer cell cycles. *Science*. 1996;274(5293):1672-1677.
25. Piezzo M, Cocco S, Caputo R, et al. Targeting cell cycle in breast cancer: CDK4/6 Inhibitors. *Int J Mol Sci*. 2020;21(18):6479.
26. Hao JF, Ren KM, Bai JX, et al. Identification of potential biomarkers for clear cell renal cell carcinoma based on microRNA-mRNA pathway relationships. *J Cancer Res Ther*. 2014;10(Suppl):C167-177.
27. Jalali S, Shi J, Buko A, et al. Increased glutathione utilization augments tumor cell proliferation in Waldenstrom Macroglobulinemia. *Redox Biol*. 2020;36:101657.
28. Desideri E, Ciccarone F, Ciriolo MR. Targeting glutathione metabolism: partner in crime in anticancer therapy. *Nutrients*. 2019;11(8):1926.
29. Grimm D, Bauer J, Wise P, et al. The role of SOX family members in solid tumours and metastasis. *Semin Cancer Biol*. 2020;67(Pt 1):122-153.
30. Taniguchi K, Karin M. NF-κB, inflammation, immunity and cancer: coming of age. *Nat Rev Immunol*. 2018;18(5):309-324.
31. Djordjević G, Matusan-Ilijas K, Sinozić E, et al. Relationship between vascular endothelial growth factor and nuclear factor-kappaB in renal cell tumors. *Croat Med J*. 2008;49(5):608-617.
32. Meng J, Liu K, Shao Y, et al. ID1 confers cancer cell chemoresistance through STAT3/ATF6-mediated induction of autophagy. *Cell Death Dis*. 2020;11(2):137.
33. Gao Z, Zhang D, Duan Y, et al. A five-gene signature predicts overall survival of patients with papillary renal cell carcinoma. *PLoS One*. 2019;14(3):e0211491.
34. Gao X, Chen Y, Chen M, Wang S, Wen X, Zhang S. Identification of key candidate genes and biological pathways in bladder cancer. *PeerJ*. 2018;6:e6036.
35. Wang X, Xiao H, Wu D, Zhang D, Zhang Z. miR-335-5p regulates cell cycle and metastasis in lung adenocarcinoma by targeting CCNB2. *Onco Targets Ther*. 2020;13:6255-6263.
36. Qiu Z, Zhu W, Meng H, et al. CDYL promotes the chemoresistance of small cell lung cancer by regulating H3K27 trimethylation at the CDKN1C promoter. *Theranostics*. 2019;9(16):4717-4729.
37. Na H, Li X, Zhang X, et al. lncRNA STEAP3-AS1 modulates cell cycle progression via affecting CDKN1C expression through STEAP3 in colon cancer. *Mol Ther Nucleic Acids*. 2020;21:480-491.
38. Cao J, Wang H, Liu G, et al. LBX2-AS1 promotes ovarian cancer progression by facilitating E2F2 gene expression via miR-455-5p and miR-491-5p sponging. *J Cell Mol Med*. 2021;25(2):1178-1189.
39. Liao P, Han S, Qu H. Expression, prognosis, and immune infiltrates analyses of E2Fs in human brain and CNS cancer. *Biomed Res Int*. 2020;2020:6281635.
40. Zhou Q, Zhang F, He Z, Zuo MZ. E2F2/5/8 Serve as potential prognostic biomarkers and targets for human ovarian cancer. *Front Oncold*. 2019;9:161.
41. Yang Q, Chu W, Yang W, et al. Identification of RNA transcript makers associated with prognosis of kidney renal clear cell carcinoma by a competing endogenous RNA network analysis. *Front Genet*. 2020;11:540094.
42. Gnarra JR, Tory K, Weng Y, et al. Mutations of the VHL tumour suppressor gene in renal carcinoma. *Nat Genet*. 1994;7(1):85-90.
43. Zhang J, Zhang Q. VHL and hypoxia signaling: beyond HIF in cancer. *Biomedicines*. 2018;6(1):35.
44. Quail DF, Joyce JA. Microenvironmental regulation of tumor progression and metastasis. *Nat Med*. 2013;19(11):1423-1437.
45. Chen X, Song E. Turning foes to friends: targeting cancer-associated fibroblasts. *Nat Rev Drug Discovery*. 2019;18(2):99-115.
46. Chandran SS, Klebanoff CA. T cell receptor-based cancer immunotherapy: emerging efficacy and pathways of resistance. *Immunol Rev*. 2019;290(1):127-147.
47. van Montfoort N, Borst L, Korrer MJ, et al. NKG2A blockade potentiates CD8 T cell immunity induced by cancer vaccines. *Cell*. 2018;175(7):1744-1755.e15.
48. Myers JA, Miller JS. Exploring the NK cell platform for cancer immunotherapy. *Nat Rev Clin Oncol*. 2021;18(2):85-100.
49. Mehla K, Singh PK. Metabolic regulation of macrophage polarization in cancer. *Trends Cancer*. 2019;5(12):822-834.
50. Najafi M, Hashemi Goradel N, Farhood B, et al. Macrophage polarity in cancer: a review. *J Cell Biochem*. 2019;120(3):2756-2765.
51. Mollinedo F. Neutrophil degranulation, plasticity, and cancer metastasis. *Trends Immunol*. 2019;40(3):228-242.
52. Nurmik M, Ullmann P, Rodriguez F, Haan S, Letellier E. In search of definitions: cancer-associated fibroblasts and their markers. *Int J Cancer*. 2020;146(4):895-905.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Jing X, Xu G, Gong Y, et al. A five-gene methylation signature predicts overall survival of patients with clear cell renal cell carcinoma. *J Clin Lab Anal*. 2021;35:e24031. https://doi.org/10.1002/jcla.24031