Cancer Biol Med 2018. doi: 10.20892/j.issn.2095-3941.2018.0027

ORIGINAL ARTICLE

OPEN ACCESS

Five-CpG-based prognostic signature for predicting survival in hepatocellular carcinoma patients

Feng Fang*, Xiaoqing Wang*, Tianqiang Song

Department of Hepatobiliary Cancer, Tianjin Medical University Cancer Institute and Hospital, National Clinical Research Center for Cancer, Key Laboratory of Cancer Prevention and Therapy, Tianjin, Tianjin's Clinical Research Center for Cancer, Tianjin 300060, China

ABSTRACT

Objective: Hepatocellular carcinoma (HCC) is a common malignancy associated with high morbidity and mortality rates worldwide. Early diagnosis plays an important role in the improvement of HCC prognosis.

Methods: In this study, we conducted a comprehensive analysis of HCC DNA methylation and gene expression datasets in The Cancer Genome Atlas (TCGA), to identify a prognostic signature for HCC diagnosis and survival prediction. First, we identified differential methylation CpG (dmCpG) sites in HCC samples and compared them with those in adjacent normal liver tissues; this was followed by univariate analysis and Sure Independence Screening (SIS) in the training set. The robustness of the identified prognostic signature was evaluated using the testing set. To explore the biological processes involved in HCC progression, we also performed functional enrichment analysis for overlapping genes between genes containing dmCpG sites (DMGs) and differential expression genes (DEGs) in HCC patients, using data from the Database for Annotation, Visualization, and Integrated Discovery (DAVID).

Results: As a result, we identified five CpG sites that were significantly associated with HCC survival through univariate analysis and SIS. Univariate analysis of clinical characteristics identified age and risk factors (including alcohol consumption and smoking) as independent factors that indicated HCC survival. Multivariate analysis indicated that the integrated prognostic signature (weighted combination of the five CpG sites) that took age and risk factors into consideration resulted in more accurate survival prediction.

Conclusions: This study provides a novel signature for predicting HCC survival, and should be helpful for early HCC diagnosis and personalized treatment.

KEYWORDS Hepatocellular carcinoma; methylation; prognosis; prognostic signature; TCGA

Introduction

Hepatocellular carcinoma (HCC) is the third cause of death from cancer, and one of the few cancers for which upward trends are observed in both sexes worldwide[1]. Several risk factors have been identified to induce HCC, including chronic infections with hepatitis B virus (HBV) and hepatitis C virus (HCV), alcohol abuse, diabetes, and metabolic syndrome[2,3]. Clinical statistical analysis showed that early-stage HCC patients have a relatively favorable prognosis, with a 5-year survival rate of 75%[4]. However, after resection, recurrence would be observed in half of these patients, causing the 5-year survival rate to decrease to 30%. In essence, hepatocarcinogenesis is essentially a slow process, accompanied by genomic and epigenetic changes that produce cellular intermediates; this eventually evolves into hepatocellular carcinoma[5]. Recent studies have explored genomic alterations occurring during HCC using high-throughput analysis of gene microarrays, and have identified frequently mutated genes as molecular markers for tumor detection[6-8]. However, the knowledge regarding the association between the genomic phenotype and clinical outcome of HCC prediction remains extremely limited.

Genomic DNA has relatively few CpG dinucleotides (5%–10%) in which a cytosine nucleotide is followed by a guanine nucleotide in the linear sequence of bases along its 5' → 3' direction. In addition, methylating the cytosine of CpG di-nucleotides within a gene can cause heritable genomic changes without altering the DNA sequence, which is an important epigenetic pattern[9,10]. Moreover, DNA

*These authors have contributed equally to this work.

Correspondence to: Feng Fang
E-mail: xyfangfeng@hotmail.com

Received February 13, 2018; accepted July 26, 2018.
Available at www.cancerbiomed.org

Copyright © 2018 by Cancer Biology & Medicine

hypermethylation in CpG sites can lead to the repression of tumor-suppressor genes and inactivation of tumor-repair genes, resulting in a loss of tumor suppression and increased genetic damage[11,12]. Recent reports have indicated that DNA hypermethylation in promoter CpG islands was related to epidemiological characteristics, histological features, precancerous lesions, molecular characteristics, and HCC prognosis, in RASSF1A, p16, p53, DLC-1, and GSTP1[13-17]. However, further molecular investigations are still needed to obtain more information for predicting recurrence and HCC patient classification.

In the present study, DNA methylation profiles of HCC patients (n = 317) in The Cancer Genome Atlas (TCGA) database were analyzed, to determine the relationship between the aberrant methylation of cancer-specific methylation sites and clinicopathologic features. Five differential methylation CpG (dmCpG) sites that were associated with HCC prognosis were distinctly identified by a combination of univariate Cox analysis and Sure Independence Screening (SIS). Furthermore, the integration of specific clinical features with the prognostic signature (weighted combination of the five dmCpG sites) resulted in a more accurate prediction of survival. This study provides more insights about HCC survival prediction.

Material and methods

HCC dataset preprocessing

We downloaded HCC datasets from TCGA containing genome-wide scale DNA methylation and gene expression profiling data for 317 samples. Fifty patients had DNA methylation data in both HCC tissues and adjacent normal tissues. For DNA methylation datasets, we filtered out CpG sites with a detection P value > 0.05 in more than 75% samples. Samples were also excluded from analysis if more than 75% CpG sites were unreliably detected (i.e. detection P value $< 1 \times 10^{-5}$). Then, the quantile normalization method was applied for correcting background noises before conducting a comparison analysis between the rest of the normal and HCC samples. The dmCpG sites in HCC samples that were compared with those in adjacent normal tissues were identified *via* the Illumina Methylation Analyzer (IMA)[18] bioconductor package, which has been specifically designed for exploratory analysis and summarization of site- and region-level methylation changes based on the Illumina Infinium HumanMethylation450 BeadChip, using the criteria: absolute delta beta > 0.3 and adjusted P value < 0.001.

For gene expression datasets, we first conducted the quartile normalization of raw read count data, followed by logarithmic transformation, to obtain normal distribution expression values. Differential expression genes (DEGs) in HCC samples that were compared with those in adjacent normal tissues were obtained through the DESeq2[19] bioconductor package, with a thresholds of absolute log2-based fold change > 1 and adjusted P value < 0.01.

Prognostic signature construction

To construct a methylation profile-based prognostic signature for HCC survival prediction, we conducted a combination analysis of univariate and SIS. First, the samples were randomly divided into two groups, i.e. training set (158 samples) and testing set (159 samples); their clinicopathologic characteristics have been shown in **Table 1**. Second, univariate survival analysis was conducted for dmCpG sites in the training set to identify CpG sites that were significantly associated with HCC survival, which have been abbreviated as SurvCG hereafter. Thirdly, we performed SIS, which is a variable selection technique for model selection and estimation in high-dimensional statistical models, for SurvCG identification using LASSO regression analysis *via* the SIS R package, to identify a reliable CpG combination for HCC survival prediction. The prognostic signature is a weighted combination of SIS identified CpG sites.

Multivariate survival analysis

Clinicopathologic features might also prove to be important indicators for HCC survival. We have studied four features, i.e. age, sex, risk factors (including alcohol consumption, smoking, and HBV infection), and stage, to analyze their associations with HCC survival in the training and testing sets. Prognostic score, which was based on the prognostic signature for every sample, was calculated through the distinct values for every CpG site included in the prognostic signature, and its association with HCC survival was evaluated through univariate survival analysis in the training and testing sets. Besides, we also conducted multivariate survival analysis for clinicopathologic features and prognostic scores to obtain the most robust combination of features for HCC survival prediction.

Functional enrichment analysis

We intersected DEGs in HCC patients with genes containing dmCpG sites (DMGs), to obtain genes whose dysregulation might be affected by aberrant DNA methylation. The

Table 1 Clinical characteristics of samples used as training and testing sets.

Characteristics	Training set	Testing set	*P*
Censor rate	63.9%	70.1%	NA
Age, median years (range)	65.4 (17.9–86.0)	59.5 (16.1–84.8)	NA
Gender, *n* (%)			0.329
Male	96 (60.8)	106 (73.0)	
Female	62 (39.2)	53 (27.0)	
Risk factor, *n* (%)			0.290
Yes	30 (19.0)	39 (24.5)	
No	128 (81.0)	120 (75.5)	
Stage, *n* (%)			0.480
I, II	120 (75.9)	127 (79.9)	
III, IV	38 (24.1)	32 (20.1)	

NA represents not available, *P* value was obtained through Chi-square test.

functions of the involved overlapping genes should represent biological processes involved in HCC progression. Hence, we conducted a functional enrichment analysis for overlapping genes using the Database for Annotation, Visualization and Integrated Discovery (DAVID)[20]. Gene Ontology (GO) terms and KEGG pathways with P values < 0.05 were identified, and biological process (BP) terms were further clustered through the Enrichment Map Plugin of Cytoscape[21,22].

Results

Differential methylation analysis

Preprocessing the DNA methylation dataset retained 479,036 of 485,577 CpG sites and samples for the following analysis. **Figure 1A** illustrated the average beta value for every CpG site in adjacent normal tissues (X-axis) and HCC samples (Y-axis). Differential methylation analysis identified 10,803 dmCpG sites in HCC samples, as compared with those in adjacent normal tissues; of which 9,373 and 1,430 sites were hypomethylation and hypermethylation sites, respectively. **Figure 1B** showed the heatmap of beta values of dmCpG sites (row) in HCC and adjacent normal samples (column), in which green and red colors represent low and high methylation values, respectively. We then explored the location-wise distribution of hypermethylation and hypomethylation CpG sites relative to the CpG island and gene. As a result, hypermethylation CpG sites tend to be located in a CpG island (80.13%) and regulatory elements, such as TSS200 (200bp upstream/downstream of transcription start site) and TSS1500 (**Figure 1C**). Hypomethylated CpG sites tend to be located in the OpenSea (70.15%, i.e. area far away from CpG island, and usually 4,000 bp or more) and gene body area (62.09%) (**Figure 1D**). This was consistent with the fact that the cancer methylome is characterized by hypermethylation in CpG islands of the promoter region and hypomethylation in diffuse CpG sites.

Prognostic signature

Univariate analysis showed that 100 CpG sites were significantly associated with HCC survival. SIS identified 5 CpG sites, whose combination could robustly predict HCC survival. **Figure 2A** illustrated the beta value in adjacent normal tissues and HCC samples and regression coefficients of the 5 CpG sites. Four of the five CpG sites were hypermethylated in HCC samples compared to those in adjacent normal tissues. Prognostic score = $1.10 \times \text{cg05971966} - 0.61 \times \text{cg08833577} + 1.23 \times \text{cg14826425} + 1.92 \times \text{cg20980783} + 1.17 \times \text{cg24085930}$. Kaplan-Meier plots were prepared and compared using the log-rank test. Hence, we found that a higher prognostic score is significantly associated with poor HCC survival in the training ($P = 3.60 \times 10^{-4}$) and testing sets ($P = 7.62 \times 10^{-3}$) (**Figure 2B**). To evaluate the robustness of our prognostic signature further, we assigned the samples in the testing set into four groups with the same sample size in an increasing order of prognostic score. It was found that HCC survival became poorer with an increase in the prognostic score (**Figure 2C**, $P = 0.017$).

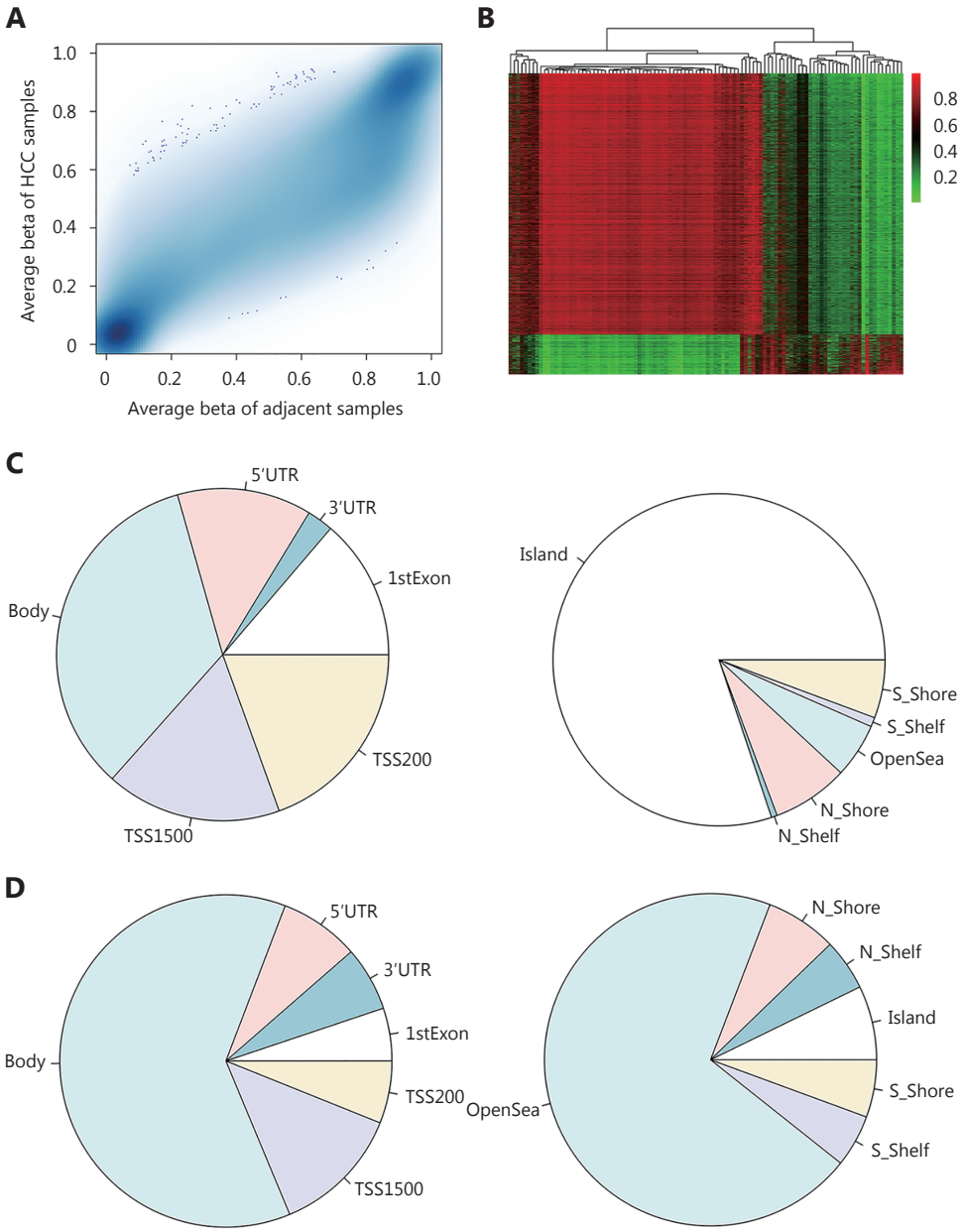

Figure 1 Differential methylation analysis. (A) Average beta value of all CpG sites in HCC (Y-axis) and adjacent normal samples (X-axis). (B) Heatmap of beta values of dmCpG sites in HCC patients and their corresponding adjacent normal samples. Rows represent dmCpG sites and columns represent samples; green and red colors represent the low and high beta values, respectively. (C) Distribution of hypermethylation CpG site locations relative to the gene (left panel) and CpG island (right panel). (D) Distribution of hypomethylation CpG site locations relative to the gene (left panel) and CpG island (right panel).

Multivariate survival analysis

As shown in **Table 2**, age (> 60 and < 60) and risk factors (with and without) were identified as independent survival factors in training and testing sets. Combining the prognostic score and all the four clinicopathologic features could further separate samples with better survival from those with poorer survival in training (P = 0.023) and testing sets (P = 0.042) (**Figure 3**). Multivariate survival analysis indicates that a combination of prognostic scores and ages could provide the most robust prediction method for HCC survival.

Supplementary Figure S1 showed the prediction

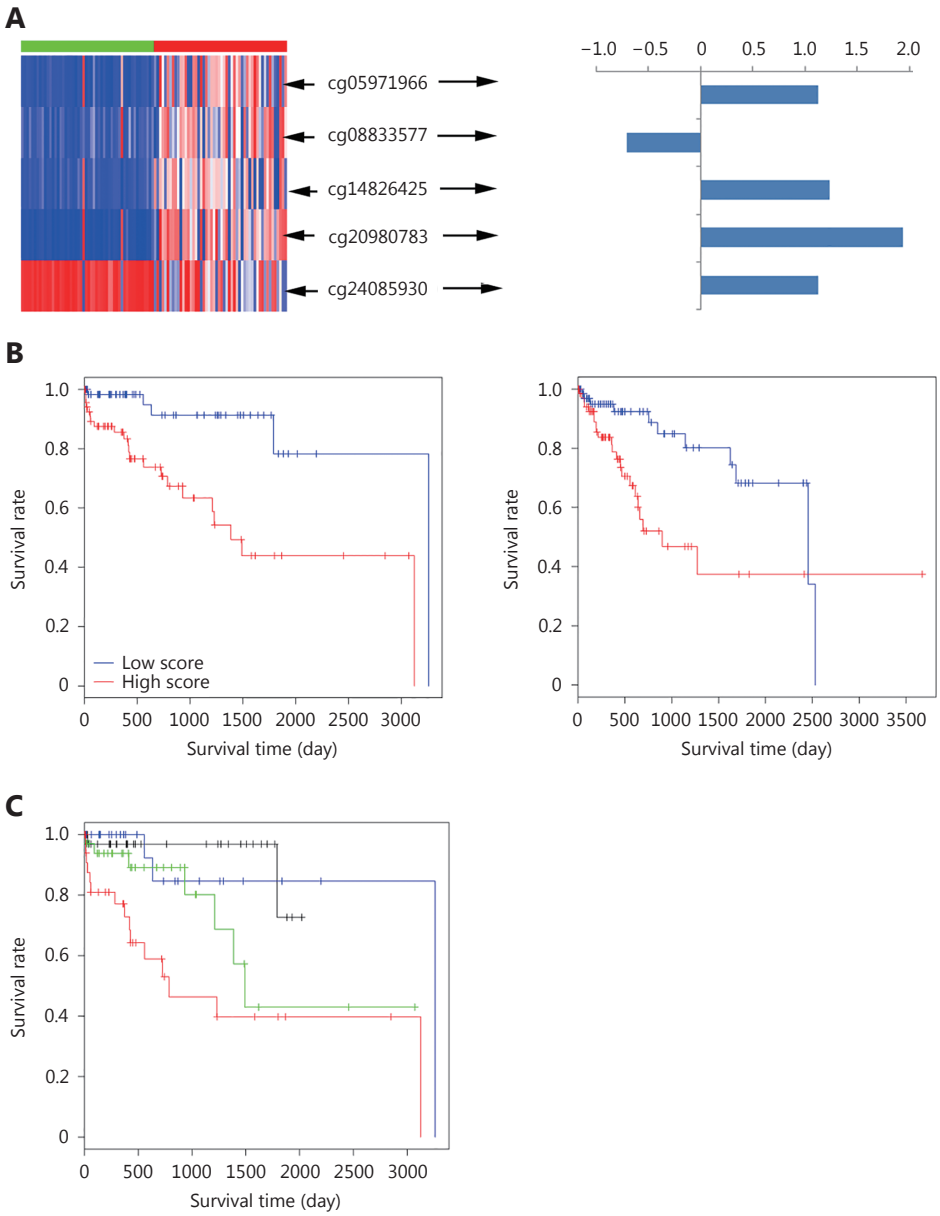

Figure 2 Prognostic signature and Kaplan-Meier survival analysis. (A) Beta value (left panel) and the regression coefficients (right panel) of the five CpG sites, identified through a combination of univariate survival analysis and SIS. Blue and red bars at the top of the heatmap (left panel) represent adjacent normal and HCC samples. (B) Kaplan-Meier plot of samples with lower (green line) and higher (red line) prognostic scores (divided by the median prognostic score) in the training set (left panel) and testing set (right panel). Plus signs are censored values. (C) Kaplan-Meier plot of samples with different prognostic scores in the testing set. Prognostic scores were sorted in the decreasing order. The red line represents the first quarter samples; blue line represents the second quarter samples; green line represents the third quarter samples and black line represents the last quarter samples. Plus signs are censored values.

performance of prognostic score in the testing set after adjustment for age (A), stage (B), gender (C), and risk factors (D), respectively, from which we conclude that the prognosis score could effectively predict HCC prognosis independent of the main clinicopathologic features. Additionally, receiver operating characteristic curve (ROC) analysis was conducted to compare differences in predicting HCC prognosis using a combination of the prognostic score and different clinicopathologic features. As shown in **Supplementary Figure S2**, a combination of prognostic score and age

Table 2 Cox regression analysis of clinical characteristics and risk scores.

Characteristics	Training set		Testing set	
	Univariable	Univariable	Univariable	Univariable
	P (HR, 95% CI)	*P* (HR, 95% CI)	*P* (HR, 95% CI)	*P* (HR, 95% CI)
Risk score	0.00036 (0.74, 0.56–0.89)	0.000135 (0.51, 0.46–0.76)	0.00762 (0.81, 0.68–0.87)	0.00504 (0.79, 0.69–0.85)
Age	0.00192 (0.86, 0.75–0.89)	3.94e–05 (0.45, 0.41–0.69)	0.00783 (0.82, 0.71–0.90)	0.0015 (0.75, 0.62–0.79)
Gender	0.0955 (0.91, 0.85–0.96)	0.000653 (0.42, 0.32–0.50)	0.000202 (0.40, 0.29–0.50)	0.00645 (0.81, 0.75–0.86)
Risk factor	0.0268 (0.82, 0.75–0.90)	6.79e–05 (0.51, 0.43–0.60)	0.0723 (0.89, 0.75–0.92)	0.00467 (0.73, 0.65–0.78)
Stage	0.0777 (0.93, 0.82–0.96)	0.000622 (0.53, 0.42–0.63)	0.0724 (0.88, 0.78–0.92)	0.0107 (0.74, 0.68–0.79)

CI represents confidence interval.

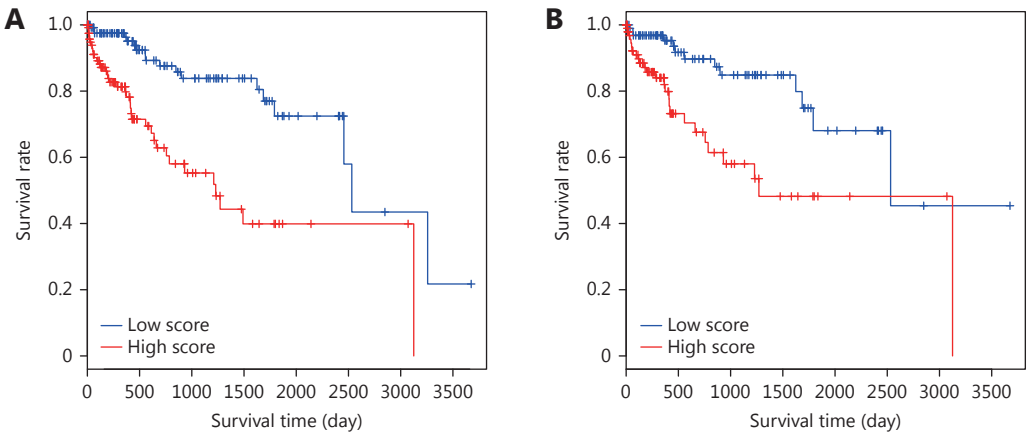

Figure 3 Kaplan-Meier plot of multivariate survival analysis of prognostic score and clinicopathologic features. (A) Kaplan-Meier plot of samples with lower (green line) and higher (red line) combined scores (divided by the median combined score) in the training set. Plus signs are censored values. (B) Kaplan-Meier plot of samples with lower (green line) and higher (red line) combined score (divided by the median combined score) in the testing set. Plus signs are censored values.

distinctly outperformed other combinations, as well as the prognostic score alone.

Functional enrichment analysis

A total of 2,538 DEGs were identified in HCC samples compared with adjacent normal tissues. The dmCpG sites were found to be located in 3,618 genes, i.e. DMGs. We identified 580 overlaps between DEGs and DMGs, and those overlapping genes were significantly associated with biological processes and pathways of cancer progression and neurodegenerative diseases. **Figure 4** top-right panel shows the cluster analysis of significantly enriched BP terms. A main cluster that was associated with extracellular matrix and vasculature development was obtained. **Figure 4** bottom-right panel shows the significantly enriched pathways with gene number shown. Strikingly, we found a significantly enrichment of nervous system disease-related pathways, which indicated the potential associations between cancer and nervous system diseases. Consistent with this result, many studies have previously proven associations between cancer and nervous system disorders[23,24].

Discussion

The value of clinical characteristics and initial performance status as prognostic factors for HCC survival has long been recognized[25,26]. However, due to the great individual differences and complicated influential factors, the traditional analytic strategies are often unable to predict the prognosis of HCC patients. DNA methylation is a type of covalent chemical modification and a stable (replication-coupled) epigenetic marker. It can be detected in biological fluids and fresh-frozen and paraffin-embedded tissue samples, by methylome profiling in the clinical setting[27]. Thus, the high-throughput detection of genetic alterations

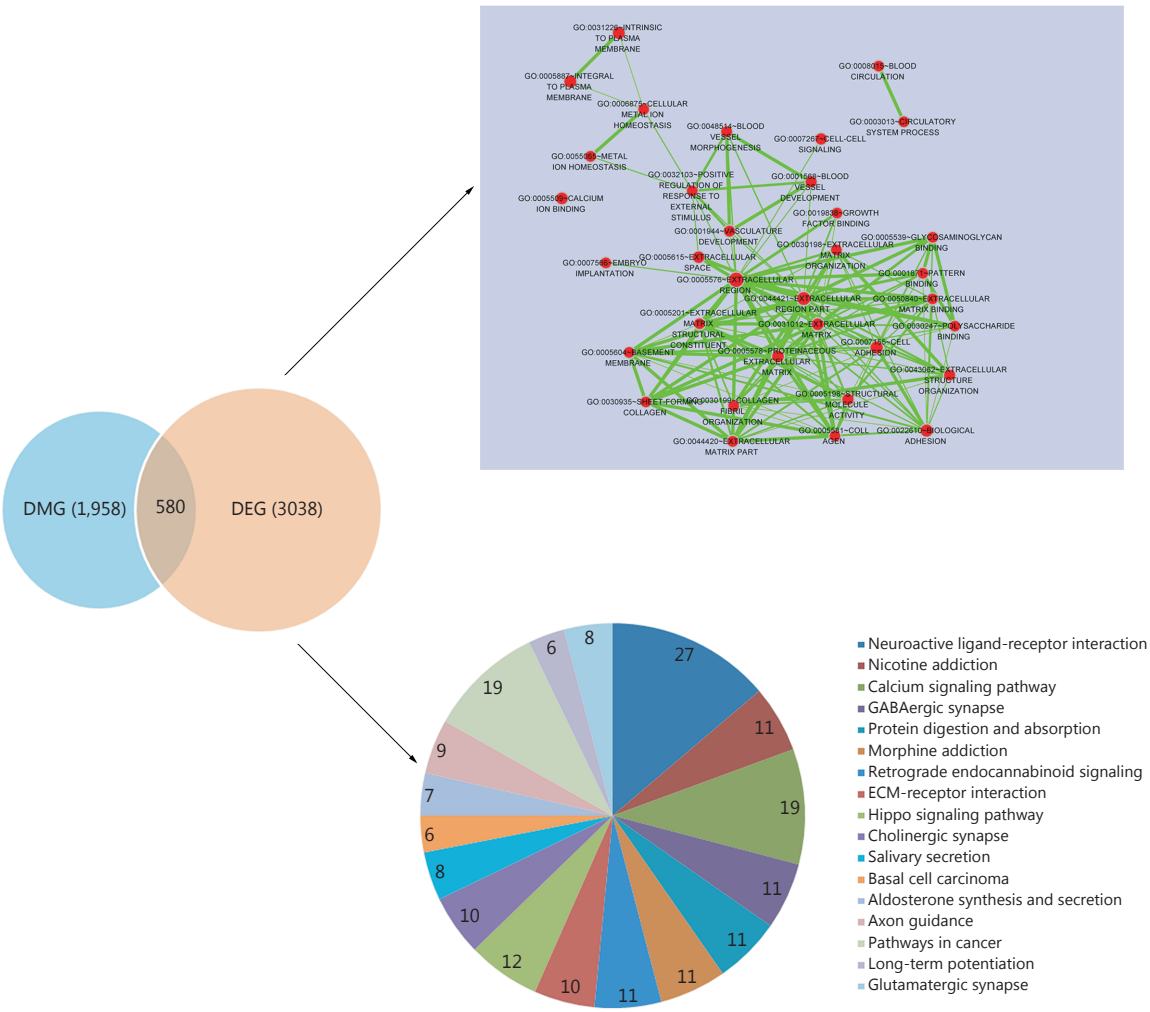

Figure 4 Cross-analysis of DEGs and DMGs. A total of 580 overlaps were identified between DEGs and DMGs (left panel). The top-right panel shows the cluster analysis of significantly enriched BP terms of overlapping genes. Nodes are BP terms and lines are interactions among them. The larger node represents more genes contained in the terms and darker color i.e. smaller *P* value is more significant. Line thickness represents the number of overlapping genes between two terms. The bottom-right panel represents the significantly enriched KEGG pathways of overlapping genes. The gene number involved in each pathway is also shown in the pie chart.

has been widely used in the early diagnosis, individual treatment, and prognosis prediction of various cancers[28]. CpG island methylation, a common molecular tumor marker, has already been confirmed as a prognostic indicator in lung, prostate, and esophageal cancers, and acute leukemia[12].

Aberrant DNA methylation changes in CpG islands as well as hypermethylation and diffuse genomic hypomethylation are common phenomena in multiple human cancers, and there is a good chance that these two changes are mechanically linked[29,30]. In this study, differential methylation CpG sites (dmCpG) in HCC samples were compared to those in adjacent normal tissues and were screened using DNA methylation microarrays in TCGA samples. The probes have been annotated based on their relationship with the nearest gene and the probes might belong to any of the following genomic elements: TSS1500, TSS200, 5'UTR, 1st exon, gene body, 3'UTR, or intergenic regions. Genomic regions close to the transcription start site showed a relatively high level of methylation (TSS1500, TSS200, 5'UTR and 1stexon), and 80% of the hypermethylation sites were located in the CpG island, which indicated that CpG islands in promoters tend to be hypermethylated in HCC patients. Most of the hypomethylation dmCpG sites were located in regions further away from the transcription start site (gene body and

3'UTR) and CpG islands (OpenSea). This is consistent with the reported associations between DNA methylation and cancer development.

Promoter CpG island hypermethylation has been reported to occur in chronic liver diseases[31,32], whereas genome-wide hypomethylation takes place at the HCC stage[33]. As shown in the previous study, p16 hypermethylation in HCC was found to be more frequently caused by cirrhotic than non-cirrhotic inducements, suggesting that the epigenetic modulations of HCC may be influenced by the disease state of the background liver[34]. However, the underlying mechanism between CpG island hypermethylation and the background liver condition remains unclear. The GO term and KEGG analysis results of overlapping genes between DMG and DEG in HCC samples were compared to those in para-carcinoma tissues, and they indicated a significant enrichment of molecule activity, extracellular matrix formation, cell-matrix and cell-cell interactions, cell secretion, cell adhesion, and vessel morphogenesis, suggesting the main signal transduction pathway regulated by DNA methylation during the development of HCC. Tumor cells need to grow in the appropriate target organ microenvironment, and cancer progression depends on the interplay between transformed cells and their microenvironment, particularly the surrounding extracellular matrix (ECM)[35]. In human livers, fibro genes underlie the development of HCC in at least 90% of cases[36], and the overexpression of matrix components and MMP2 activity were strikingly associated in HCCs[37]. Microenvironment of the background liver is undoubtedly a decisive factor that influences tumor recurrence and metastasis[38].

Considering all these factors, we identified five dmCpG sites that exhibited a significant association between their aberrant methylation and patient prognosis using a combination of univariate analysis and sure independence screening. A prognostic signature of the combination of the five dmCpG sites was obtained and the prognostic score was negatively correlated with HCC prognosis. Furthermore, multivariate analysis revealed that the prognostic score showed significant relevance together with clinical characteristics such as age and risk factors, which affected the overall survival of HCC patients. It has been proven that CpG island hypermethylation in HCC tumors was closely associated with the condition of the background liver and sex[39]. In addition, our study indicates that the combination of age, risk factors, and the prognostic score provides a more robust prediction for HCC prognosis than any one of these alone.

Conclusions

In this study, we performed a comprehensive analysis of DNA methylation and gene expression datasets of HCC data obtained from TCGA and identified an important prognostic signature. This should be helpful for HCC survival prediction and personalized treatment.

Acknowledgements

This work was supported by the National Nature Science Foundation of China (Grant No. 81201644 and 81572858) and National Key Clinical Specialist Construction Programs of China (Grant No. 2013-544).

Conflict of interest statement

No potential conflicts of interest are disclosed.

References

1. Bertuccio P, Turati F, Carioli G, Rodriguez T, La Vecchia C, Malvezzi M, et al. Global trends and predictions in hepatocellular carcinoma mortality. J Hepatol. 2017; 67: 302-9.
2. Bosetti C, Turati F, La Vecchia C. Hepatocellular carcinoma epidemiology. Best Pract Res Clin Gastroenterol. 2014; 28: 753-70.
3. Farazi PA, DePinho RA. Hepatocellular carcinoma pathogenesis: from genes to environment. Nat Rev Cancer. 2006; 6: 674-87.
4. Johnson P, Berhane S, Kagebayashi C, Satomura S, Teng M, Fox R, et al. Impact of disease stage and aetiology on survival in hepatocellular carcinoma: implications for surveillance. Br J Cancer. 2017; 116: 441-7.
5. Thorgeirsson SS, Lee JS, Grisham JW. Molecular prognostication of liver cancer: end of the beginning. J Hepatol. 2006; 44: 798-805.
6. Zhang YJ, Wu HC, Shen J, Ahsan H, Tsai WY, Yang HI, et al. Predicting hepatocellular carcinoma by detection of aberrant promoter methylation in serum DNA. Clin Cancer Res. 2007; 13: 2378-84.
7. Hill VK, Ricketts C, Bieche I, Vacher S, Gentle D, Lewis C, et al. Genome-wide DNA methylation profiling of CpG islands in breast cancer identifies novel genes associated with tumorigenicity. Cancer Res. 2011; 71: 2988-99.
8. The Cancer Genome Atlas Research Network. Comprehensive and integrative genomic characterization of hepatocellular carcinoma. Cell. 2017; 169(7): 1327-1341.e23.
9. Razin A, Riggs AD. DNA methylation and gene function. Science. 1980; 210: 604-10.
10. Jabbari K, Bernardi G. Cytosine methylation and CpG, TpG (CpA) and TpA frequencies. Gene. 2004; 333: 143-9.

11. Das PM, Singal R. DNA methylation and cancer. J Clin Oncol. 2004; 22: 4632-42.
12. Issa JP. CpG island methylator phenotype in cancer. Nat Rev Cancer. 2004; 4: 988-93.
13. Wong IHN, Lo YMD, Zhang J, Liew CT, Ng MHL, Wong N, et al. Detection of aberrant *p16* methylation in the plasma and serum of liver cancer patients. Cancer Res. 1999; 59: 71-3.
14. Shen LL, Ahuja N, Shen Y, Habib NA, Toyota M, Rashid A, et al. DNA methylation and environmental exposures in human hepatocellular carcinoma. J Natl Cancer Inst. 2002; 94: 755-61.
15. Zhang YJ, Ahsan H, Chen Y, Lunn RM, Wang LY, Chen SY, et al. High frequency of promoter hypermethylation of *RASSF1A* and *p16* and its relationship to aflatoxin B_1-DNA adduct levels in human hepatocellular carcinoma. Mol Carcinog. 2002; 35: 85-92.
16. Wong CM, Lee JM, Ching YP, Jin DY, Ng IO. Genetic and epigenetic alterations of *DLC-1* gene in hepatocellular carcinoma. Cancer Res. 2003; 63: 7646-51.
17. Hussain SP, Schwank J, Staib F, Wang XW, Harris CC. TP53 mutations and hepatocellular carcinoma: insights into the etiology and pathogenesis of liver cancer. Oncogene. 2007; 26: 2166-76.
18. Wang D, Yan L, Hu Q, Sucheston LE, Higgins MJ, Ambrosone CB, et al. IMA: an R package for high-throughput analysis of Illumina's 450K Infinium methylation data. Bioinformatics. 2012; 28: 729-30.
19. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014; 15: 550
20. Huang DW, Sherman BT, Tan QN, Kir J, Liu D, Bryant D, et al. DAVID Bioinformatics resources: expanded annotation database and novel algorithms to better extract biology from large gene lists. Nucleic Acids Res. 2007; 35: W169-75.
21. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003; 13: 2498-504.
22. Merico D, Isserlin R, Stueker O, Emili A, Bader GD. Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. PLoS One. 2010; 5: e13984
23. Realmuto S, Cinturino A, Arnao V, Mazzola MA, Cupidi C, Aridon P, et al. Tumor diagnosis preceding Alzheimer's disease onset: is there a link between cancer and Alzheimer's disease? J Alzheimers Dis. 2012; 31: 177-82.
24. Ibáñez K, Boullosa C, Tabarés-Seisdedos R, Baudot A, Valencia A. Molecular evidence for the inverse comorbidity between central nervous system disorders and cancers detected by transcriptomic meta-analyses. PLoS Genet. 2014; 10: e1004173
25. Al-Sarraf M, Go TS, Kithier K, Vaitkevicius VK. Primary liver cancer: a review of the clinical features, blood groups, serum enzymes, therapy, and survival of 65 cases. Cancer. 1974; 33: 574-82.
26. Okuda K, Ohtsuki T, Obata H, Tomimatsu M, Okazaki N, Hasegawa H, et al. Natural history of hepatocellular carcinoma and prognosis in relation to treatment study of 850 patients. Cancer. 1985; 56: 918-28.
27. How Kit A, Nielsen HM, Tost J. DNA methylation based biomarkers: practical considerations and applications. Biochimie. 2012; 94: 2314-37.
28. Soto J, Rodriguez-Antolin C, Vallespin E, de Castro Carpeno J, Ibanez de Caceres I. The impact of next-generation sequencing on the DNA methylation-based translational cancer research. Transl Res. 2016; 169: 1-18.e1.
29. Feinberg AP, Ohlsson R, Henikoff S. The epigenetic progenitor origin of human cancer. Nat Rev Genet. 2006; 7: 21-33.
30. Cho NY, Kim JH, Moon KC, Kang GH. Genomic hypomethylation and CpG island hypermethylation in prostatic intraepithelial neoplasm. Virchows Arch. 2009; 454: 17-23.
31. Lee S, Lee HJ, Kim JH, Lee HS, Jang JJ, Kang GH. Aberrant CpG island hypermethylation along multistep hepatocarcinogenesis. Am J Pathol. 2003; 163: 1371-8.
32. Um TH, Kim H, Oh BK, Kim MS, Kim KS, Jung G, et al. Aberrant CpG island hypermethylation in dysplastic nodules and early HCC of hepatitis B virus-related human multistep hepatocarcinogenesis. J Hepatol. 2011; 54: 939-47.
33. Lin CH, Hsieh SY, Sheen IS, Lee WC, Chen TC, Shyu WC, et al. Genome-wide hypomethylation in hepatocellular carcinogenesis. Cancer Res. 2001; 61: 4238-43.
34. Zang JJ, Xie F, Xu JF, Qin YY, Shen RX, Yang JM, et al. *P16* gene hypermethylation and hepatocellular carcinoma: a systematic review and meta-analysis. World J Gastroenterol. 2011; 17: 3043-8.
35. Gupta GP, Massagué J. Cancer metastasis: building a framework. Cell. 2006; 127: 679-95.
36. Affo S, Yu LX, Schwabe RF. The role of cancer-associated fibroblasts and fibrosis in liver cancer. Annu Rev Pathol. 2017; 12: 153-86.
37. Théret N, Musso O, Turlin B, Lotrian D, Bioulac-Sage P, Campion JP, et al. Increased extracellular matrix remodeling is associated with tumor progression in human hepatocellular carcinomas. Hepatology. 2001; 34: 82-8.
38. Nguyen DX, Bos PD, Massagué J. Metastasis: from dissemination to organ-specific colonization. Nat Rev Cancer. 2009; 9: 274-84.
39. Lee HS, Kim BH, Cho NY, Yoo EJ, Choi M, Shin SH, et al. Prognostic implications of and relationship between CpG island hypermethylation and repetitive DNA hypomethylation in hepatocellular carcinoma. Clin Cancer Res. 2009; 15: 812-20.

Cite this article as: Fang F, Wang X, Song T. Five-CpG-based prognostic signature for predicting survival in hepatocellular carcinoma patients. Cancer Biol Med. 2018; 15: 425-33. doi: 10.20892/j.issn.2095-3941.2018.0027

Supplementary materials

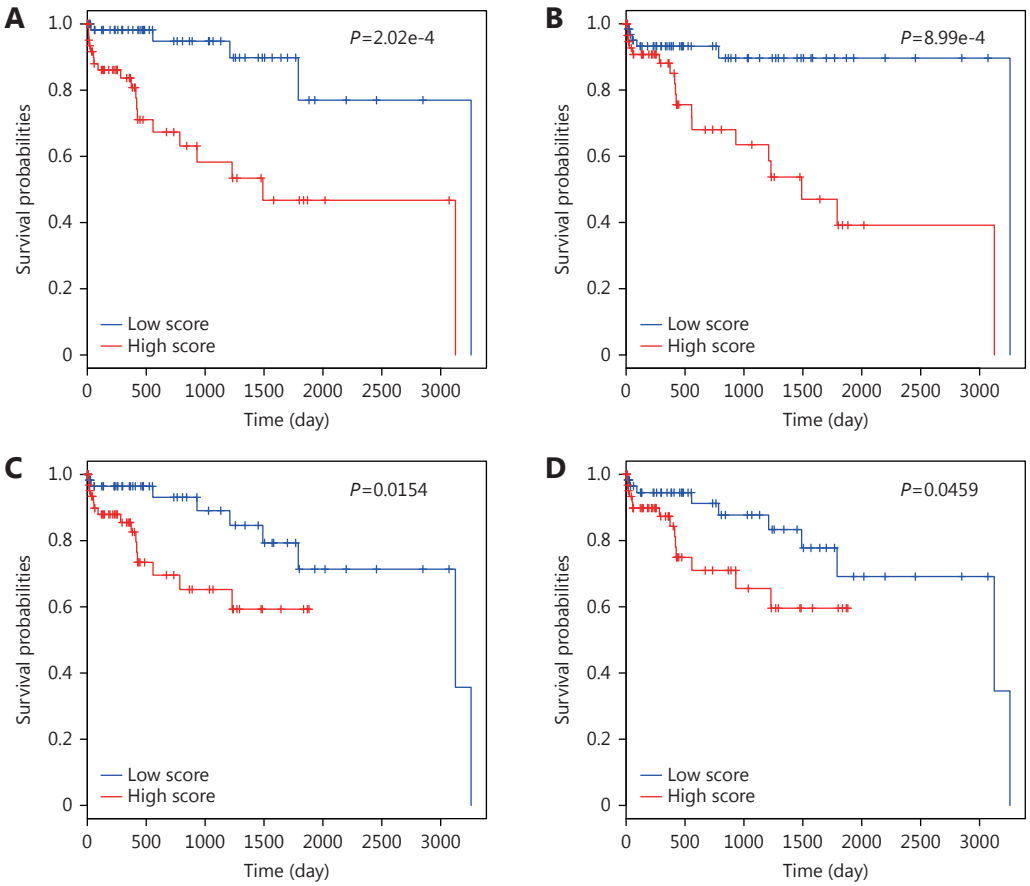

Figure S1 The prediction performance of prognostic score in the testing set after adjustment for age (A), stage (B), gender (C), and risk factors (D), respectively.

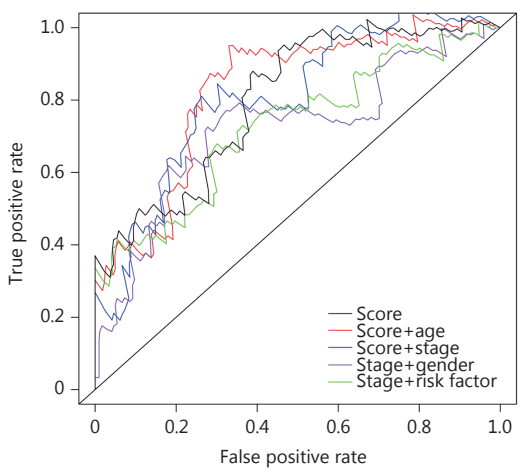

Figure S2 Receiver operating characteristic curve (ROC) analysis was conducted to compare differences inpredicting HCC prognosis using a combination of the prognostic score and different clinicopathologic features.