Integrative Analysis of DNA Methylation Identified 12 Signature Genes Specific to Metastatic ccRCC

Siwei Qian[1], Si Sun[1], Lei Zhang[1,2], Shengwei Tian[1], Kai Xu[3], Guangyuan Zhang[1,2]* and Ming Chen[1,2]*

[1] Department of Urology, Zhongda Hospital, Southeast University, Nanjing, China, [2] Institute of Urology, School of Medicine, Southeast University, Nanjing, China, [3] Department of Urology, Changzhou No. 2 People's Hospital, Changzhou, China

Background: Abnormal epigenetic alterations can contribute to the development of human malignancies. Identification of these alterations for early screening and prognosis of clear cell renal cell carcinoma (ccRCC) has been a highly sought-after goal. Bioinformatic analysis of DNA methylation data provides broad prospects for discovery of epigenetic biomarkers. However, there is short of exploration of methylation-driven genes of ccRCC.

Methods: Gene expression data and DNA methylation data in metastatic ccRCC were sourced from the Gene Expression Omnibus (GEO) database. Differentially methylated genes (DMGs) at 5′-C-phosphate-G- 3′ (CpG) sites and differentially expressed genes (DEGs) were screened and the overlapping genes in DMGs and DEGs were then subject to gene set enrichment analysis. Next, the weighted gene co-expression network analysis (WGCNA) was used to search hub DMGs associated with ccRCC. Cox regression and ROC analyses were performed to screen potential biomarkers and develop a prognostic model based on the screened hub genes.

Results: Three hundred and fourteen overlapping DMGs were obtained from two independent GEO datasets. The turquoise module contained 79 hub DMGs, which represent the most significant module screened by WGCNA. Furthermore, a total of 12 hub genes (*CETN3*, *DCAF7*, *GPX4*, *HNRNPA0*, *NUP54*, *SERPINB1*, *STARD5*, *TRIM52*, *C4orf3*, *C12orf51*, and *C17orf65*) were identified in the TCGA database by multivariate Cox regression analyses. All the 12 genes were then used to generate the model for diagnosis and prognosis of ccRCC. ROC analysis showed that these genes exhibited good diagnostic efficiency for metastatic and non-metastatic ccRCC. Furthermore, the prognostic model with the 12 methylation-driven genes demonstrated a good prediction of 5-year survival rates for ccRCC patients.

Conclusion: Integrative analysis of DNA methylation data identified 12 signature genes, which could be used as epigenetic biomarkers for prognosis of metastatic ccRCC. This prognostic model has a good prediction of 5-year survival for ccRCC patients.

Keywords: clear cell renal cell carcinoma, data integration, DNA methylation, gene expresssion, ROC (receiver operating characteristic curve)

BACKGROUND

Clear cell renal carcinoma (ccRCC) is the major type of renal tumor in the human urinary system (1). Many patients with ccRCC have distant metastasis to lymphoid or other organs (2). The 5-year survival rate was lower in patients with distant metastatic ccRCC than with non-metastatic ccRCC (3). Over the past decade, the therapeutic strategies for advanced ccRCC have evolved rapidly from a non-specific immune approach to targeted therapy against vascular endothelial growth factor (VEGF), and recently to novel immunotherapy. Multiple agents like immune checkpoint inhibitors (ICIs) and tyrosine kinase inhibitors (TKIs) show promising results in the clinical trials and have been approved as the second-line even the first-line treatment for advanced ccRCC (4–7). Moreover, targeted combination therapies have been explored as the first-line treatment in pioneer trials (8).

However, to enhance the efficacy and prognosis always underscores a need for new therapeutic targets and drug combinations (9). An increasing number of studies prove that epigenetic alterations such as DNA methylation and histone modifications are associated with cancer progression and occurrence. Therefore, they could be exploited as biomarkers for cancer diagnosis and prognosis (10–14). However, epigenetic biomarkers underlying metastatic ccRCC remain to be elucidated. Epigenetic biomarkers have been testified in certain malignancies and allowed their potential use in differentiation between metastatic and non- metastatic ccRCC. In this study, we profiled methylation patterns of metastatic and primary non-metastatic ccRCC and identified potential biomarkers associated with the cancer progression. We measured the gene methylation levels at CpG sites throughout the genome and found that a subset of genes presented with aberrant methylation and expression in metastatic ccRCC when compared with primary non-metastatic ccRCC. We further created the criteria to screen hub DMGs and defined the pathological stages of ccRCC accordingly. Finally, we applied the survival analysis to evaluate the hub DMGs as potential biomarkers for the prognosis of ccRCC patients.

METHODS

Data Collection

The gene methylation datasets (GSE113501 and GSE105260) were downloaded from the GEO database. GSE113501 and GSE105260 were generated on the platform GPL13534 (Illumina HumanMethylation450 BeadChip). In total, 54 metastatic tumor samples and 96 non-metastatic tumor samples were included in the study. GSE105261 was used for screening of DEGs in non-metastatic and metastatic ccRCC. TCGA-KIRC dataset was used for analysis of correlation between gene expression and methylation levels. All ccRCC-related clinical data in this study were obtained from the GEO and the Cancer Genome Atlas (TCGA) database.

Differential Methylation Analysis

First, the missing methylation values among all samples were removed. To identify differentially methylated CpGs, two-sample independent t-test was applied to compare primary non-metastatic ccRCC and metastatic ccRCC data. Bonferroni procedure, together with transformed β values, was used to adjust crude p-values in multiple comparisons. The β value was calculated by the formula $\beta = M/(M + U + a)$, where M and U represent the methylated and unmethylated signal intensities, respectively. a is usually set as 100 to stabilize β values when M and U are too small (15). β average were used for evaluation of global and regional CpG methylations. The delta β value was calculated to evaluate the methylation difference between metastatic and non-metastatic ccRCC. M and β values were used to identify the DMGs and intergenic CpG sites. The adjusted $p < 0.05$ and delta β values in 4/10 quantile were used as the cutoff criteria to screen candidate hub genes or intergenic CpG sites. R packages "limma" and "lumi" were used for differential analysis and conversion between β and M values, respectively (16–22). The DMGs were classified into six types according to the methylation patterns.

GENE SET ENRICHMENT ANALYSIS

Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses were implemented by DAVID online tools (23, 24). $p < 0.05$ was considered statistically significant. The GO term enrichment analysis included three categories: biological process (BP), molecular function (MF), and cellular component (CC). R package "Venn Diagram" was used to illustrate the overlapping genes in DMGs and DEGs.

Correlation Between DNA Methylation and Gene Expression

We examined the correlation between DNA methylation and gene expression based on the data from the TCGA database. Spearman correlation coefficients were used to evaluate the expression value and the methylation level at all genes CpG sites. The correlation was regarded significant if the absolute value of correlation coefficient was >0.3 and FDR was <0.05. Only the overlapping genes in DMGs and DEGs were evaluated for their biological functions by gene regulation analysis.

Weighted Co-expression Network Construction Analysis (WCGNA)

R package "WGCNA" was used for constructing co-expression networks among genes across microarray samples (25). To begin with, the power of β (soft thresholding) was defined to ensure a standard scale-free network. An adjacency matrix was then constructed with the values of adjacencies between each pair of node genes in the network and their Pearson's correlation coefficients. The adjacency matrix was used to calculate the topological overlap matrix (TOM) and the
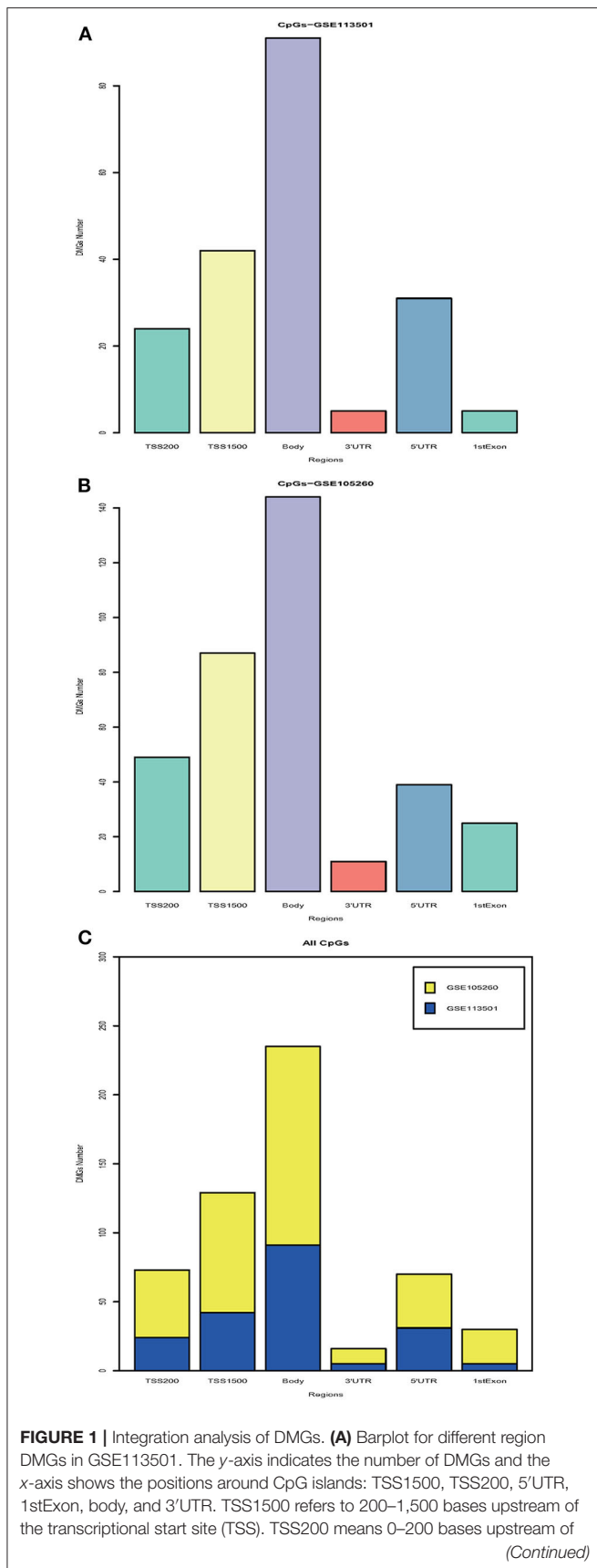
FIGURE 1 | TSS. 5′UTR stands for the 5′untranslated region located between the TSS and the ATG start site. 1stExon is short for the first exon of the gene. Body is the region between ATG start site and stop codon. 3′UTR is short for 3′untranslated region that is between stop codon and poly-A tail. **(B)** Barplot for different region DMGs in GSE105260. **(C)** Barplot for all DMGs based on different regions. The blue module stands for DMGs in GSE113501. The yellow module represents for DMGs in GSE105260.

corresponding dissimilarity (1-TOM). Finally, based on TOM-based dissimilarity measures, the dendrogram was constructed by average linkage hierarchical clustering. Highly similar modules were clustered and then merged with a height cut-off of 0.25.

Identification of Modules Corresponding to Clinical Traits

By WGCNA, the modules most relevant to the clinical phenotypes of interest could be identified. Here, we differentiated clinical phenotypes by pathological stages. As a principal component of the gene expression matrix, module eigengene (ME) was used to identify modules corresponding to clinically significant traits. Gene significance was evaluated by log p-value of each gene in the linear regression between gene expression and pathological progression. The module significance was used to evaluate the correlation between module expression patterns and clinical traits (age, gender, survival time, and disease stage, etc.). In general, a module was regarded significantly correlating with certain clinical traits if its absolute module significance ranked the highest among others.

Identification of DMGs Associated With Overall Survival Prognosis and Validation of Potential Epigenetic Biomarkers by Receiver Operating Characteristic (ROC) Analysis

To identify module genes associated with prognosis of ccRCC patients, we analyzed gene methylation data in the turquoise module from the Cancer Genome Atlas Kidney Renal Clear Cell Carcinoma (TCGA-KIRC) data collection. Univariate Cox regression was used to evaluate the prognostic power of 79 individual probes at CpG sites. Furthermore, multivariate Cox regression analysis was performed to evaluate whether all genes corresponding to individual CpG sites could be used as independent prognostic factors for patient's survival. $p < 0.05$ indicates statistical significance. ROC analysis was used to validate the predictive accuracy of potential epigenetic biomarkers, which were remarkably methylated in the multivariable Cox regression model. The area under the curve (AUC) was calculated to measure the quality of the classifier with R package "pROC".

Development and Evaluation of Prognostic Model

The genes with significant p-values by multivariate Cox regression were selected to develop a prognostic model. The risk

FIGURE 1 | Integration analysis of DMGs. **(A)** Barplot for different region DMGs in GSE113501. The y-axis indicates the number of DMGs and the x-axis shows the positions around CpG islands: TSS1500, TSS200, 5′UTR, 1stExon, body, and 3′UTR. TSS1500 refers to 200–1,500 bases upstream of the transcriptional start site (TSS). TSS200 means 0–200 bases upstream of
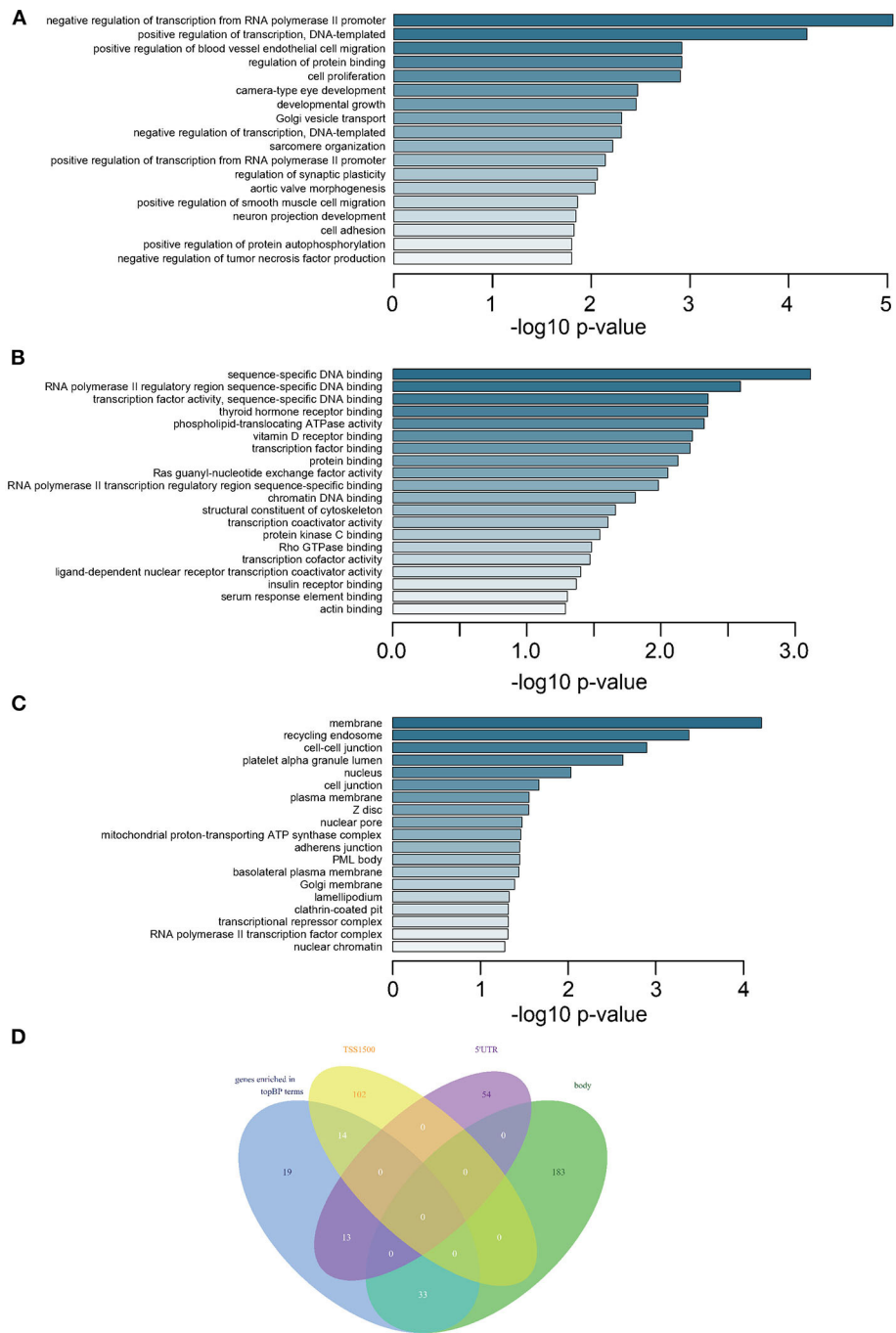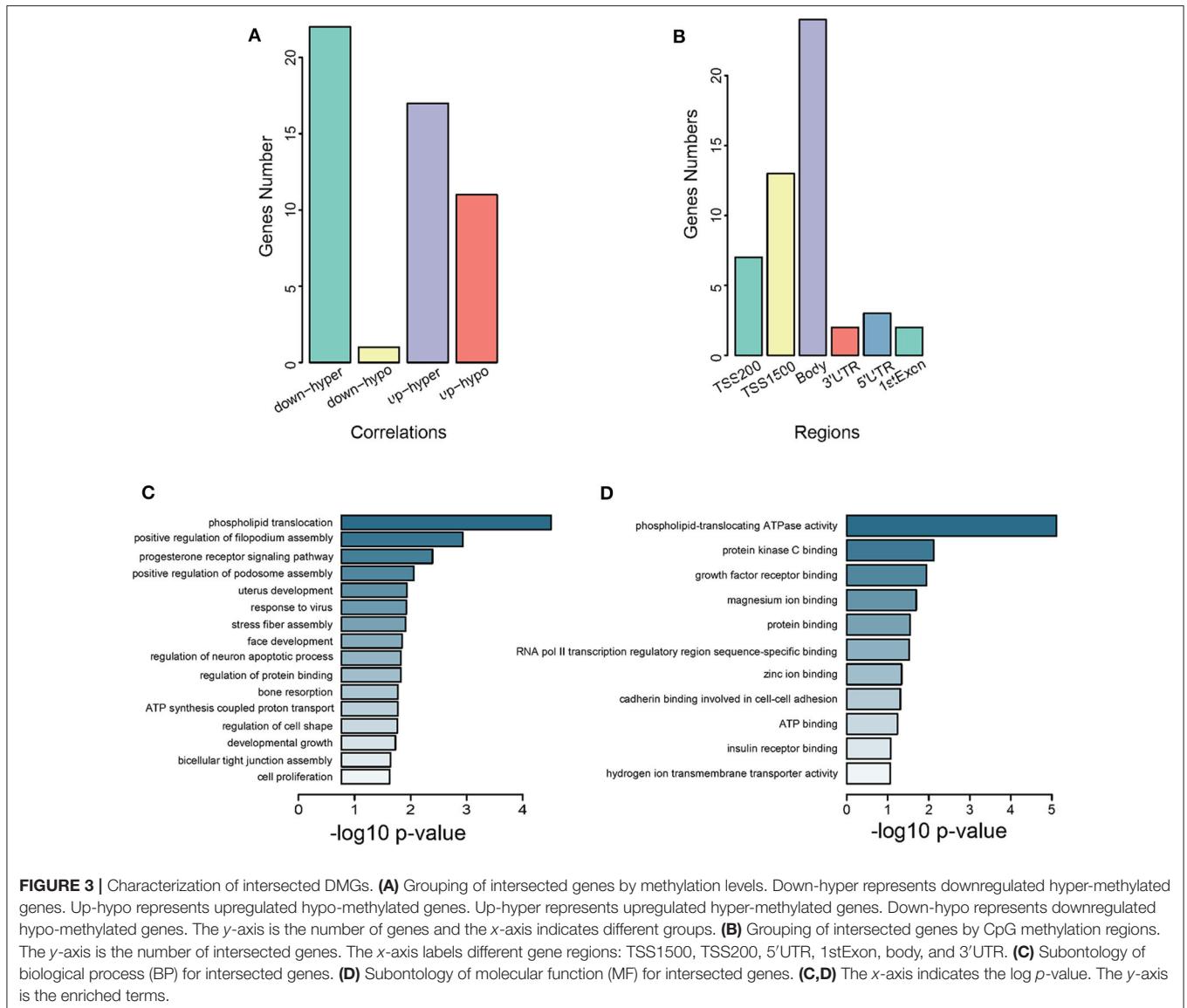
(Continued)

FIGURE 2 | Enrichment analysis of DMGs and intersection of methylated CpGs in different regions. **(A)** Subontology of biological process (BP). **(B)** Subontology of molecular function (MF). **(C)** Subontology of cellular components (CC). **(A–C)** The x-axis indicates the log p-value. The y-axis is the enriched terms. **(D)** Venn plot for DMGs under BP terms based on different CpG methylation regions. The numbers on the diagram represent the DMG numbers in a specific region or multiple regions. The regions is as indicated.

score was calculated with regression coefficients multiplied by methylation levels. The median risk score was set as the cutoff value, by which 307 ccRCC patients from the TCGA database were divided into the high-risk group and the low-risk group. Clinical traits were used as an independent variable and the

overall survival (OS) as the dependent variable to calculate the hazard ratio (HR). ROC curve was plotted with the R package "survivalROC" to confirm the predictive power of the 12-gene prognostic model and assess the probability of 1, 3, and 5-year OS for ccRCC patients.

FIGURE 3 | Characterization of intersected DMGs. **(A)** Grouping of intersected genes by methylation levels. Down-hyper represents downregulated hyper-methylated genes. Up-hypo represents upregulated hypo-methylated genes. Up-hyper represents upregulated hyper-methylated genes. Down-hypo represents downregulated hypo-methylated genes. The y-axis is the number of genes and the x-axis indicates different groups. **(B)** Grouping of intersected genes by CpG methylation regions. The y-axis is the number of intersected genes. The x-axis labels different gene regions: TSS1500, TSS200, 5′UTR, 1stExon, body, and 3′UTR. **(C)** Subontology of biological process (BP) for intersected genes. **(D)** Subontology of molecular function (MF) for intersected genes. **(C,D)** The x-axis indicates the log p-value. The y-axis is the enriched terms.

RESULTS

Identification of DMGs in Metastatic ccRCC

To identify potential DMG biomarkers for prognosis of ccRCC, we performed a DNA methylation profiling on 150 tumor tissues. All the relevant clinical data were collected from the GEO datasets: GSE113501 (including 28 metastatic ccRCC tissues and 87 non-metastatic ccRCC tissues) and GSE105260 (including 26 metastatic ccRCC tissues and 9 non-metastatic ccRCC tissues). In total, 237 CpG sites in GSE113501 and 462 CpG sites in GSE105260 were identified to have differential methylations (**Supplementary Tables 1, 2**), which corresponded to 198 and 355 genes, respectively (**Supplementary Tables 3, 4**). GO analysis showed that 552 DMGs were linked to the biological pathways that involved in transcriptional regulation of cancers, cell cycle and metabolism, such as bladder cancer (**Supplementary Table 5**).

With the DNA methylation annotation file (GPL13534), we examined CpG methylation in different regions: TSS1500, TSS200, 5′UTR, 1stExon, body, and 3′UTR. Over 80 methylated CpG sites located in the body region (**Figure 1A**, GSE113501 and **Figure 1B**, GSE105260). Totally, more than 200 differentially methylated CpG sites located in the TSS1500 and body region (**Figure 1C**). By contrast, <25 methylated CpG sites appeared at 3′UTR region.

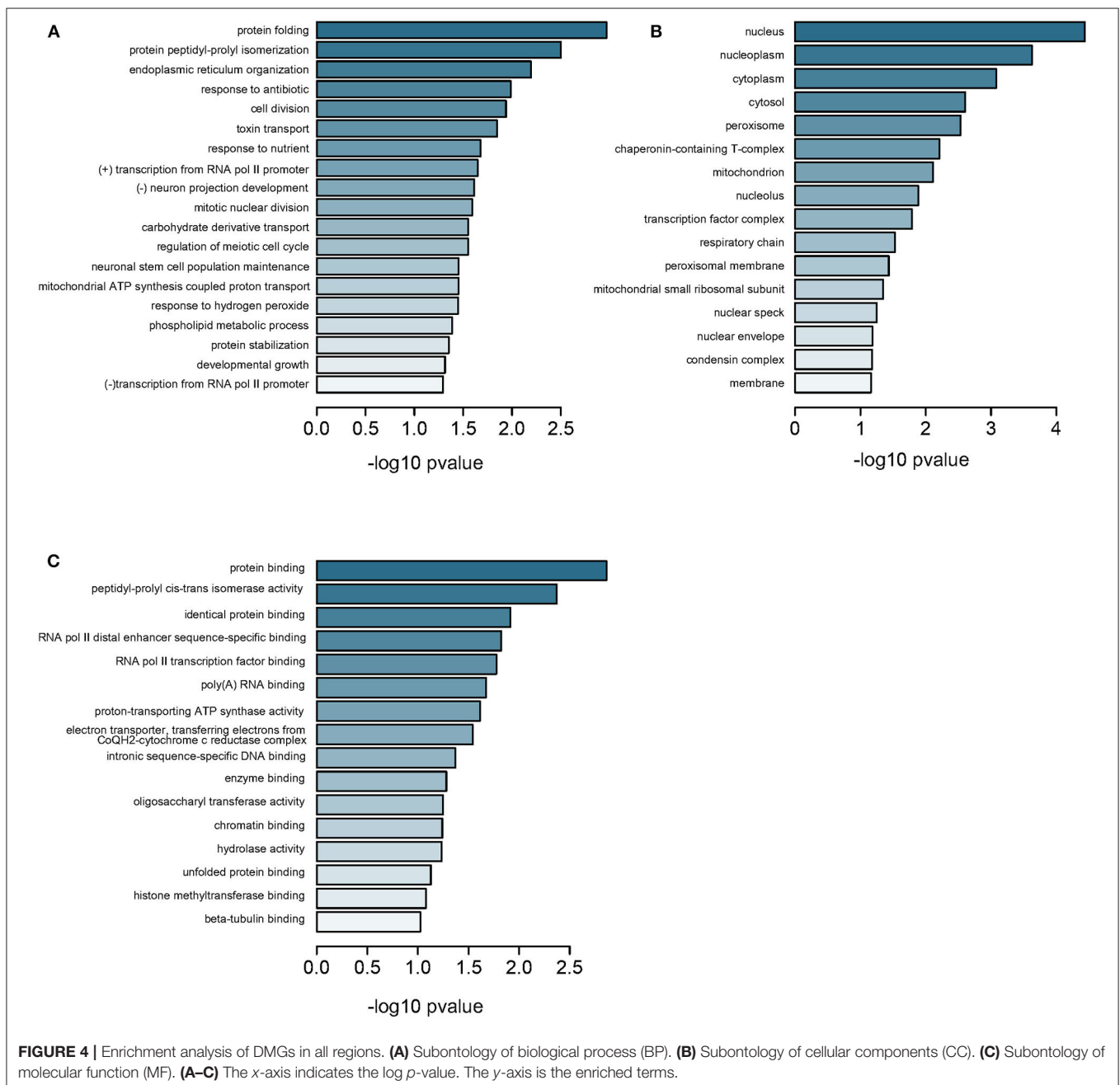### DMGs Involved in Transcriptional Regulation

GO analysis revealed that most DMGs were annotated to biological process (BP) and molecular function (MF) terms indicating transcriptional regulation: positive or negative regulation of transcription for various BPs (**Figure 2A** and **Supplementary Table 6**) and transcription factor bindings

(**Figure 2B**). In addition, the ontology of cell component (CC) pointed to key cellular structures responsible for cell-cell interaction and cell activities (**Figure 2C**). Venn diagram showed that genes representing top BP terms were mostly methylated in the regions of body, 5′UTR, and TSS1500 (**Figure 2D**).

Overlapping Genes in DMGs and DEGs

We identified 51 genes overlapping between DEGs from GSE105261 and DMGs from GSE105260 (**Supplementary Figure 1** and **Supplementary Table 7**), which were classified into four types according to their methylation

and expression levels, namely downregulated hypermethylated or hypomethylated, and upregulated hypermethylated or hypomethylated. Noticeably, downregulated hypomethylated genes accounted for a much lower proportion (**Figure 3A**). The CpG methylation of these genes mainly took place in the body and TSS1500 regions (**Figure 3B**). GO analysis also revealed that they were related to transcriptional factor assembly and receptor-mediated signaling (**Figure 3C**), and key protein bindings (**Figure 3D**). Moreover, 10 genes were screened against the TCGA-KIRC database, which showed a significant correlation between methylation level and expression level in ccRCC patients (**Supplementary Figure 2**).

FIGURE 4 | Enrichment analysis of DMGs in all regions. **(A)** Subontology of biological process (BP). **(B)** Subontology of cellular components (CC). **(C)** Subontology of molecular function (MF). **(A–C)** The x-axis indicates the log p-value. The y-axis is the enriched terms.

Identification of DMGs Associated With Metastatic ccRCC Throughout the Genome

Further, we measured all CpG methylations throughout the genome and identified 6,300 genes from GSE113501 (**Supplementary Table 8**) and 799 genes from GSE105260 (**Supplementary Table 9**), respectively. Three hundred and fourteen genes were overlapped between the two subsets (**Supplementary Figure 3**). As previously, these genes were annotated to BPs relating to protein folding and transportation (**Figure 4A** and **Supplementary Table 10**) and CCs to
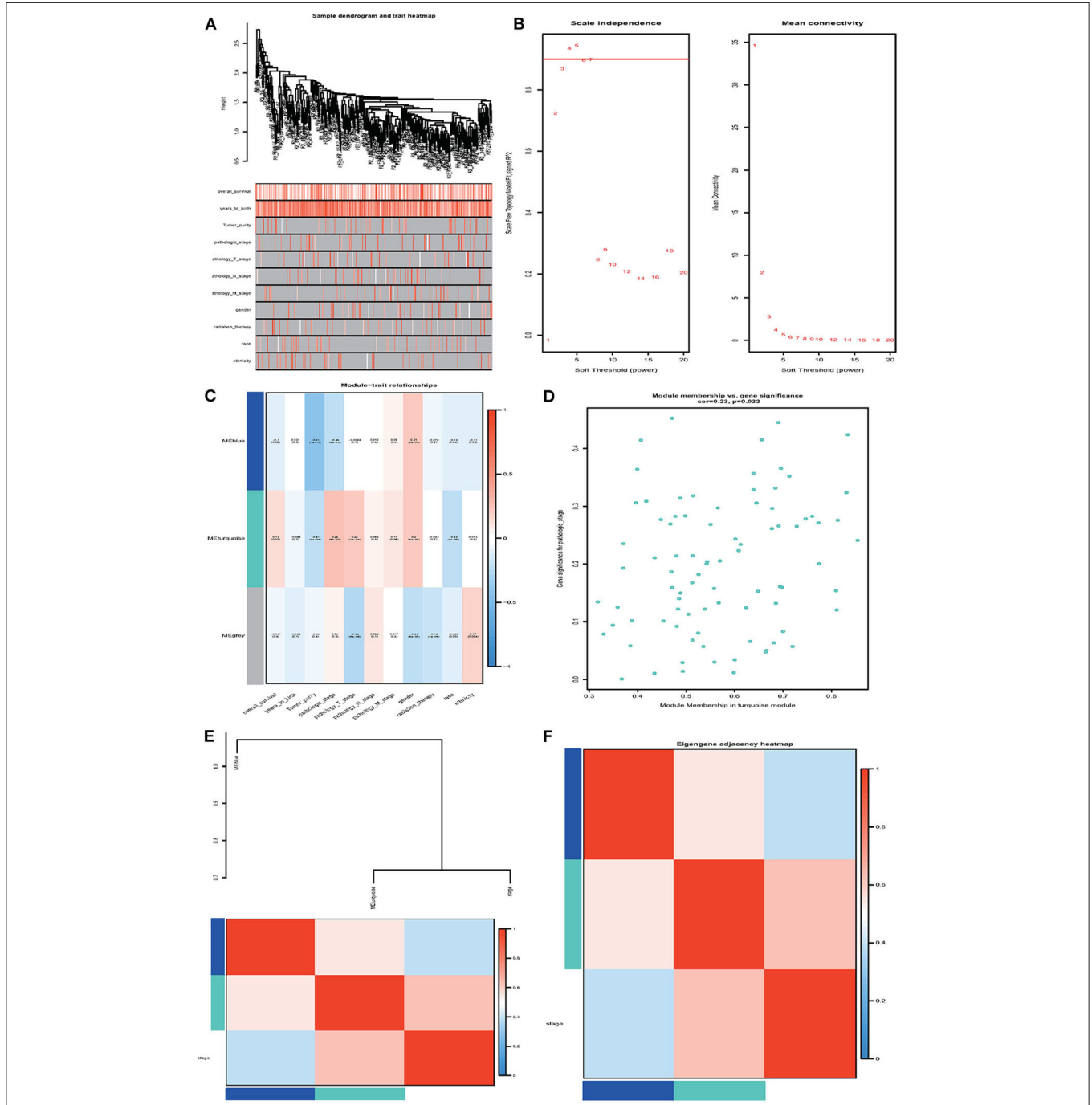

FIGURE 5 | Weighted gene co-expression network analysis (WGCNA). **(A)** Sample dendrogram and trait heatmap. **(B)** Analysis of the scale-free fit index and the mean connectivity for various soft-thresholding powers (β). **(C)** Heatmap of correlation between module eigengenes and clinical information. **(D)** Module membership in the turquoise module. **(E,F)** The correlation between each module and clinical characteristic pathological stage was demonstrated based on Eigen gene. Blue represents a negative correlation, while red represents a positive correlation.

components in nucleus (**Figure 4B**), as well as MFs to key element bindings (**Figure 4C**). When comparing the 314 DMGs with 3,002 DEGs filtered from GSE105261, we found 44 genes were overlapped (**Supplementary Figures 4** and **Supplementary Table 11**). Most of them were downregulated-hypermethylated genes (**Supplementary Figure 5**), indicating their domination in the progress of ccRCC. The TCGA-KIRC dataset and Spearman test were applied for correlation analysis, which identified top 10 genes with significant correlation between methylation and expression (**Supplementary Table 12** and **Supplementary Figure 6**).

Identification of Hub DMGs

Next, we performed the WGCNA analysis to identify potential molecular modules that could characterize the pathological stages of ccRCC. The WGCNA network was constructed with 308 DMGs screened out by methylation level changes in all CpG island regions. The DMGs irrelevant to ccRCC clinical features were removed (**Supplementary Table 13**). The dendrogram and traits of all samples were illustrated (**Figure 5A**) and the soft threshold power was set as five to ensure a scale-free network (**Figure 5B**). Three WCGNA modules were identified as blue, turquoise, and gray (**Figure 5C**). The blue module showed to be negatively correlated with pathological stages while turquoise module was positively. The turquoise module harboring 79 genes was selected for further analysis as it was the most relevant to clinical traits (**Supplementary Tables 14** and **Data Sheet 2**). The correlation coefficients with p-values, as well as heatmap of each module were calculated, respectively (**Figures 5D–F**) and the 79 relevant genes were considered as hub DMGs.

Identification of 12 Signature DMGs

Among the 79 hub DMGs, 43 CpG sites were further identified as potential DNA methylation biomarkers for prediction of overall survival of ccRCC patients by univariate Cox regression analysis. The genes such as *TGDS*, *TCTE3*, *TOPBP1*, *SNX14*, and *PHIP* showed their expression were strongly associated with the overall survival in ccRCC patients (**Supplementary Data Sheet 1**). By contrast, the multivariate cox analysis revealed 12 DMGs (*CETN3*, *DCAF7*, *GPX4*, *HNRNPA0*, *NUP54*, *SERPINB1*, *STARD5*, *TRIM52*, *C4orf3*, *C12orf51*, *C17orf65*, and *C21orf45*) significantly affected overall survival of ccRCC patients (**Table 1**). The module' risk score was significant (**Figure 6A**) and survival analysis exhibited hypermethylation signature of the 12 genes significantly correlated with poor prognosis ($p < 0.001$) (**Figure 6B**). Moreover, we examined methylation levels of the 12 genes in normal kidney tissues. The results showed that there was no significant difference between normal kidney tissues and non-metastatic ccRCC kidney tissues, but for metastatic renal cancer. Above all, all the analyses suggested that the multi-gene methylation signature may be associated with ccRCC metastasis (**Supplementary Data Sheet 3**).

Next, the ROC AUC was used to evaluate predictive powers of the prognostic models constructed with single signature gene or multiple signature genes. The ROC curves indicated that *C17orf65*, *HNRNPA0*, and S*TARD5* exhibited better

TABLE 1 | The results of multivariate cox regression analysis.

Variables	Overall Survival	
	HR (95% CI)	*P*
Clinical parameters		
Gender	1.577 (0.62598–1.19)	0.369
Age, year	1.044 (1.01861–1.046)	2.73e−06***
Race	9.396 (0.03915–2.076)	0.215
Tumor purity	0.1471 (0.27283–2.591)	0.763
Pathologic stage	0.1484 (0.24456–0.491)	2.53e−09***
Analysis parameters		
CETN3	7.82E+35 (5.835e+07–1.049e+64)	0.012380*
DCAF7	1.42E+150 (9.363e+04–2.149e+295)	0.042654*
GPX4	6.37E−40 (1.197e−67–3.385e−12)	0.005592**
HNRNPA0	2.89E−38 (2.574e−65–3.255e−11)	0.006532**
NUP54	2.18E+57 (1.668e+15–2.838e+99)	0.007622**
SERPINB1	2.36E+08 (5.448e+02–1.022e+14)	0.003598**
STARD5	1.32E+16 (1.019e+00–1.704e+32)	0.049886*
TRIM52	4.91E+45 (4.245e+13–5.684e+77)	0.005222**
C4orf3	2.85E+68 (2.846e+11–2.856e+125)	0.018582*
C12orf51	1.60E−39 (6.178e−73–4.167e−06)	0.022873*
C17orf65(ASB16)	4.40E−59 (3.745e−89–5.165e−29)	0.000143***
C21orf45	2.47E−91 (2.804e−181–2.182e−01)	0.048339*

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

diagnostic accuracy for differentiating metastatic and non-metastatic ccRCC cases in GSE105260 as the AUC value was >70% (**Supplementary Data Sheet 4**). By contrast, *C12orf51*, *C17orf65*, *SERPINBI*, and *TRIM52* demonstrated better diagnostic accuracy for differentiation of cases in GSE113501 (**Supplementary Data Sheet 5**). The ROC curve validated that the 12 signature genes as a whole had better diagnostic accuracy than individual signature genes in GSE105260 (**Figure 7A**). Also, the 12-gene signature could effectively diagnose the patients with metastatic ccRCC case in GSE113501 (**Figure 7B**).

The prediction powers for 1-, 3-, and 5-year overall survival were 0.87, 0.84, and 0.81, respectively, indicating that the prognostic model with the 12 signature genes had high sensitivity and specificity (**Figure 7C**). The values for predicting the 3- and 5-year survival were lower than that for 1-year, suggesting that the model was more accurate for short-term prediction than for long-term prediction. Additionally, as shown in **Figure 7D**, the 12-gene prognostic model was more powerful in the prediction than other model based on pathological stages.

DISCUSSION

ccRCC is the most common type of renal carcinoma in the world. Although the treatment strategies for ccRCC have been improved during the last decades, the management of metastatic ccRCC still remain challenging due to lack of eligible molecular targets.
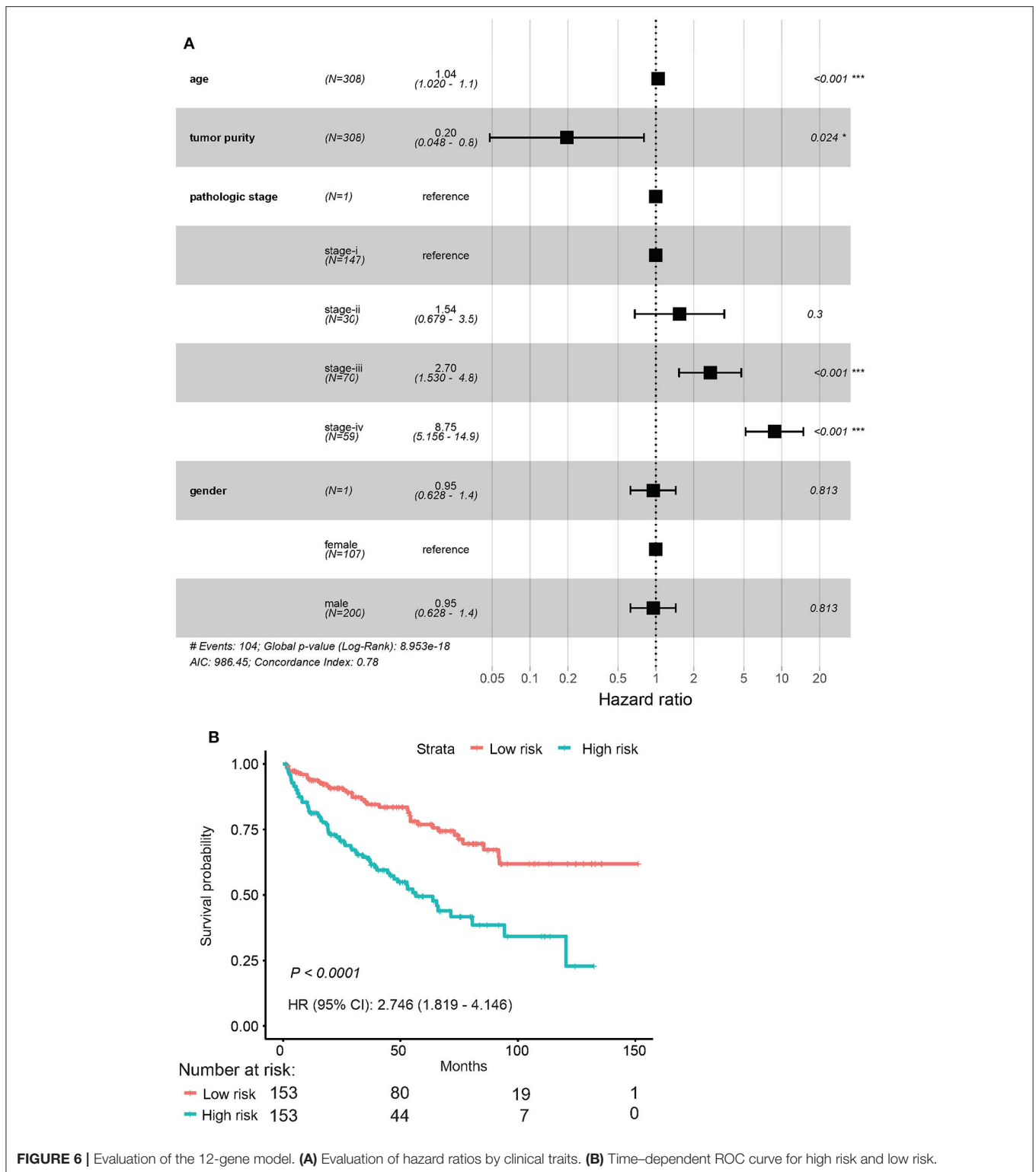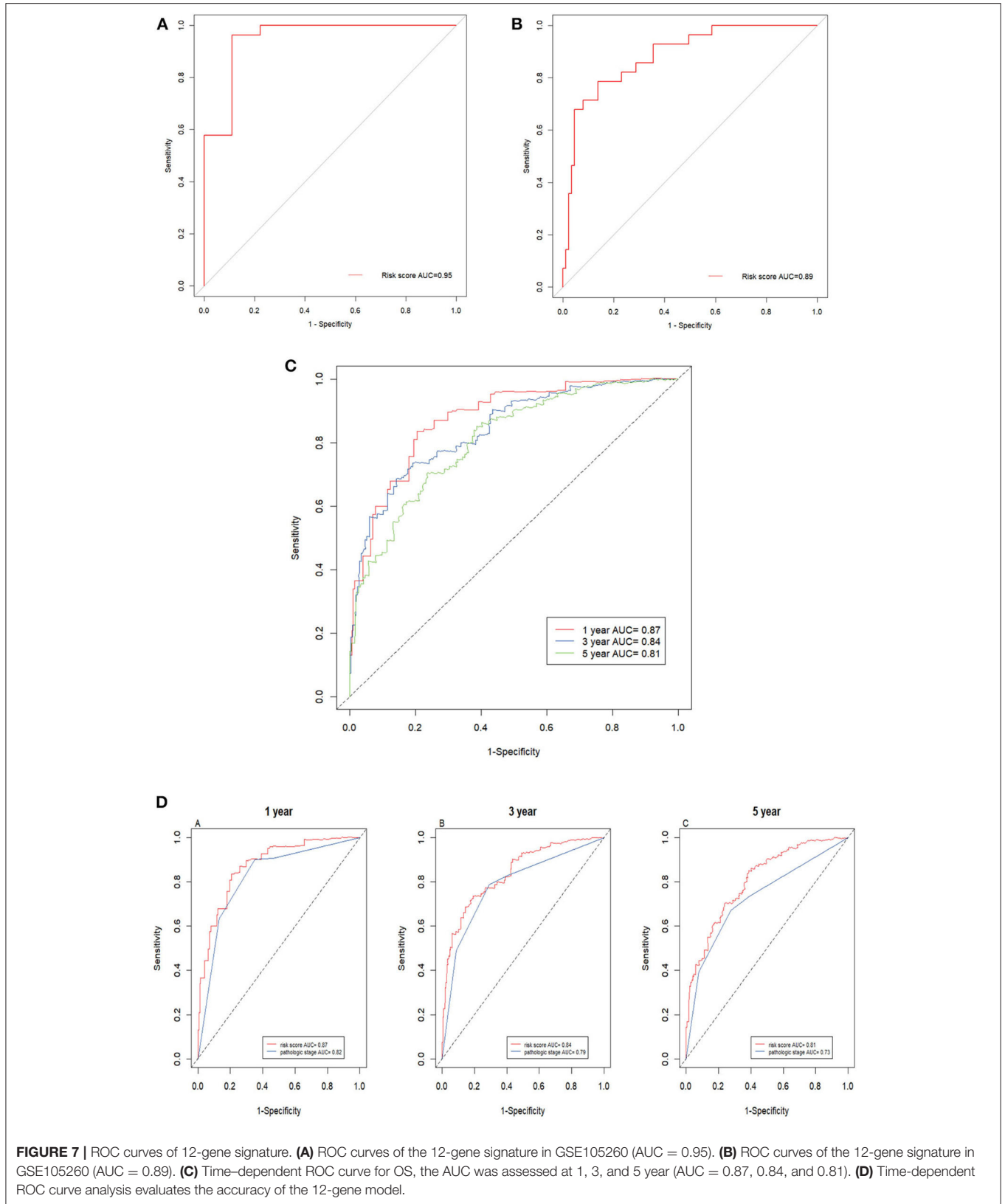
FIGURE 6 | Evaluation of the 12-gene model. **(A)** Evaluation of hazard ratios by clinical traits. **(B)** Time–dependent ROC curve for high risk and low risk.

Therefore, discovery of specific markers for early diagnosis is essential for control of ccRCC progress. DNA methylation is an important driver of many of the distinct stages of cancer, to study the molecular mechanisms underlying the pathogenesis of ccRCC may be helpful for its diagnosis and treatment. To

the best of our knowledge, there is no applicable markers for differentiating clinical stages of ccRCC. In the present study, we used the clinical data from GSE113501 and GSE105260 to profile genes associated with non-metastatic and metastatic ccRCC. Through WGCNA and ROC analysis, 12 genes were finally

FIGURE 7 | ROC curves of 12-gene signature. **(A)** ROC curves of the 12-gene signature in GSE105260 (AUC = 0.95). **(B)** ROC curves of the 12-gene signature in GSE105260 (AUC = 0.89). **(C)** Time–dependent ROC curve for OS, the AUC was assessed at 1, 3, and 5 year (AUC = 0.87, 0.84, and 0.81). **(D)** Time-dependent ROC curve analysis evaluates the accuracy of the 12-gene model.

identified in the signature module associated with the clinical traits of ccRCC.

We first screened 237 differentially methylated CpG sites from the GSE113501 dataset and 462 CpG sites from the GSE105260 dataset, which were mapped to 552 genes. We examined the methylation patterns in specific regions: TSS1500, TSS200, 5'UTR, 1stExon, body, and 3'UTR, of which TSS1500 and body region showed remarkable methylation changes. GO analysis showed these DMGs were annotated to transcription landscape. In fact, DNA methylation and transcription factors (TFs) binding represent two components of the regulatory architecture and their interplay would underlie the clinical outcome of cancer development (26–33). By the WGCNA analysis, we screened 314 intersected DMGs between metastatic and non-metastatic ccRCC, which were associated with the biological processes and functions that underlined cancer metastasis and invasion. For example, dysregulation of cell cycle was also reported to associate with cancer progress and prognosis in the number of previous studies (33–38). In view of MFs, these DMGs involved in protein binding, peptidyl-prolyl cis-trans isomerase activity, identical protein binding and poly (A) RNA binding, etc. For example, the etiology of early-onset CRC was linked to ploy (A) RNA binding endorsed by a set of hub genes (39) and the metastasis of breast cancer resulted from LSD1 demethylation (40).

From GSE105261 and GSE105260, we further screened 44 intersected genes between DEGs (in GSE105261) and DMGs (in GSE105260). We classified these intersected genes into two categories: negatively and positively correlated between their methylation and expression levels. The majority of intersected genes fell into the negative correlation category, including downregulated-hypermethylated genes and upregulated-hypomethylated genes. Among these intersected genes, the methylation changes in 10 genes were the most significantly associated to the phenotypes of ccRCC when searching against the TCGA-KIRC dataset. However, some of them have been found to associate with other cancers via diversified mechanisms, for example, nuclear LDHA promoted HPV-induced cervical cancer development using a mechanism of H3K79 hypermethylation (41–44).

Finally, univariate and multivariate Cox regression analysis identified 43 CpG sites and 12 DMGs (*CETN3*, *DCAF7*, *GPX4*, *HNRNPA0*, *NUP54*, *SERPINB1*, *STARD5*, *TRIM52*, *C4orf3*, *C12orf51*, *C17orf65*, and *C21orf45*) that were highly correlated to patients' survival. Indeed, some genes are reported to contribute to the occurrence and progress of cancers. For example, GPX4 negatively correlated with prognosis of pan-cancer patients as the low methylation level at the upstream site leads to its higher expression in cancer tissues (45). On the contrary, hypermethylation of *SERPINB1* promoter resulted in enhanced neutrophil elastase (NE) activity, which was associated with poor outcome in prostate cancer (46). Other genes like *HNRNPA0*, *DCAF7*, and *TRIM52* functioned in diverse ways including abnormal changes in alternative splicing or activation of canonical signal pathways in relation with cancer progression (47–50).

Until recently, two research groups also screened signature genes for prognosis in ccRCC patients based on TCGA database. One group constructed a prognostic model with 7 signature genes (*IFI30*, *WNT5A*, *IRF9*, *AGER*, *PLAUR*, *TEK*, and *BID*), which were all immunity-related genes (IAGs) (50). The other group built the model with 5 lncRNAs (51). While in present study, we screened the signature genes from the standpoint of methylation alteration. Our model with 12 DMGs has the AUC value of 0.81 for 5-year survival prediction compared to the IGA-based model (0.751) and the lncRNA-based model (0.91). Our model could distinguish metastatic and non-metastatic pathologies, which could accompany traditional pathological tissue typing and provide more precise diagnosis and prognosis of ccRCC patients. Meanwhile, it is necessary to validate our model with more clinical data and the mechanisms by these signature DMGs in ccRCC.

CONCLUSION

Taken together, we performed an integrative DNA methylation profiling and constructed a 12-gene model that signed in metastatic ccRCC. The model could be used for prognosis of patients with ccRCC and diagnosis of non-metastatic ccRCC and metastatic ccRCC. Aberrant methylation of the 12 genes would leave patients with ccRCC at higher risk of a poor prognosis.

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

MC and GZ designed the study and were major contributors in editing the manuscript. SQ analyzed and interpreted the data and was major contributor in writing the manuscript. SS, LZ, ST, and KX performed analysis and contributed to writing the manuscript. All authors read and approved the final manuscript.

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2020.556018/full#supplementary-material

REFERENCES

1. Baldewijns MM, van Vlodrop IJ, Schouten LJ, Soetekouw PM, de Bruïne AP, van Engeland M, et al. Genetics and epigenetics of renal cell cancer. *Biochim Biophys Acta.* (2008) 1785:133–55. doi: 10.1016/j.bbcan.2007.12.002

2. Howlader N, Noone AM, Krapcho M, Garshell J, Neyman N, Altekruse SF, et al. *SEER Cancer Statistics Review 1975–2010.* Bethesda, MD: National Cancer Institute (2013). Available online at: https://seer.cancer.gov/archive/csr/1975-2010

3. National Cancer Institute. *Cancer Stat Facts: Kidney and Renal Pelvis Cancer.* Available online at: https://seer.cancer.gov/statfacts/html/kidrp.html

4. Motzer RJ, Hutson TE, Tomczak P, Michaelson MD, Bukowski RM, Rixe O, et al. Sunitinib versus interferon alfa in metastatic renal-cell carcinoma. *N Engl J Med.* (2007) 356:115–24. doi: 10.1056/NEJMoa065044

5. Sternberg CN, Davis ID, Mardiak J, Szczylik C, Lee E, Wagstaff J, et al. Pazopanib in locally advanced or metastatic renal cell carcinoma: results of a randomized phase III trial. *J Clin Oncol.* (2010) 28:1061–8 doi: 10.1200/JCO.2009.23.9764

6. Choueiri TK, Escudier B, Powles T, Tannir NM, Mainwaring PN, Rini BI, et al. Cabozantinib versus everolimus in advanced renal cell carcinoma (METEOR): final results from a randomized, open-label, phase 3 trial. *Lancet Oncol.* (2016) 17:917–27. doi: 10.1016/S1470-2045(16)30107-3

7. Motzer RJ, Escudier B, McDermott DF, George S, Hammers HJ, Srinivas S, et al. Nivolumab versus everolimus in advanced renal-cell carcinoma. *N Engl J Med.* (2015) 373:1803–13. doi: 10.1056/NEJMoa1510665

8. Motzer RJ, Powles T, Atkins MB, Escudier B, McDermott DF, Suarez C, et al. IMmotion151: a randomized phase III study of atezolizumab plus bevacizumab vs sunitinib in untreated metastatic renal cell carcinoma. In: *Abstract (588) Presented at the Annual Genitourinary Cancers Symposium of the American Society of Clinical Oncology, February 8–10, 2018.* San Francisco, CA (2018).

9. Motzer RJ, Tannir NM, McDermott DF, Arén Frontera O, Melichar B, Choueiri TK, et al. Nivolumab plus ipilimumab versus sunitinib in advanced renal-cell carcinoma. *N Engl J Med.* (2018) 378:1277–90. doi: 10.1056/NEJMoa1712126

10. Nguyen B, Veys I, Leduc S, Bareche Y, Majjaj S, Brown DN, et al. Genomic, transcriptomic, epigenetic, and immune profiling of mucinous breast cancer. *J Natl Cancer Inst.* (2019) 111:742–6. doi: 10.1093/jnci/djz023

11. Liu T, Wu X, Chen T, Luo Z, Hu X. Downregulation of DNMT3A by miR-708-5p inhibits lung cancer stem cell-like phenotypes through repressing Wnt/β-catenin signaling. *Clin Cancer Res.* (2018) 24:1748–60 doi: 10.1158/1078-0432.CCR-17-1169

12. Kessler T, Sahm F, Sadik A, Stichel D, Hertenstein A, Reifenberger G, et al. Molecular differences in IDH wildtype glioblastoma according to MGMT promoter methylation. *Neuro Oncol.* (2018) 20:367–79. doi: 10.1093/neuonc/nox160

13. Hao X, Luo H, Krawczyk M, Wei W, Wang W, Wang J, et al. DNA methylation markers for diagnosis and prognosis of common cancers. *Proc Natl Acad Sci U S A.* (2017) 114:7414–9. doi: 10.1073/pnas.1703577114

14. Verbiest A, Couchy G, Job S, Caruana L, Lerut E, Oyen R, et al. Molecular subtypes of clear-cell renal cell carcinoma are prognostic for outcome after complete metastasectomy. *Eur Urol.* (2018) 74:474–80. doi: 10.1016/j.eururo.2018.01.042

15. Weinhold L, Wahl S, Pechlivanis S, Hoffmann P, Schmid M. A statistical model for the analysis of beta values in DNA methylation studies. *BMC Bioinform.* (2016) 17:480. doi: 10.1186/s12859-016-1347-4

16. Du P, Kibbe WA, Lin SM. Lumi: a pipeline for processing Illumina microarray. *Bioinformatics.* (2008) 24:1547–8. doi: 10.1093/bioinformatics/btn224

17. Dennis G, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, et al. DAVID: Database for annotation, visualization, and integrated discovery. *Genome Biol.* (2003) 4:R60. doi: 10.1186/gb-2003-4-9-r60

18. Li W, Huang K, Wen F, Cui G, Guo H, He Z, et al. LINC00184 silencing inhibits glycolysis and restores mitochondrial oxidative phosphorylation in esophageal cancer through demethylation of PTEN. *EBioMedicine.* (2019) 44:298–310. doi: 10.1016/j.ebiom.2019.05.055

19. Zheng S, Lin F, Zhang M, Fu J, Ge X, Mu N, et al. AK001058 promotes the proliferation and migration of colorectal cancer cells by regulating methylation of ADAMTS12. *Am J Transl Res.* (2019) 11:5869–78.

20. Liang Y, Zhang C, Dai DQ. Identification of differentially expressed genes regulated by methylation in colon cancer based on bioinformatics analysis. *World J Gastroenterol.* (2019) 25:3392–407. doi: 10.3748/wjg.v25.i26.3392

21. Zeng WJ, Yang YL, Liu ZZ, Wen ZP, Chen YH, Hu XL, et al. Integrative analysis of DNA methylation and gene expression identify a three-gene signature for predicting prognosis in lower-grade gliomas. *Cell Physiol Biochem.* (2018) 47:428–39. doi: 10.1159/000489954

22. Zhang X, Sui S, Wang L, Li H, Zhang L, Xu S, et al. Inhibition of tumor propellant glutathione peroxidase 4 induces ferroptosis in cancer cells and enhances anticancer effect of cisplatin. *J Cell Physiol.* (2020) 235:3425–37 doi: 10.1002/jcp.29232

23. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc.* (2009) 4:44–57. doi: 10.1038/nprot.2008.211

24. Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol.* (2005) 4:2194. doi: 10.2202/1544-6115.1128

25. Zhao J, Wang L, Kong D, Hu G, Wei B. Construction of novel DNA methylation-based prognostic model to predict survival in glioblastoma. *J Comput Biol.* (2019) 27:718–28. doi: 10.1089/cmb.2019.0125

26. Ung MH, Varn FS, Lou S, Cheng C. Regulators associated with clinical outcomes revealed by DNA methylation data in breast cancer. *PLoS Comput Biol.* (2015) 11:e1004269. doi: 10.1371/journal.pcbi.1004269

27. Yuan S, Yu Z, Liu Q, Zhang M, Xiang Y, Wu N, et al. GPC5, a novel epigenetically silenced tumor suppressor, inhibits tumor growth by suppressing Wnt/β-catenin signaling in lung adenocarcinoma. *Oncogene.* (2016) 35:6120–6131. doi: 10.1038/onc.2016.149

28. Han HJ, Kwon N, Choi MA, Jung KO, Piao JY, Ngo HK, et al. Peptidyl prolyl isomerase PIN1 directly binds to and stabilizes hypoxia- inducible factor-1α. *PLoS One.* (2016) 11:e0147038. doi: 10.1371/journal.pone.0147038

29. Zhang J, Zhang Y, Li Z, Wu H, Xun J, Feng H. Bioinformatics analysis of Ewing's sarcoma: seeking key candidate genes and pathways. *Oncol Lett.* (2019) 18:6008–16. doi: 10.3892/ol.2019.10936

30. He Z, Duan X, Zeng G. Identification of potential biomarkers and pivotal biological pathways for prostate cancer using bioinformatics analysis methods. *PeerJ.* (2019) 7:e7872. doi: 10.7717/peerj.7872

31. Xue J, Gao HX, Sang W, Cui WL, Liu M, Zhao Y, et al. Identification of core differentially methylated genes in glioma. *Oncol Lett.* (2019) 18:6033–45. doi: 10.3892/ol.2019.10955

32. Liu ZK, Zhang RY, Yong YL, Zhang ZY, Li C, Chen ZN, et al. Identification of crucial genes based on expression profiles of hepatocellular carcinomas by bioinformatics analysis. *PeerJ.* (2019) 7:e7436. doi: 10.7717/peerj.7436

33. Yu C, Hao X, Zhang S, Hu W, Li J, Sun J, et al. Characterization of the prognostic values of the NDRG family in gastric cancer. *Therap Adv Gastroenterol.* (2019) 12:1756284819858507. doi: 10.1177/1756284819858507

34. Meng J, Gao L, Zhang M, Gao S, Fan S, Liang C. Systematic investigation of the prognostic value of cell division cycle-associated proteins for clear cell renal cell carcinoma patients. *Biomark Med.* (2020) 14:223–38. doi: 10.2217/bmm-2019-0498

35. Zhang C, Zhang B, Meng D, Ge C. Comprehensive analysis of DNA methylation and gene expression profiles in cholangiocarcinoma. *Cancer Cell Int.* (2019) 19:352. doi: 10.1186/s12935-019-1080-y

36. Li ZX, Zheng ZQ, Wei ZH, Zhang LL, Li F, Lin L, et al. Comprehensive characterization of the alternative splicing landscape in head and neck squamous cell carcinoma reveals novel events associated with tumorigenesis and the immune microenvironment. *Theranostics.* (2019) 9:7648–65. doi: 10.7150/thno.36585

37. Jiao X, Zhang S, Jiao J, Zhang T, Qu W, Muloye GM, et al. Promoter methylation of SEPT9 as a potential biomarker for early detection of cervical cancer and its overexpression predicts radioresistance. *Clin Epigenet.* (2019) 11:120. doi: 10.1186/s13148-019-0719-9

38. Mo X, Su Z, Yang B, Zeng Z, Lei S, Qiao H. Identification of key genes involved in the development and progression of early-onset colorectal cancer by co-expression network analysis. *Oncol Lett.* (2020) 19:177–86. doi: 10.3892/ol.2019.11073

39. Liu J, Feng J, Li L, Lin L, Ji J, Lin C, et al. Arginine methylation-dependent LSD1 stability promotes invasion and metastasis of breast cancer. *EMBO Rep.* (2020) 21:e48597. doi: 10.15252/embr.201948597

40. Liu Y, Guo JZ, Liu Y, Wang K, Ding W, Wang H, et al. Nuclear lactate dehydrogenase A senses ROS to produce α-hydroxybutyrate for HPV-induced cervical tumor growth. *Nat Commun.* (2018) 9:4429. doi: 10.1038/s41467-018-06841-7

41. Zhang N, Chen S, Wu L, Wu Y, Jiang G, Shao J, et al. Identification of cancer-specific methylation of gene combination for the diagnosis of bladder cancer. *J Cancer.* (2019) 10:6761–6. doi: 10.7150/jca.28192

42. Lai AY, Fatemi M, Dhasarathy A, Malone C, Sobol SE, Geigerman C, et al. DNA methylation prevents CTCF-mediated silencing of the oncogene BCL6 in B cell lymphomas. *J Exp Med.* (2010) 207:1939–50. doi: 10.1084/jem.20100204

43. Liu B, Liu J, Liao Y, Jin C, Zhang Z, Zhao J, et al. Identification of SEC61G as a novel prognostic marker for predicting survival and response to therapies in patients with glioblastoma. *Med Sci Monit.* (2019) 25:3624–35. doi: 10.12659/MSM.916648

44. Pan H-X, Bai H-S, Guo Y, Cheng Z-Y. Bioinformatic analysis of the prognostic value of ZNF860 in recurrence-free survival and its potential regulative network in gastric cancer. *Eur Rev Med Pharmacol Sci.* (2019) 23:162–70. doi: 10.26355/eurrev_201901_16760

45. Chen T, Zheng W, Chen J, Lin S, Zou Z, Li X, et al. Systematic analysis of survival-associated alternative splicing signatures in clear cell renal cell carcinoma. *J Cell Biochem.* (2019) 121:4074–84. doi: 10.1002/jcb.29590

46. Sircoulomb F, Nicolas N, Ferrari A, Finetti P, Bekhouche I, Rousselet E, et al. ZNF703 gene amplification at 8p12 specifies luminal B breast cancer. *EMBO Mol Med.* (2011) 3:153–66. doi: 10.1002/emmm.201100121

47. Lerman I, Ma X, Seger C, Maolake A, Garcia-Hernandez ML, Rangel-Moreno J, et al. Epigenetic suppression of SERPINB1 promotes inflammation-mediated prostate cancer progression. *Mol Cancer Res.* (2019) 17:845–59. doi: 10.1158/1541-7786.MCR-18-0638

48. Mu X, Li H, Zhou L, Xu W. TRIM52 regulates the proliferation and invasiveness of lung cancer cells via the Wnt/β-catenin pathway. *Oncol Rep.* (2019) 41:3325–34. doi: 10.3892/or.2019.7110

49. Yang W, Liu L, Li C, Luo N, Chen R, Li L, et al. TRIM52 plays an oncogenic role in ovarian cancer associated with NF-kB pathway. *Cell Death Dis.* (2018) 9:908 doi: 10.1038/s41419-018-0881-6

50. Shen C, Liu J, Wang J, Zhong X, Dong D, Yang X, et al. Development and validation of a prognostic immune-associated gene signature in clear cell renal cell carcinoma. *Int Immunopharmacol.* (2020) 81:106274. doi: 10.1016/j.intimp.2020.106274

51. Zhang C, Huang D, Liu A, Xu Y, Na R, Xu D. Genome-wide screening and cohorts validation identifying novel lncRNAs as prognostic biomarkers for clear cell renal cell carcinoma. *J Cell Biochem.* (2020) 121:2559–70. doi: 10.1002/jcb.29478