

Identifying Methylation-Driven Immune-related Molecular Subtypes Predicts the Prognosis of Hepatocellular Carcinoma

Qian Yan

Guangzhou University of Chinese Medicine

Baoqian Ye

Guangzhou University of Chinese Medicine

Boqing Wang

Guangzhou University of Chinese Medicine

Wenjiang Zheng

Guangzhou University of Chinese Medicine

Xiongwen Wang (✉ awen681029@163.com)

Guangzhou University of Chinese Medicine

Research Article

Keywords: hepatocellular carcinoma, DNA methylation, molecular subtype, immune-related genes, prognosis model

Posted Date: July 28th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-737824/v1>

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

[Read Full License](#)

Identifying Methylation-Driven Immune-related Molecular subtypes Predicts the Prognosis of Hepatocellular Carcinoma

Qian Yan^{1a}, Baoqian Ye^{1a}, Boqing Wang^{1a}, Wenjiang Zheng^{1*}, Xiongwen Wang^{*2}

¹The First Clinical Medical College, Guangzhou University of Chinese Medicine, Guangzhou, China

²Department of Oncology, The First Affiliated Hospital, Guangzhou University of Chinese Medicine, Guangzhou, China

^aThese authors have the same contribution and should be considered as co-first authors

*Correspondence:

Xiongwen Wang, Tel: 13711448399, Email: awen681029@163.com

Wenjiang Zheng, Tel: 16760205187, Email: 20202120031@stu.gzucm.edu.cn

Abstract

The purpose of this study is to analyze the DNA methylation and gene expression profiles of immune-related CpG sites to identify the molecular subtypes and CpG sites related to the prognosis of HCC. In this study, the DNA methylation and gene expression datasets were downloaded from The Cancer Genome Atlas database, together with immune-related genes downloaded from the immunology database and analysis portal database to explore the prognostic molecular subtypes of HCC. By performing consistent clustering analysis on 830 immune-related CpG sites, we identified seven subgroups with significant differences in overall survival. Finally, 16 classifiers of immune-related CpG sites were constructed and used in the testing set to verify the prognosis of DNA methylation subgroups, and the results were consistent with the training set. Using the TIMER database, we analyzed 16 immune-related CpG sites expression with the abundance of six types of immune infiltrating cells and found that most are positively correlated with the level of infiltration of multiple immune cells in HCC. This study screened potential immune-related prognostic methylation sites and established a new prognosis model of HCC based on DNA methylation molecular subtype, which may help in the early

diagnosis of HCC and developing more effective personalized treatments.

Keywords: hepatocellular carcinoma, DNA methylation, molecular subtype, immune-related genes; prognosis model.

Introduction

Cancer is a global health problem. The incidence of primary liver cancer (PLC) ranks sixth among all cancer types, and PLC is the fourth most common cause of cancer-related deaths [1]. Hepatocellular carcinoma (HCC), the most important pathological subtype of PLC, accounts for about 90% of diagnosed cases [2]. According to the results of the latest statistics from the International Agency for Research on Cancer of the World Health Organization, the incidence of PLC will increase from 841,080 in 2018 to 1,361,836 in 2040, while the mortality number will increase from 78,136 to 1,284,252 from 2018 to 2040 [3]. Due to the significant heterogeneity of HCC, fewer than 30% of patients meet the indications for radical treatment [4-6]. Over the past 10 years, the level of clinical management of HCC has improved significantly, especially in advanced patients [2, 4, 5]. Although systemic treatments show a certain effect, especially in molecular targeted drugs and immune checkpoint blockers, the overall prognosis of patients with HCC remains unsatisfactory. Thus, we still need to continue to explore on the basis of previous studies and seek better treatment options.

Because of the particularity of its anatomical structure and immunological state, the liver is a common target for primary and secondary tumors. PLC is often secondary to the development of various chronic inflammations, such as nonalcoholic steatohepatitis, alcoholic liver disease, hepatitis B virus (HBV), and hepatitis C virus infection, which can promote its development [2]. Moreover, the liver is a crucial immune organ that contains a large number of immune cells, including natural killer cells, natural killer T cells, and CD56⁺, CD4⁺, and CD8⁺ T cells and macrophages, thus maintaining and regulating the homeostasis between the liver and body [7]. On the one hand, the immune system can destroy the viruses that promote the growth of cancer cells, disrupt tumor cells as well as inhibit the proliferation of cancer cells. On the other hand, it can inhibit the formation and development of an inflammatory microenvironment, which could ultimately promote cancer growth. In this case, we

can see that the immune system plays a dual role in the process of tumor development [8]. A growing body of studies indicated that the immune system can be considered a significant factor influencing tumorigenesis and development of HCC [9-11]. In recent years, immunotherapies (such as anti-PD-1, anti-PD-L1 and anti-CTLA-4 antibodies) have been widely used in advanced HCC and have shown potential therapeutic effects [12]. Furthermore, there is a high degree of heterogeneity among different tumor types and patients, even among different patients with the same tumor. Hence, the identification of immune-related biomarkers helps to stratify patients who benefit from long-term treatment with immune checkpoints.

With the continuous advancement of high-throughput sequencing technology, multi-omics research including genomics, transcriptomics, and epigenomics has gained critical insights into the molecular biology of disease [6]. From the perspective of cancer genomics, HCC is derived from the accumulation of somatic genome and epigenome changes in primitive tissues over time. Unfortunately, at least for now, most of the driver genes detected in HCC are not clinically feasible[13]. Nevertheless, attempts are being made at investigating more new treatment approaches.

In the global epigenomic changes, DNA methylation is one of the important epigenetic regulators, which can adjust the expression of related genes and the phenotype of cells [14] and is closely related to the inactivation of the X chromosome [15], tissue differentiation [16], cellular development [17], and genetic imprinting [18]. Unlike genetic mutations, changes in DNA methylation can be reversed and occur early in the body. Abnormal changes can be detected through body fluids and have the potential for early diagnosis and prognosis of cancer [14]. To date, there is already conclusive evidence that epigenetic markers can be used for prognosis and predictive biomarkers in oncology, which will lead to breakthroughs in the prevention and treatment of human cancer and will expand into other diseases in the future [19]. A comprehensive analysis of DNA methylation and gene expression also contributes to identify epigenetic drivers in cancer [20]. Xu [21] constructed and validated a diagnostic prediction model for HCC based on ten specific methylation markers. Zheng et al. [22] screened 222 epigenetic driving genes in HCC patients, whose expression was strongly negatively regulated by promoter methylation, and the

methylation of SFN, SPP1, and TK1 was significantly correlated with overall survival (OS). Additionally, Cheng et al. [23] comprehensively analyzed the data of DNA methylation and gene expression profiles and identified six specific abnormal methylation sites, providing promising biomarkers for the diagnosis of HCC. Shen et al. [24] demonstrated the importance of abnormal DNA methylation in HCC tumorigenesis through genome-wide methylation analysis.

Future studies may require combination regimens including both immunotherapies and molecularly matched targeted treatments [13, 25, 26]. Genomic studies on the immune mechanism of HCC can already predict molecular biomarkers related to the prognosis of the disease. In addition, studies have shown that epigenetic therapy has the effect of regulating immunity, and the two have the potential to be combined[27]. Immune checkpoint inhibitors have now become one of the conventional therapies for HCC, and the combination of epigenetic therapy and immunotherapy is becoming a tool to enhance the immune response[28, 29]. However, more research is needed to determine the best combination of epigenetic therapy and immunotherapy. To gain insight into the prognostic value of immune-related genes (IRGs) and DNA methylation in HCC, we established an HCC classification based on integrated gene expression data and DNA methylation sites related to IRGs. This will allow the further development comprehensive DNA methylation biomarker prognosis prediction model to improve clinical prognosis and personalized treatment.

Materials and Methods

Data acquisition and pre-processing

On March 6, 2020, the RNA-sequencing data of HCC were downloaded from The Cancer Genome Atlas (TCGA, <https://portal.gdc.cancer.gov/>) database, including 374 cases of HCC tissues and 50 cases of normal tissue, together with clinical follow-up data. The data of the Illumina Infinium HumanMethylation450 Bead-Chip array in TCGA database were retrieved from the UCSC Cancer Genomics Browser (<https://genome-cancer.ucsc.edu/>), which is an application that gathers various -omic data with clinical and phenotype data to study the correlation between them. The data downloaded from this website contained 380 tumor tissues and 50 adjacent normal

tissues [30]. In addition, a total of 1811 IRGs were derived from the immunology database and analysis portal (ImmPort, <https://www.immport.org/>) [31].

Preliminary screening of DNA methylation loci in hepatocellular carcinoma

The data of TCGA-LIHC DNA methylation array contain 485,578 methylation sites, and the methylation level of each probe is represented by the β value, which ranges from 0 to 1, with 0 indicating unmethylation and 1 indicating complete methylation. On the one hand, we deleted samples with more than 70% missing methylation sites or containing “NA” sites and removed unstable genomic sites, such as methylation sites on sex chromosomes. Furthermore, the regulation of gene silencing/expression by abnormal DNA methylation in the promoter region, which is located 2,500 bp upstream to 500 bp downstream of the gene, is one of the key features of cancer [32]. Hence, we selected the methylation sites in the promoter region. Most importantly, we combined the annotation files with the IRGs obtained in the ImmPort database and further screened the methylation sites for subsequent analysis. On the other hand, we discarded the samples whose OS was less than 30 days or when there were vacancy values in the OS and survival state.

Univariate and multivariate survival analysis of methylation sites in the training set

According to the methylation level of each CpG methylation site, we used the survival package and coxph function in the R software to construct a univariate Cox proportional hazards regression model for each methylation site in the training set, hence obtaining methylation loci related to prognosis for further multivariate analysis [33]. OS was our endpoint, and $p < 0.05$ was our screening standard. Afterward, we performed a multivariate analysis of these prognostic methylation sites and clinical data, such as age, gender, stage, and grade. The significant methylation sites independent from these clinical shapes and prognostic methylation sites were used for further molecular subtype screening.

Screening of molecular subtypes and specific methylation sites in hepatocellular carcinoma

The above-identified methylation sites associated with IRGs are related to the prognosis of HCC, so the hub-methylation sites were used for the next analysis. Using the ConsensusClusterPlus package in the R software, we performed a consistent clustering analysis of the above-mentioned methylation sites [34]. The software package uses the consensus clustering algorithm to classify the methylation sites in the sample; that is, the repeated subsampling and custom clustering methods provided by the algorithm can provide stable evidence. At the same time, the results were visualized using a consensus matrix, cumulative distribution function curve plots, and the delta area curve of consensus clustering to find the best molecular subtype groupings with strong correlations within the group and weak correlations between groups [34].

Survival and clinical characteristics analyses

To investigate the differences in survival among various molecular subtypes, we applied survival package in the R software to illustrate the survival rate among HCC subgroups defined by DNA methylation profiles. The log-rank test was used to evaluate the significance of differences among the clusters. Associations between both clinical and biological characteristics and DNA methylation clustering were analyzed using the chi-squared test. All tests were two-sided; $p < 0.05$ was considered statistically significant for all tests.

Functional enrichment analysis and genome annotation

By further annotating the above-mentioned methylated sites with significant differences, a series of prognostic-associated IRGs can be obtained. To explore the potential functions and the enriched terms of these IRGs, we analyzed the possible mechanisms and functions of IRGs via gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis, enabled by the clusterProfiler package in R software [35], with a false discovery rate value < 0.05 . GO enrichment

analyses were mainly composed of the biological process (BP), cellular component (CC), and molecular function (MF); the KEGG analysis was an annotation of the pathway that the protein may participate in.

Construction and verification of the prognosis model of hepatocellular carcinoma

The least absolute shrinkage and selection operator (LASSO) algorithm is a widely accepted method of shrinkage estimation, which establishes a penalty function to obtain a relatively complete model [36]. We deleted genes with high correlation to prevent overfitting of the model and performed cross-validation after 1,000 simulations. Therefore, we obtained the correlation coefficient of each methylation site in the model. Hence, the above-mentioned prognosis-related methylation sites in each subtype were incorporated to construct a prognostic prediction model based on the LASSO algorithm using the survival and glmnet package in R software, and 16 CpG sites were contained in the model. The risk score formula is as follows:

$$\begin{aligned} \text{Risk score} = & -0.96 * \text{cg00536939} \pm 1.27 * \text{cg00630958} + 5.48 * \text{cg01914037} \pm 9.72 * \text{cg09959112} + \\ & -0.86 * \text{cg10395772} \pm 0.51 * \text{cg11839863} \pm 1.39 * \text{cg13615963} \pm 3.12 * \text{cg14046477} + \\ & -0.14 * \text{cg14076258} \pm 0.56 * \text{cg15140465} \pm 0.58 * \text{cg15929078} \pm 0.31 * \text{cg19476647} + \\ & -0.05 * \text{cg21282997} + -1.69 * \text{cg23165623} + -0.70 * \text{cg24065044} + -1.91 * \text{cg26822501}. \end{aligned}$$

The expression profile data of specific CpG methylation sites were subsequently extracted from both the training and testing group, followed by substitution into the model for calculation. According to the above-mentioned formula, we can get the risk score of each methylation site. The training set is divided into a high-risk group and a low-risk group based on the median risk score. We used the Kaplan-Meier survival curve to show the difference in survival rate between high- and low-risk groups, and also showed the heatmap of the methylation level of the CpG sites, survival status, and risk value between the two groups. The ROC curve was used to analyze and validate the predictive accuracy of the prognostic classification model and identify the stability of the methylation characteristics.

Verification of immune-related signatures

To further verify the application value of methylation sites in the model, we annotated the 16 methylation sites and obtained the hub genes driven by each methylation site. The cBio Cancer Genomics Portal (cBioPortal, <http://www.cbioportal.org/>) database was used for somatic mutation analysis of these prognosis IRGs in HCC, including amplification, deep deletion, and mRNA high, and so on. In this database, we also performed the Pearson correlation analysis between the DNA methylation status and mRNA expression levels of IRGs [37].

Correlations between 16 immune-related signatures expression and immune cells in TIMER

The relationship between 16 IRGs expression and immune infiltration was determined using the TIMER (<http://cistrome.org/TIMER/>) database[38]. TIMER is an ideal resource for the systematic analysis of immune infiltration across diverse cancer types. TIMER applies a previously published statistical deconvolution method to infer the abundance of tumor-infiltrating immune cells from gene expression profiles [38]. The TIMER database contains 371 samples across HCC from TCGA to allow the evaluation of the abundance of immune infiltration. We analyzed 16 IRGs expression with the abundance of all six types of immune infiltrating cells, including B cells, CD4⁺ T cells, CD8⁺ T cells, neutrophils, macrophages, and dendritic cells. The relationship between the expression level of 16 IRGs and tumor purity was also determined.

Results

Identification of potential prognostic methylation sites associated with overall survival in training dataset patients

First, we preprocessed the DNA methylation and clinical data as detailed in the *Materials and Methods* section. After removing the normal samples, we divided the samples into two groups, one is the training group and the other is validation set. There were 14,192 methylation sites related to IRGs identified in both groups. Finally, the training group contained 231 samples, and the validation group included 118

samples. The corresponding clinical information for the two groups is shown in **Table 1**, respectively. Then, we performed univariate and multivariate cox regression analysis on the methylation sites related to OS in the training set, and clinical factors were included in the analysis as well. In univariate analysis, we obtained 1,494 methylation sites associated with prognosis. In addition, clinical factors such as stage ($p = 4.228 \times 10^{-6}$), T staging ($p = 8.314 \times 10^{-6}$), and M staging ($p = 0.008$) were also significant. Yet, when we combined all clinical parameters with critical methylation sites into a multivariate analysis, only grade ($p = 0.0195$) and 830 CpG sites were independent prognostic factors for OS, which further indicated the superiority of our model. Ultimately, the information of 830 methylation sites was used for further analysis. The data of the top 20 CpG sites are shown in **Table 2**.

Using characteristic DNA methylation sites to achieve consistent clustering of hepatocellular carcinoma molecular subgroups

To discover the molecular subgroups related to the prognosis of HCC, we performed a consensus cluster analysis on the methylation levels of the 830 methylation sites obtained in the univariate and multivariate analysis mentioned above. Taking advantage of the ConsensusClusterPlus package in R software, we re-sampled the qualified data 1,000 times and judged the optimal molecular subtype according to the following criteria: strong correlation within each subgroup, weak correlation between subgroups, and stable area under the consensus cumulative distribution function (CDF) curve. Based on the quasi-flat growth rate of the CDF value, we chose $k = 7$ group as the final tumor classification (**Figure 1**). To judge the correlation between groups or within groups, a consensus matrix was used to display a heatmap of the consensus for $k = 7$, further clarifying the classification structure of $k = 7$ (**Figure 2**).

Cluster analysis and clinical characteristics analysis of methylation expression profile based on hepatocellular carcinoma molecular subgroup

Based on the above-mentioned consistent clustering analysis results, we chose a stable clustering result of $k = 7$. As shown in **Figure 3**, 231 samples from the training

set were allocated to seven subgroups. The overall methylation level of each subgroup was distinct. At the same time, we included all clinical factors in our study to display the distribution of different clinical factors in different subgroups in the form of a heatmap. Kaplan–Meier survival analysis showed that the prognosis differed significantly among the seven subgroups ($p = 7.53 \times 10^{-5}$) (**Figure 4A**). The prognosis of clusters 3 and 5 was the best, and that of cluster 6 was the worst. In addition, we further analyzed the distribution of age, gender, grade, stage, T staging, N staging, and M staging in each molecular subgroup (**Figure 4B–H**). The results suggested that cluster 1 was associated with the size, location, degree of invasion, and distant metastasis of the primary tumor; cluster 2 was associated with lymph node metastasis and distant metastasis; cluster 3, which was related to the earlier grade and stage, included most of the males; cluster 4 was associated with later grade and lymph node metastasis; clusters 5 and 6 had earlier grade and stage, there was no lymph node and distant metastasis, and the patients in cluster 6 were all over 65 years old; cluster 7 had the late grade stage, accompanied by lymph node and distant metastasis, further suggesting the differences among various molecular subtypes.

Screening and functional enrichment analysis of specific methylation sites based on molecular subtypes of hepatocellular carcinoma

First, we annotated the 830 specific methylation sites and obtained 513 genes located in the promoter region. Combined with the data of the transcriptome, we extracted the expression of these genes in each sample and plotted a heatmap of gene expression among subgroups (**Figure 5A**). The expression levels of these genes in different subgroups were significantly different, indicating that there is a certain regulatory relationship between the methylation modification in the promoter region of these genes and their expression.

We further analyzed the differences in 830 methylation sites in HCC subtypes, obtaining 238 subtype-specific CpG sites and 189 genes in the promoter region. The heatmap in **Figure 5B** shows that cluster 4 has the most differential methylation sites, and all of them are hypomethylation sites. Clusters 5 and 3, with a good prognosis, were enriched to 34 and 48 differential methylation sites, most of which were

hypermethylation sites. Compared with other clusters, cluster 5 has the highest methylation level (**Figure 6**).

To study the biological mechanism of these specific methylation sites in the pathogenesis and progression of HCC, we used the clusterProfiler R package to perform GO and KEGG enrichment analysis on the 513 genes, with a threshold of $p < 0.05$. **Figure 7** and **Supplementary Tables 1-4** summarize the top 30 items enriched by GO and the top 20 items enriched by pathway enrichment. Among them, the BP mainly included regulation of leukocyte activation, positive regulation of cytokine production, and T cell activation (**Figure 7A**); CC included the receptor complex, side of the membrane, and plasma membrane protein complex (**Figure 7B**); molecular function (MF) was mainly enriched in receptor-ligand activity, receptor regulator activity, and growth factor activity (**Figure 7C**). KEGG enrichment analysis results (**Figure 7D**) showed that these genes are mainly involved in the immune process of cells and the regulation of classical tumor signals, such as cytokine-cytokine receptor interaction, Th17 cell differentiation, the Mitogen-activated protein kinase (MAPK) signaling pathway, and the Janus kinase/signal transducers and activators of transcription (JAK-STAT) signaling pathway. The enrichment analysis of differential methylation sites showed that these genes mainly play a role in immune-related pathways and classical tumor pathways, which is basically consistent with the above-mentioned conclusions.

Establishment and test evaluation of the prognostic model for patients with hepatocellular carcinoma

Cluster 5 had 34 specific methylation sites, most of which were hypermethylation sites. It also contained a large number of samples and was related to a good prognosis. In this case, we chose cluster 5 as the seed cluster to evaluate the prognostic ability of specific methylation sites.

First, we analyzed the screening process of methylation sites in this model. The abscissa in **Figure 8** is the penalty coefficient (lambda [λ] value), which controls the complexity of the model. The larger the value of λ , the greater the penalty. In **Figure 8A**, the ordinate is the correlation coefficient. As the value of λ increases, the

correlation coefficient gradually becomes 0. The ordinate in **Figure 8B** is the error of cross-validation. When the cross-validation error is the smallest, we can see that the optimal coefficient to be included in the model is 16.

Subsequently, we used the formula in the Materials and Methods section to calculate the risk score of each sample, ranked the samples according to the risk score, and selected the median value of the risk score to divide the training set into two groups of high and low risk. As shown in **Figure 9**, the high-risk group has a worse prognosis (**Figure 9A**), and the risk score is between -11 and -1 (**Figure 9B**). As the risk score increases, the patient's OS rate decreases (**Figure 9C**). As shown in **Figure 9D**, patients in the high-risk group have a worse prognosis, which is associated with lower methylation levels. At the same time, the model was verified by the receiver operating characteristic (ROC) curve and the area under the curve (AUC). When the value of AUC is greater than 0.7, the model has a good prognostic performance. As shown in **Figure 9E**, the AUC of the training set is 0.825, indicating that the model has a good prognostic performance.

Finally, we verified the model in the testing set, used the methylation levels of these 16 CpG sites for the test dataset samples, and applied the prognostic model to calculate the risk score. The samples were classified according to the risk score, and the median risk score in the training set was used to divide the test dataset samples into a high-risk group and a low-risk group. The prognosis of the two groups was again significantly different (**Figure 10A**, $p = 2.075 \times 10^{-2}$). These results are consistent with the results obtained from the training set, indicating that the model has good accuracy and stability (**Figure 10B-D**).

Verification of immune-related genes in model

Given the crucial potential clinical meaning, we annotated the 16 CpG sites in the model and obtained the following 16 IRGs: *NR1H3*, *LEPR*, *PGRMC2*, *CIITA*, *FGF14*, *FGF1*, *CCR6*, *NR6A1*, *PAK1*, *TAPBP*, *RGL2*, *LTBR*, *IL18RAP*, *ESR1*, *BDNF*, and *LTBP1*. We analyzed the frequency of gene mutations in 349 patients in TCGA database, and **Figure 11A** shows the overall mutation landscape of a series of samples in HCC. Among them, mRNA high (32.66%) and multiple alternations (15.76%)

accounted for the largest proportion, and deep deletion (0.86%) was the rarest. The mutation rate of 14 genes out of 16 genes could reach more than 5%, of which the mutation rate of *TAPBP* and *RGL2* >10% (**Figure 11B**). Additionally, we performed a Pearson correlation analysis on the DNA methylation level and mRNA expression of these 16 IRGs. The analysis results showed that, except for the *ESR1* gene, the mRNA expression of the other genes was negatively correlated with the DNA methylation level (**Figure 12**), which implied that methylation in the promoter region might have a regulatory effect on gene expression.

The levels of 16 immune-related genes expression correlate with the infiltration levels of immune cells in hepatocellular carcinoma

Several studies suggest that the survival times of patients in HCC is determined by the quantity and activity status of tumor-infiltrating lymphocytes [39, 40]. Therefore, it makes sense for us to explore the association between immune infiltration and 16 IRGs expression. We explored the relationship between 16 IRGs expression and the infiltrating immune cells in HCC using the TIMER database. The results show that *BDNF*, *CCR6*, *CIITA*, *ESR1*, *FGF1*, *FGF14*, *IL18RAP*, *LTBP1*, *NR1H3*, *NR6A1* and *TAPBP* expression significantly correlate with tumor purity in HCC. However, *LEPR*, *LTBR*, *PKA1*, *PGRMC2*, *RGL2* have the opposite results (**Figure 13**). The *BDNF*, *CCR6*, *CIITA*, *FGF14*, *IL18RAP*, *LTBP1*, *NR6A1*, *PKA1*, *RGL2* and *TAPBP* expression level had significant positive correlations with the infiltration levels of B cells, CD8⁺ T cells, CD4⁺ T cells, macrophages, neutrophils, and dendritic cells (**Figure 13A, B, C, F, G, I, L, M, O, P**).

The other 6 IRGs had some different outcomes. *ESR1* expression has significant negative correlation with infiltrating levels of B cell and macrophage but no relation with infiltrating levels of CD8⁺ T cell, CD4⁺ T cell, neutrophil and dendritic cell (**Figure 13D**). *FGF1* expression has significant positive correlation with infiltrating levels of CD8⁺ T cell, CD4⁺ T cell, macrophage, neutrophil and dendritic cell but no relation with infiltrating levels of B cell (**Figure 13E**). *LEPR* expression has no relation with infiltrating levels of neutrophil and significant negative correlation with infiltrating levels of B cell, CD8⁺ T cell, CD4⁺ T cell, macrophage and dendritic cell

(**Figure 13H**). LTBR expression has no relation with infiltrating levels of CD8⁺ T cell and significant positive correlation with infiltrating levels of B cell, CD4⁺ T cell, macrophage, neutrophil and dendritic cell (**Figure 13J**). NR1H3 expression has significant positive correlation with infiltrating levels of B cell, CD8⁺ T cell, neutrophil and no relation with infiltrating levels of CD4⁺ T cell, macrophage and dendritic cell (**Figure 13K**). PGRMC2 expression has no relation with infiltrating levels of B cell, CD4⁺ T cell, macrophage and significant positive correlation with infiltrating levels of CD8⁺ T cell, neutrophil and dendritic cell (**Figure 13N**).

Although the results are slightly different, it strongly suggests that most IRGs are positively correlated with the level of infiltration of multiple immune cells in HCC.

Discussion

HCC is a highly heterogeneous disease with different clinical, pathological, and molecular subtypes, which may lead to different sensitivities to therapies and result in distinct prognostic results [41, 42]. Immune cells, one of the main CCs of the tumor microenvironment and one of the main causes of tumor heterogeneity, play a significant role in the formation and development of cancer [43, 44]. To date, immunotherapeutic approaches for HCC have shown encouraging results. Programmed cell death protein-1 inhibitors, including nivolumab and pembrolizumab, have received accelerated approval for the treatment of advanced HCC based on the promising outcomes of phase I/II clinical trials [45, 46]. However, recent phase III studies have failed to demonstrate statistically significant improvement compared to other treatment options [47, 48]. Although systemic therapy shows a certain effect, the improvement of the patient's prognosis is not significant. More recently, epigenetic mechanisms, including DNA methylation and histone modification, have received much attention in cancer [49, 50]. Dawson MA et al. demonstrated that epigenetic alterations can regulate the gene activity and influence the cellular phenotype, which also plays an important role in tumorigenesis [51]. In addition, the regulation of gene expression by abnormal methylation of DNA in tumor cells and lymphocytes may affect the signaling and expression of important proteins in innate and acquired immune system functions [52]. Several studies have reported that DNA methylation

can regulate the expression of programmed cell death protein-1 and programmed cell death protein - Ligand 1 [53-55]. Yonghong Zhang et al. [56] tested and verified that the immune system has a distinct DNA methylation signature in HCC, and as cancer progresses, the signature intensifies. Therefore, the identification of these immune-related abnormal methylation sites is crucial for further research on the clinical potential of combining epigenetic drugs with immunotherapy strategies. The establishment of this model can not only stratify the prognosis of different patients according to the level of DNA methylation but also provide a reference for the early diagnosis and new treatment of HCC.

DNA methylation is a covalent chemical modification that can be inherited but is more reversible than gene mutation, which is thought to be an effective predictive tool for tumor patients [14]. Genome-wide hypomethylation and the hypermethylation of tumor suppressor genes (TSGs) are general features of tumors. These methylation changes can be used as potential molecular markers for tumorigenesis and development [57-59]. CpG islands have always been the focus of DNA methylation research [60], covering approximately 60%–70% of the promoter regions of human protein-coding genes, usually with low methylation levels, and the hypermethylation of CpG island in the promoter region is associated with transcriptional silencing [61]. Studies have indicated that abnormal DNA methylation of TSGs, such as cyclin-dependent kinase 4 [62], Ras association domain family 1 isoform A [63], and Runt-related transcription factor 3 [64], is considered to be an early indication of cancer and maybe a potential biomarker of HCC [65]. Wang et al. conducted a systematic review to find that CpG island methylator phenotype was significantly associated with the prognosis of HCC [66]; Cheng et al. identified six HCC-specific hypermethylation sites as potential diagnostic biomarkers, and this combination had a 92% sensitivity in predicting HCC [23]; Neumann et al. determined three TSGs, including period homolog 3, insulin-like growth-factor-binding protein, acid-labile subunit, and protein Z, indicating that promoter hypermethylation is indeed the cause of gene silencing [67]. In this study, we selected methylation sites located in the promoter region to construct molecular subtypes related to the prognosis of HCC and

tried to integrate multiple DNA methylation sites related to immunity to stratify the HCC patients, with a view to providing a promising target for clinical practice.

Hence, we integrated the gene expression and methylation data associated with IRGs to explore the aberrant DNA loci that can drive and regulate the IRGs. Subsequently, a consensus clustering analysis was performed on key methylation sites to discover molecular subgroups related to the prognosis of HCC. We found that the survival difference among each subtype is statistically significant (Figure 4A), and every subtype has different clinical characteristics. In addition, we annotated the 830 specific methylation sites and obtained 513 genes located in the promoter region. Combining this with transcriptome data, we extracted the expression of these genes in each sample. As is well known, the level of gene expression has a profound influence on the regulation of cell's vital activities. Our results (Figure 5A) indicated that the methylation modification in the promoter region of these genes and their expression have a certain regulatory relationship. To obtain a deeper understanding of the function of these genes, we performed GO and KEGG enrichment analysis on these 513 genes. GO enrichment analysis provides a simple annotation of gene products from BP, CC and MF, and the results of KEGG can help us decipher gene pathways. GO enrichment results showed that these genes are mainly involved in the regulation of leukocyte activation, positive regulation of cytokine production, and T cell activation (Figure 7A), which are closely related to immunity and inflammation in the BP. Studies have shown that cytokines are the main cause of HBV infection, and the HBV-infected liver can form a complex immune response network under the interaction of various cytokines and immune cells, thereby promoting the occurrence and development of hepatitis B [68]. In MF, these genes primarily regulate receptor ligand activity, receptor modulating activity, and growth factor activity (Figure 7C). In CC, they are mainly involved in the composition of various complexes. Meanwhile, KEGG enrichment analysis results (Figure 7D) indicated that these genes are mainly involved in cellular immune processes and the regulation of classical tumor signals, such as cytokine–cytokine receptor interaction, Th17 cell differentiation, the MAPK signaling pathway, and the JAK-STAT signaling pathway (Figure 7D).

Because HCC is a tumor with high heterogeneity at the molecular and histological levels, the molecular subtypes identified by this model may help to gradually move the HCC molecular subtypes from the purely academic context to the clinical setting. At present, the molecular subtypes of breast cancer [69, 70], ovarian cancer [71], cervical cancer [72], and some other common tumors are used to guide the selection of clinical treatment, while many other tumors are also gradually included in clinical studies. Molecular classification can reflect different biological backgrounds and has potential significance in the early diagnosis and treatment of patients. Based on transcriptome data, two major molecular subtypes of PLC have been proposed: proliferative and non-proliferative [73, 74]. Because HCC is a typical inflammatory disease, tumor microenvironment is getting more and more attention in HCC. Hence, a classification method of HCC based on immune status has gradually been developed. Shimada et al. analyzed the gene expression and mutation information in postoperative tissues, combined with the multi-omics data in TCGA database, successfully dividing HCC into three subtypes, and this classification method was closely related to the immune spectrum [75]. Sia et al. identified two molecular subtypes related to HCC immunity through pathology and immunohistochemistry [76]. Huang et al. combined transcriptome and methylation data to identify four HCC subclasses with significant prognostic differences and high sensitivity and specificity [77]. Compared with previous studies, our research not only combined transcriptome data and DNA methylation profiles but also extracted a large number of IRGs from specialized immunological databases. The LASSO regression analysis based on the consistent clustering analysis was performed on methylation sites that can drive IRGs to construct a prognosis model of HCC. To our knowledge, this is the first model to explore the relationship between immune-related methylation sites and gene expression levels and the prognosis of HCC. Our investigation created a new prognosis for DNA methylation sites of the IRG model. At the same time, we verified the model in the testing group, and the results showed that the model can be used as a prognostic tool for HCC. Finally, we obtained the IRGs driven by the 16 CpG sites in the model, and the verification of these 16 IRGs further demonstrated the accuracy of our model. We also demonstrated that most IRG-related DNA

methylation sites are significantly correlated with the level of infiltration of multiple immune cells in HCC.

Similar to other research works on prognosis models, our study also has some limitations. First of all, there are only 450k methylation data in TCGA database, and although the internal validation results demonstrated the accuracy of the model, further validation is still needed for better clinical application. Furthermore, we did not compare the methylation sites in this model with specific biomarkers of HCC such as serum AFP. If the model can be further validated in the clinic, this may be a worthy research direction. Notably, the efficacy of immunotherapy is based on the degree of infiltration of immune cells in the tumor, and we did not perform immune infiltration analysis on immune genes driven by methylation sites in the model. Nevertheless, we systematically and comprehensively analyzed the application value of the IRG-related DNA methylation prognostic model in HCC and provided a new perspective for the individualized treatment of HCC. We believe that individualized precision treatment in HCC will progress significantly in the coming years.

Conclusion

Based on the DNA methylation data and mRNA expression profiles related to IRGs in TCGA database, we constructed a model of 16 IRG-related methylation sites and further verified their prognostic value in the validation set. The model can help in the identification of new biomarkers as well as precise therapeutic targets in HCC patients. Furthermore, it may be helpful for the prognosis prediction, clinical diagnosis, and the treatment of patients with different epigenetic subtypes of HCC.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

Yes.

Competing interests

The authors declare that they have no conflicts of interest.

Funding

This work is supported by Guangzhou University of Chinese Medicine, Project of Administration of Traditional Chinese Medicine of Guangdong Province of China (20180330164630), First-class Discipline Research Key Project (A1-AFD018181A29), the Student Learning Team Incubation Project of Innovation Academy from The First Affiliated Hospital of Guangzhou University of Chinese Medicine (Award ID: 2018XXTD003), and Guangzhou University of Chinese Medicine National University Student Innovation and Entrepreneurship Training Project (Award ID: 201810572038).

Authors' contributions

Xiongwen Wang and Wenjiang Zheng: Conceptualization, Methodology, Software; Qian Yan: Data curation, Writing- Original draft preparation; Baoqian Ye: Visualization, Investigation; Boqing Wang: Supervision; Wenjiang Zheng: Software, Validation; Qian Yan: Writing- Reviewing and Editing. All authors reviewed the manuscript.

Acknowledgments

We thank Accdon (www.accdon.com) for its linguistic assistance during the preparation of this manuscript.

Figure Legends

Figure 1. The consensus cluster analysis of various DNA methylation prognosis subgroups in hepatocellular carcinoma. (A) The CDF curve among clusters for every category number k . (B) The Delta area curves for consensus clustering (y-axis represents the relative change in area under the CDF curve, and the abscissa means

the category number k), which indicates the relative change in area under the CDF curve for each category number k compared to $k - 1$.

Figure 2. The color-coded heatmap of the consensus matrix for seven molecular subgroups classification obtained by applying the consensus cluster (1 to 7 in the legend represent Cluster 1 to 7); the color gradient indicates the consensus value from 0 to 1; white indicates 0, and dark blue indicates 1).

Figure 3. The heatmap of 840 methylation sites of the seven clusters (the heatmap combined DNA methylation classification with age, gender, TNM stage, and clinicopathological stage as the annotations).

Figure 4. The prognostic difference and the distribution of different clinical factors among seven clusters. (A) Prognosis difference among seven clusters. The horizontal axis represents the survival time (year), the vertical axis represents the survival rate, and the p-value represents the significance of the difference between the clusters. Distribution proportion of age (B), gender (C), grade (D), stage (E), T staging (F), N staging (G), and M staging (H) among seven clusters.

Figure 5. The heatmap of the gene expression of specific CpG sites and the methylation level of differential methylation sites among the seven clusters. (A) Heatmap of 513 gene expressions in seven DNA methylation clusters. (B) Heatmap of differential methylation levels of 238 specific CpG sites for each DNA methylation prognosis subtype. The red bars represent hypermethylated CpG sites or hypomethylated CpG sites, and the blue bars indicate there was no significance in the methylation level of CpG sites among all clusters.

Figure 6. Box plot of the methylation level (z-score) of CpG sites in the seven clusters (compared to other clusters, cluster 5 has the highest methylation level).

Figure 7. Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis of 840 annotated genes associated with prognosis methylation sites (the size of the dots is proportional to the number of genes enriched; the color of the dot is the degree of significance, the degree of significance gradually increases from red to blue, and the line indicates that there is a correlation between the two entries). (A) The first 30 items enriched by the biological process in GO analysis. (B) The first 30 items enriched by cellular component in GO analysis. (C) The first 30

items enriched in molecular function in GO analysis. (D) The first 20 items enriched in KEGG analysis.

Figure 8. The least absolute shrinkage and selection operator regression constructs a prognostic model of DNA methylation sites associated with immune genes in hepatocellular carcinoma. (A) Change track of each independent variable. The horizontal axis represents the logarithmic value of the independent variable λ , and the vertical axis represents the coefficient of the independent variable. (B) The confidence interval of each λ ; the x-axis represents the range of λ values, and the y-axis represents the partial likelihood deviance; when the y-axis takes the minimum values, the λ value and the number above the x-axis are the optimal numbers of prognostic models.

Figure 9. The verification of the stability and reliability of the prognosis prediction model for patients with hepatocellular carcinoma in the training sets. (A) The prognostic differences between high- and low-risk groups in the training set. (B) The distribution of risk scores in high- and low-risk groups in the training set. (C) The relationship between risk score and survival status in the training set (as the risk score increases, the number of deaths gradually increases). (D) The distribution of methylation sites between high- and low-risk groups in the prognostic model (from the low-risk group to the high-risk group, the methylation level of each methylation site gradually decreases). (E) The receiver operating characteristic (ROC) curve in the training set. AUC: area under the curve.

Figure 10. The verification of the stability and reliability of the prognosis prediction model for patients with hepatocellular carcinoma in the test set. (A) The prognostic differences between high- and low-risk groups in the test set. (B) The distribution of risk scores in high- and low-risk groups in the test set. (C) The relationship between risk score and survival status in the test set (as the risk score increases, the number of deaths gradually increases). (D) The distribution of methylation sites between high- and low-risk groups in the prognostic model (from the low-risk group to the high-risk group, the methylation level of each methylation site gradually decreases).

Figure 11. Genetic alterations of immune-related genes driven by 16 CpG sites in the model.

Figure 12. Pearson analysis between the DNA methylation levels of 16 CpG sites in the model and the expression of immune-related genes driven by DNA methylation sites.

Figure 13. Correlation of 16 immune-related genes expression with immune infiltration level in LIHC (From top to bottom: BDNF, CCR6, CIITA, ESR1, FGF1, FGF14, 1L18RAP, LEPR, LTBP1, LTBR, NR1H3, NR6A1, PKA1, PGRMC2, RGL2 and TAPBP).

Table 1: Clinical information of patients with hepatocellular carcinoma in the training and validation set

Table 2: The data of the top 20 CpG sites in 830 methylation sites associated with the overall survival of hepatocellular carcinoma.

Supplementary Tables 1: The top 30 items enriched Biological process GO annotations

Supplementary Tables 2: The top 30 items enriched Cellular component GO annotations

Supplementary Tables 3: The top 30 items enriched Molecular function GO annotations

Supplementary Tables 4: The top 20 items enriched KEGG pathways.

References

1. Ferlay J EM, Lam F, Colombet M, Mery L, Piñeros M, Znaor A, Soerjomataram I, Bray F: **Global Cancer Observatory: Cancer Today**. Lyon, France: International Agency for Research on Cancer. 2018.
2. Villanueva A: **Hepatocellular Carcinoma**. *The New England journal of medicine* 2019, **380**(15):1450-1462.
3. **Global Cancer Observatory: Cancer Tomorrow**. Lyon, France: International Agency for Research on Cancer [<https://gco.iarc.fr/tomorrow>]
4. Kanwal F, Singal AG: **Surveillance for Hepatocellular Carcinoma: Current Best Practice and Future Direction**. *Gastroenterology* 2019, **157**(1):54-64.
5. Yarchoan M, Agarwal P, Villanueva A, Rao S, Dawson LA, Llovet JM, Finn RS, Groopman JD, El-Serag HB, Monga SP *et al*: **Recent Developments and Therapeutic Strategies against Hepatocellular Carcinoma**. *Cancer research* 2019, **79**(17):4326-4330.
6. Llovet JM, Villanueva A, Lachenmayer A, Finn RS: **Advances in targeted therapies for hepatocellular carcinoma in the genomic era**. *Nat Rev Clin Oncol* 2015, **12**(7):408-424.
7. Kubes P, Jenne C: **Immune Responses in the Liver**. *Annu Rev Immunol* 2018, **36**:247-277.
8. Sima P, Vannucci L, Vetvicka V: **Immunity in cancer and atherosclerosis**. *Ann Transl Med* 2019, **7**(9):204.
9. Endig J, Buitrago-Molina LE, Marhenke S, Reisinger F, Saborowski A, Schutt J, Limbourg F, Konecke C, Schreder A, Michael A *et al*: **Dual Role of the Adaptive Immune System in Liver Injury and Hepatocellular Carcinoma Development**. *Cancer Cell* 2016, **30**(2):308-323.
10. Zhou RS, Wang XW, Sun QF, Ye ZJ, Liu JW, Zhou DH, Tang Y: **Anticancer Effects of Emodin on HepG2 Cell: Evidence from Bioinformatic Analysis**. *BioMed research international* 2019, **2019**:3065818.
11. Dong LQ, Peng LH, Ma LJ, Liu DB, Zhang S, Luo SZ, Rao JH, Zhu HW, Yang SX, Xi SJ *et al*: **Heterogeneous immunogenomic features and distinct escape mechanisms in multifocal hepatocellular carcinoma**. *J Hepatol* 2020, **72**(5):896-908.
12. Zongyi Y, Xiaowu L: **Immunotherapy for hepatocellular carcinoma**. *Cancer letters* 2020, **470**:8-17.
13. Llovet JM, Montal R, Sia D, Finn RS: **Molecular therapies and precision medicine for hepatocellular carcinoma**. *Nature reviews Clinical oncology* 2018, **15**(10):599-616.
14. Kulis M, Esteller M: **DNA methylation and cancer**. *Adv Genet* 2010, **70**:27-56.
15. Pollex T, Heard E: **Recent advances in X-chromosome inactivation research**. *Curr Opin Cell Biol* 2012, **24**(6):825-832.
16. Laurent L, Wong E, Li G, Huynh T, Tsirigos A, Ong CT, Low HM, Kin Sung KW, Rigoutsos I, Loring J *et al*: **Dynamic changes in the human methylome during differentiation**. *Genome Res* 2010, **20**(3):320-331.
17. Smith ZD, Meissner A: **DNA methylation: roles in mammalian development**. *Nat Rev Genet* 2013, **14**(3):204-220.

18. Ferguson-Smith AC, Bourc'his D: **The discovery and importance of genomic imprinting.** *Elife* 2018, **7**.
19. Heyn H, Esteller M: **DNA methylation profiling in the clinic: applications and challenges.** *Nat Rev Genet* 2012, **13**(10):679-692.
20. De Carvalho DD, Sharma S, You JS, Su SF, Taberlay PC, Kelly TK, Yang X, Liang G, Jones PA: **DNA methylation screening identifies driver epigenetic events of cancer cell survival.** *Cancer Cell* 2012, **21**(5):655-667.
21. Xu RH, Wei W, Krawczyk M, Wang W, Luo H, Flagg K, Yi S, Shi W, Quan Q, Li K *et al*: **Circulating tumour DNA methylation markers for diagnosis and prognosis of hepatocellular carcinoma.** *Nature materials* 2017, **16**(11):1155-1161.
22. Zheng Y, Huang Q, Ding Z, Liu T, Xue C, Sang X, Gu J: **Genome-wide DNA methylation analysis identifies candidate epigenetic markers and drivers of hepatocellular carcinoma.** *Briefings in bioinformatics* 2018, **19**(1):101-108.
23. Cheng J, Wei D, Ji Y, Chen L, Yang L, Li G, Wu L, Hou T, Xie L, Ding G *et al*: **Integrative analysis of DNA methylation and gene expression reveals hepatocellular carcinoma-specific diagnostic biomarkers.** *Genome medicine* 2018, **10**(1):42.
24. Shen J, Wang S, Zhang YJ, Wu HC, Kibriya MG, Jasmine F, Ahsan H, Wu DP, Siegel AB, Remotti H *et al*: **Exploring genome-wide DNA methylation profiles altered in hepatocellular carcinoma using Infinium HumanMethylation 450 BeadChips.** *Epigenetics* 2013, **8**(1):34-43.
25. Liu Z, Lin Y, Zhang J, Zhang Y, Li Y, Liu Z, Li Q, Luo M, Liang R, Ye J: **Molecular targeted and immune checkpoint therapy for advanced hepatocellular carcinoma.** *Journal of experimental & clinical cancer research : CR* 2019, **38**(1):447.
26. Khemlina G, Ikeda S, Kurzrock R: **The biology of Hepatocellular carcinoma: implications for genomic and immune therapies.** *Molecular cancer* 2017, **16**(1):149.
27. Jones PA, Ohtani H, Chakravarthy A, De Carvalho DD: **Epigenetic therapy in immune-oncology.** *Nature reviews Cancer* 2019, **19**(3):151-161.
28. Neureiter D, Stintzing S, Kiesslich T, Ocker M: **Hepatocellular carcinoma: Therapeutic advances in signaling, epigenetic and immune targets.** *World journal of gastroenterology* 2019, **25**(25):3136-3150.
29. Hong YK, Li Y, Pandit H, Li S, Pulliam Z, Zheng Q, Yu Y, Martin RCG: **Epigenetic modulation enhances immunotherapy for hepatocellular carcinoma.** *Cellular immunology* 2019, **336**:66-74.
30. Goldman M, Craft B, Swatloski T, Cline M, Morozova O, Diekhans M, Haussler D, Zhu J: **The UCSC Cancer Genomics Browser: update 2015.** *Nucleic acids research* 2015, **43**(Database issue):D812-817.
31. Bhattacharya S, Andorf S, Gomes L, Dunn P, Schaefer H, Pontius J, Berger P, Desborough V, Smith T, Campbell J *et al*: **ImmPort: disseminating data to the public for the future of immunology.** *Immunol Res* 2014, **58**(2-3):234-239.
32. Baylin SB, Jones PA: **A decade of exploring the cancer epigenome - biological and translational implications.** *Nat Rev Cancer* 2011, **11**(10):726-734.
33. Zhang Z: **Semi-parametric regression model for survival data: graphical visualization with R.** *Ann Transl Med* 2016, **4**(23):461.

34. Wilkerson MD, Hayes DN: **ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking.** *Bioinformatics (Oxford, England)* 2010, **26**(12):1572-1573.
35. Yu G, Wang LG, Han Y, He QY: **clusterProfiler: an R package for comparing biological themes among gene clusters.** *Omics : a journal of integrative biology* 2012, **16**(5):284-287.
36. Tibshirani R: **The lasso method for variable selection in the Cox model.** *Stat Med* 1997, **16**(4):385-395.
37. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E *et al*: **The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data.** *Cancer discovery* 2012, **2**(5):401-404.
38. Li B, Severson E, Pignon JC, Zhao H, Li T, Novak J, Jiang P, Shen H, Aster JC, Rodig S *et al*: **Comprehensive analyses of tumor immunity: implications for cancer immunotherapy.** *Genome biology* 2016, **17**(1):174.
39. Xu X, Tan Y, Qian Y, Xue W, Wang Y, Du J, Jin L, Ding W: **Clinicopathologic and prognostic significance of tumor-infiltrating CD8+ T cells in patients with hepatocellular carcinoma: A meta-analysis.** *Medicine* 2019, **98**(2):e13923.
40. Ding W, Xu X, Qian Y, Xue W, Wang Y, Du J, Jin L, Tan Y: **Prognostic value of tumor-infiltrating lymphocytes in hepatocellular carcinoma: A meta-analysis.** *Medicine* 2018, **97**(50):e13301.
41. Calderaro J, Ziol M, Paradis V, Zucman-Rossi J: **Molecular and histological correlations in liver cancer.** *Journal of hepatology* 2019, **71**(3):616-630.
42. Schulze K, Nault JC, Villanueva A: **Genetic profiling of hepatocellular carcinoma using next-generation sequencing.** *Journal of hepatology* 2016, **65**(5):1031-1042.
43. Reina-Campos M, Moscat J, Diaz-Meco M: **Metabolism shapes the tumor microenvironment.** *Curr Opin Cell Biol* 2017, **48**:47-53.
44. Lu P, Weaver VM, Werb Z: **The extracellular matrix: a dynamic niche in cancer progression.** *J Cell Biol* 2012, **196**(4):395-406.
45. El-Khoueiry AB, Sangro B, Yau T, Crocenzi TS, Kudo M, Hsu C, Kim TY, Choo SP, Trojan J, Welling THR *et al*: **Nivolumab in patients with advanced hepatocellular carcinoma (CheckMate 040): an open-label, non-comparative, phase 1/2 dose escalation and expansion trial.** *Lancet (London, England)* 2017, **389**(10088):2492-2502.
46. Zhu AX, Finn RS, Edeline J, Cattan S, Ogasawara S, Palmer D, Verslype C, Zagonel V, Fartoux L, Vogel A *et al*: **Pembrolizumab in patients with advanced hepatocellular carcinoma previously treated with sorafenib (KEYNOTE-224): a non-randomised, open-label phase 2 trial.** *Lancet Oncol* 2018, **19**(7):940-952.
47. **Bristol-Myers Squibb Announces Results from CheckMate-459 Study Evaluating Opdivo (nivolumab) as a First-Line Treatment for Patients with Unresectable Hepatocellular Carcinoma**
[<https://news.bms.com/press-release/bmy/bristol-myers-squibb-announces-results-checkmate-459-study-evaluating-opdivo-nivol> (accessed June 25, 2019).]
48. Finn RS, Ryoo BY, Merle P, Kudo M, Bouattour M, Lim HY, Breder V, Edeline J,

- Chao Y, Ogasawara S *et al*: **Pembrolizumab As Second-Line Therapy in Patients With Advanced Hepatocellular Carcinoma in KEYNOTE-240: A Randomized, Double-Blind, Phase III Trial.** *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2020, **38**(3):193-202.
49. Hardy T, Mann DA: **Epigenetics in liver disease: from biology to therapeutics.** *Gut* 2016, **65**(11):1895-1905.
 50. Toh TB, Lim JJ, Chow EK: **Epigenetics of hepatocellular carcinoma.** *Clinical and translational medicine* 2019, **8**(1):13.
 51. Dawson MA, Kouzarides T: **Cancer epigenetics: from mechanism to therapy.** *Cell* 2012, **150**(1):12-27.
 52. Sigalotti L, Coral S, Fratta E, Lamaj E, Danielli R, Di Giacomo AM, Altomonte M, Maio M: **Epigenetic modulation of solid tumors as a novel approach for cancer immunotherapy.** *Semin Oncol* 2005, **32**(5):473-478.
 53. Goltz D, Gevensleben H, Dietrich J, Dietrich D: **PD-L1 (CD274) promoter methylation predicts survival in colorectal cancer patients.** *Oncoimmunology* 2017, **6**(1):e1257454.
 54. Goltz D, Gevensleben H, Dietrich J, Ellinger J, Landsberg J, Kristiansen G, Dietrich D: **Promoter methylation of the immune checkpoint receptor PD-1 (PDCD1) is an independent prognostic biomarker for biochemical recurrence-free survival in prostate cancer patients following radical prostatectomy.** *Oncoimmunology* 2016, **5**(10):e1221555.
 55. Goltz D, Gevensleben H, Grunen S, Dietrich J, Kristiansen G, Landsberg J, Dietrich D: **PD-L1 (CD274) promoter methylation predicts survival in patients with acute myeloid leukemia.** *Leukemia* 2017, **31**(3):738-743.
 56. Zhang Y, Petropoulos S, Liu J, Cheishvili D, Zhou R, Dymov S, Li K, Li N, Szyf M: **The signature of liver cancer in immune cells DNA methylation.** *Clinical epigenetics* 2018, **10**:8.
 57. Dor Y, Cedar H: **Principles of DNA methylation and their implications for biology and medicine.** *Lancet* 2018, **392**(10149):777-786.
 58. Koch A, Joosten SC, Feng Z, de Ruijter TC, Draht MX, Melotte V, Smits KM, Veeck J, Herman JG, Van Neste L *et al*: **Analysis of DNA methylation in cancer: location revisited.** *Nature reviews Clinical oncology* 2018, **15**(7):459-466.
 59. Ozen C, Yildiz G, Dagcan AT, Cevik D, Ors A, Keles U, Topel H, Ozturk M: **Genetics and epigenetics of liver cancer.** *New biotechnology* 2013, **30**(4):381-384.
 60. Deng Q, Huang W, Peng C, Gao J, Li Z, Qiu X, Yang N, Yuan B, Zheng F: **Genomic 5-mC contents in peripheral blood leukocytes were independent protective factors for coronary artery disease with a specific profile in different leukocyte subtypes.** *Clin Epigenetics* 2018, **10**:9.
 61. Jones PA: **Functions of DNA methylation: islands, start sites, gene bodies and beyond.** *Nat Rev Genet* 2012, **13**(7):484-492.
 62. Zang JJ, Xie F, Xu JF, Qin YY, Shen RX, Yang JM, He J: **P16 gene hypermethylation and hepatocellular carcinoma: a systematic review and meta-analysis.** *World J Gastroenterol* 2011, **17**(25):3043-3048.
 63. Jain S, Xie L, Boldbaatar B, Lin SY, Hamilton JP, Meltzer SJ, Chen SH, Hu CT,

- Block TM, Song W *et al*: **Differential methylation of the promoter and first exon of the RASSF1A gene in hepatocarcinogenesis.** *Hepatology research : the official journal of the Japan Society of Hepatology* 2015, **45**(11):1110–1123.
64. Moribe T, Iizuka N, Miura T, Kimura N, Tamatsukuri S, Ishitsuka H, Hamamoto Y, Sakamoto K, Tamesa T, Oka M: **Methylation of multiple genes as molecular markers for diagnosis of a small, well-differentiated hepatocellular carcinoma.** *Int J Cancer* 2009, **125**(2):388–397.
65. Sceusi EL, Loose DS, Wray CJ: **Clinical implications of DNA methylation in hepatocellular carcinoma.** *HPB (Oxford)* 2011, **13**(6):369–376.
66. Wang Q, Wang G, Liu C, He X: **Prognostic value of CpG island methylator phenotype among hepatocellular carcinoma patients: A systematic review and meta-analysis.** *Int J Surg* 2018, **54**(Pt A):92–99.
67. Neumann O, Kesselmeier M, Geffers R, Pellegrino R, Radlwimmer B, Hoffmann K, Ehemann V, Schemmer P, Schirmacher P, Lorenzo Bermejo J *et al*: **Methylome analysis and integrative profiling of human HCCs identify novel protumorigenic factors.** *Hepatology* 2012, **56**(5):1817–1827.
68. Li X, Liu X, Tian L, Chen Y: **Cytokine-Mediated Immunopathogenesis of Hepatitis B Virus Infections.** *Clin Rev Allergy Immunol* 2016, **50**(1):41–54.
69. Sørlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS *et al*: **Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications.** *Proceedings of the National Academy of Sciences of the United States of America* 2001, **98**(19):10869–10874.
70. Tutt A, Tovey H, Cheang MCU, Kernaghan S, Kilburn L, Gazinska P, Owen J, Abraham J, Barrett S, Barrett-Lee P *et al*: **Carboplatin in BRCA1/2-mutated and triple-negative breast cancer BRCAness subgroups: the TNT Trial.** *Nature medicine* 2018, **24**(5):628–637.
71. Kommoss S, Winterhoff B, Oberg AL, Konecny GE, Wang C, Riska SM, Fan JB, Maurer MJ, April C, Shridhar V *et al*: **Bevacizumab May Differentially Improve Ovarian Cancer Outcome in Patients with Proliferative and Mesenchymal Molecular Subtypes.** *Clinical cancer research : an official journal of the American Association for Cancer Research* 2017, **23**(14):3794–3801.
72. Li C, Ke J, Liu J, Su J: **DNA methylation data-based molecular subtype classification related to the prognosis of patients with cervical cancer.** *J Cell Biochem* 2020, **121**(3):2713–2724.
73. Zucman-Rossi J, Villanueva A, Nault JC, Llovet JM: **Genetic Landscape and Biomarkers of Hepatocellular Carcinoma.** *Gastroenterology* 2015, **149**(5):1226–1239. e1224.
74. Hoshida Y, Toffanin S, Lachenmayer A, Villanueva A, Minguez B, Llovet JM: **Molecular classification and novel targets in hepatocellular carcinoma: recent advancements.** *Semin Liver Dis* 2010, **30**(1):35–51.
75. Shimada S, Mogushi K, Akiyama Y, Furuyama T, Watanabe S, Ogura T, Ogawa K, Ono H, Mitsunori Y, Ban D *et al*: **Comprehensive molecular and immunological characterization of hepatocellular carcinoma.** *EBioMedicine* 2019, **40**:457–470.

76. Sia D, Jiao Y, Martinez-Quetglas I, Kuchuk O, Villacorta-Martin C, Castro de Moura M, Putra J, Camprecios G, Bassaganyas L, Akers N *et al*: **Identification of an Immune-specific Class of Hepatocellular Carcinoma, Based on Molecular Features.** *Gastroenterology* 2017, **153**(3):812-826.
77. Huang X, Yang C, Wang J, Sun T, Xiong H: **Integrative analysis of DNA methylation and gene expression reveals distinct hepatocellular carcinoma subtypes with therapeutic implications.** *Aging (Albany NY)* 2020, **12**(6):4970-4995.

Figures

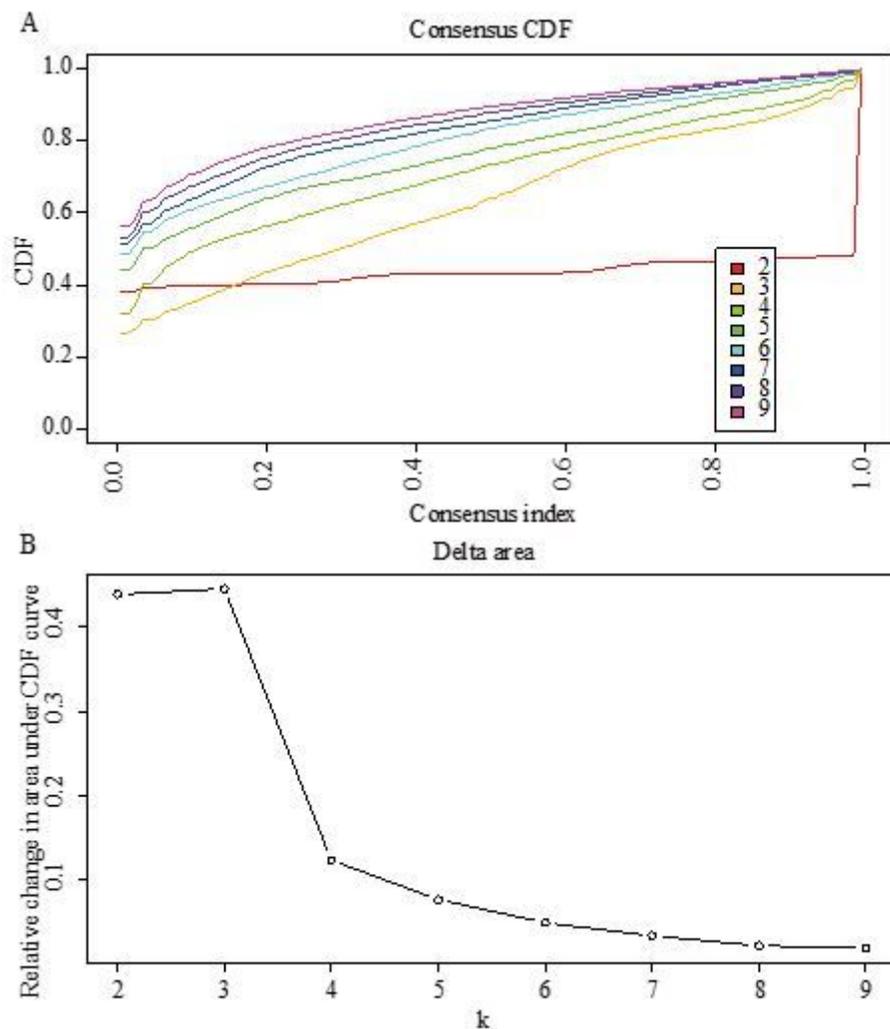


Figure 1

The consensus cluster analysis of various DNA methylation prognosis subgroups in hepatocellular carcinoma. (A) The consensus cumulative distribution function (CDF) curve among clusters for every category number k . (B) The Delta area curves for consensus clustering (y-axis represents the relative change in area under the CDF curve, and the abscissa means the category number k), which indicates the relative change in area under the CDF curve for each category number k compared to $k - 1$.

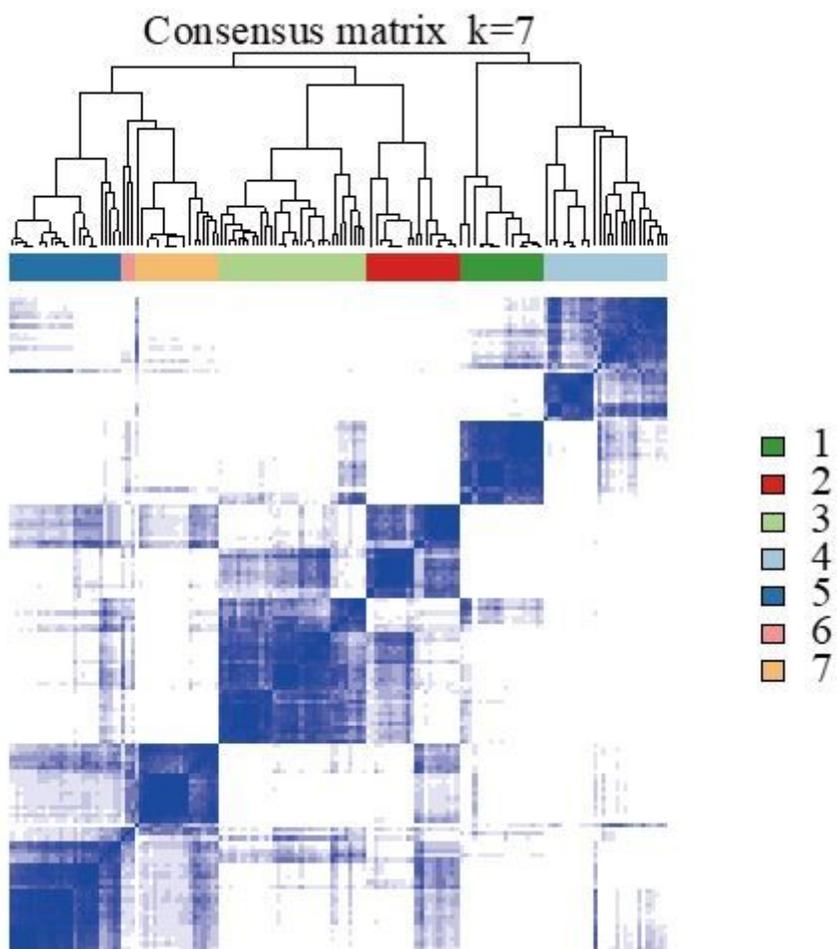


Figure 2

The color-coded heatmap of the consensus matrix for seven molecular subgroups classification obtained by applying the consensus cluster (1 to 7 in the legend represent Cluster 1 to 7); the color gradient indicates the consensus value from 0 to 1; white indicates 0, and dark blue indicates 1).

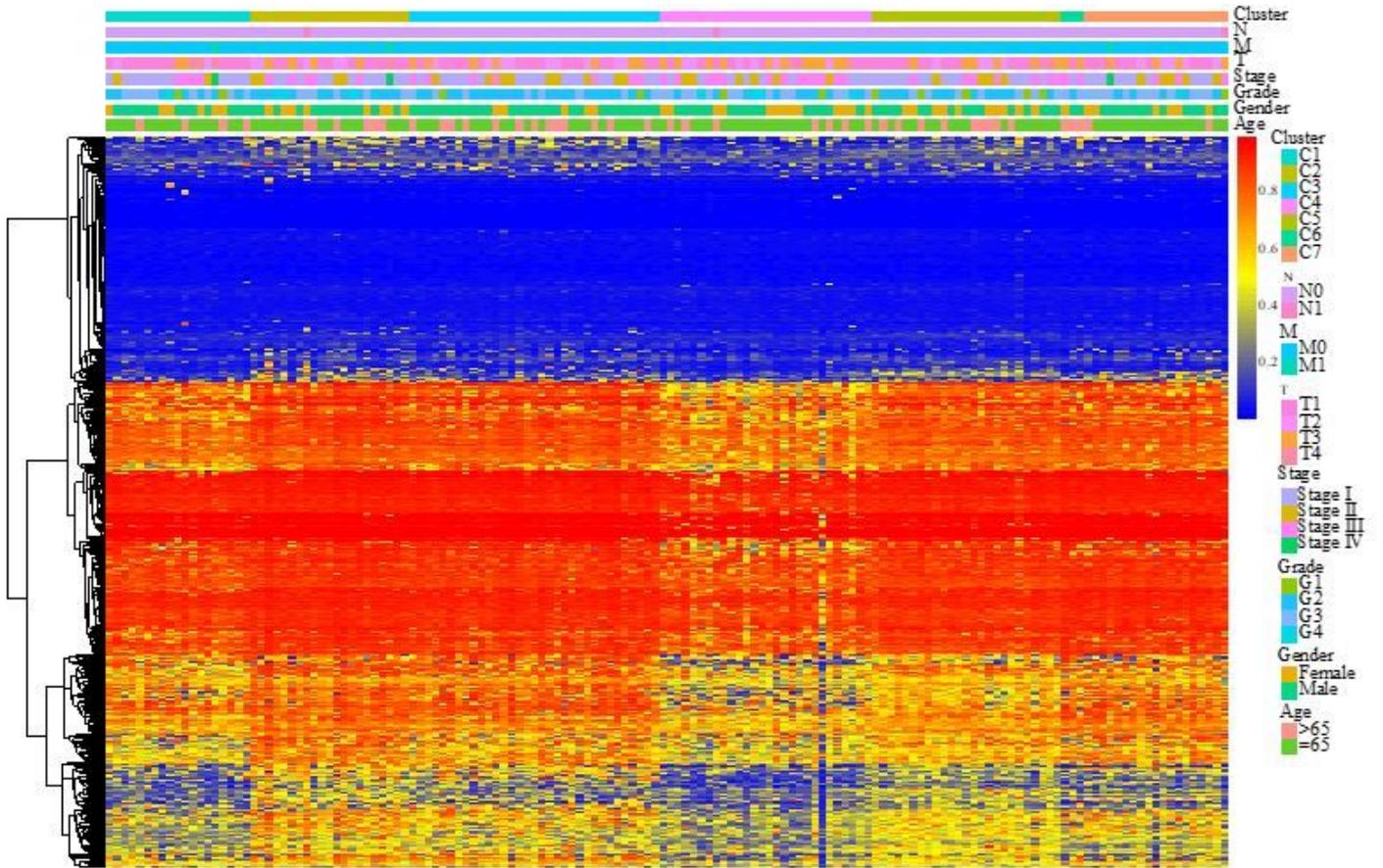


Figure 3

The heatmap of 840 methylation sites of the seven clusters (the heatmap combined DNA methylation classification with age, gender, TNM stage, and clinicopathological stage as the annotations).

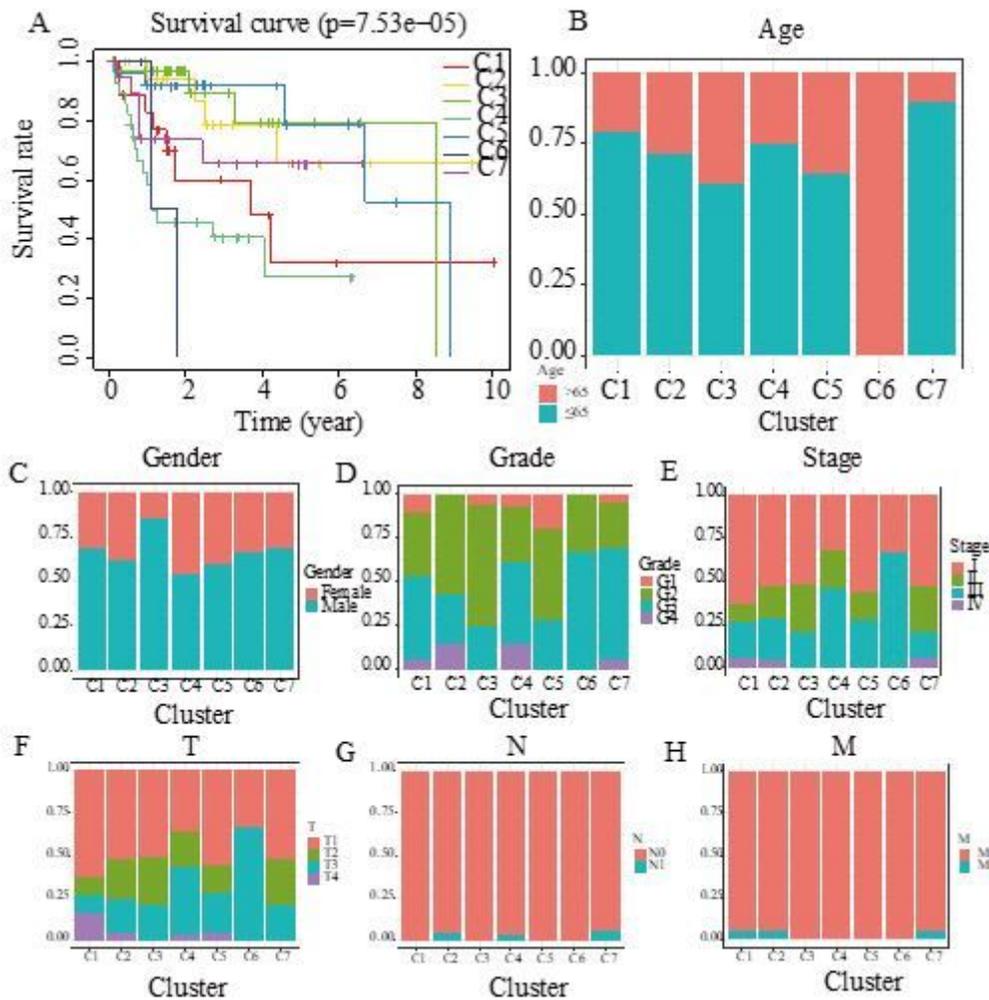


Figure 4

The prognostic difference and the distribution of different clinical factors among seven clusters. (A) Prognosis difference among seven clusters. The horizontal axis represents the survival time (year), the vertical axis represents the survival rate, and the p-value represents the significance of the difference between the clusters. Distribution proportion of age (B), gender (C), grade (D), stage (E), T staging (F), N staging (G), and M staging (H) among seven clusters.

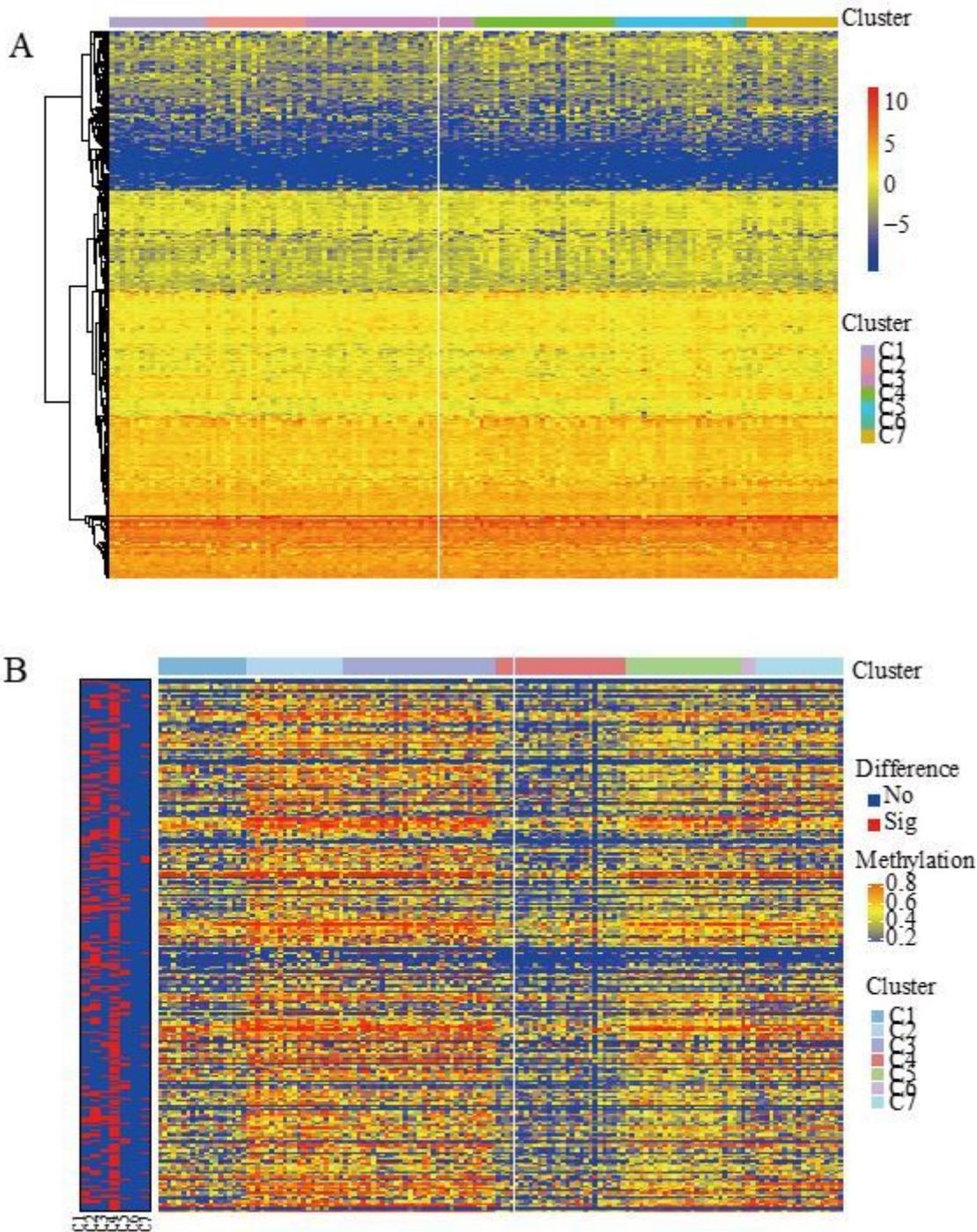


Figure 5

The heatmap of the gene expression of specific CpG sites and the methylation level of differential methylation sites among the seven clusters. (A) Heatmap of 513 gene expressions in seven DNA methylation clusters. (B) Heatmap of differential methylation levels of 238 specific CpG sites for each DNA methylation prognosis subtype. The red bars represent hypermethylated CpG sites or hypomethylated CpG sites, and the blue bars indicate there was no significance in the methylation level of CpG sites among all clusters.

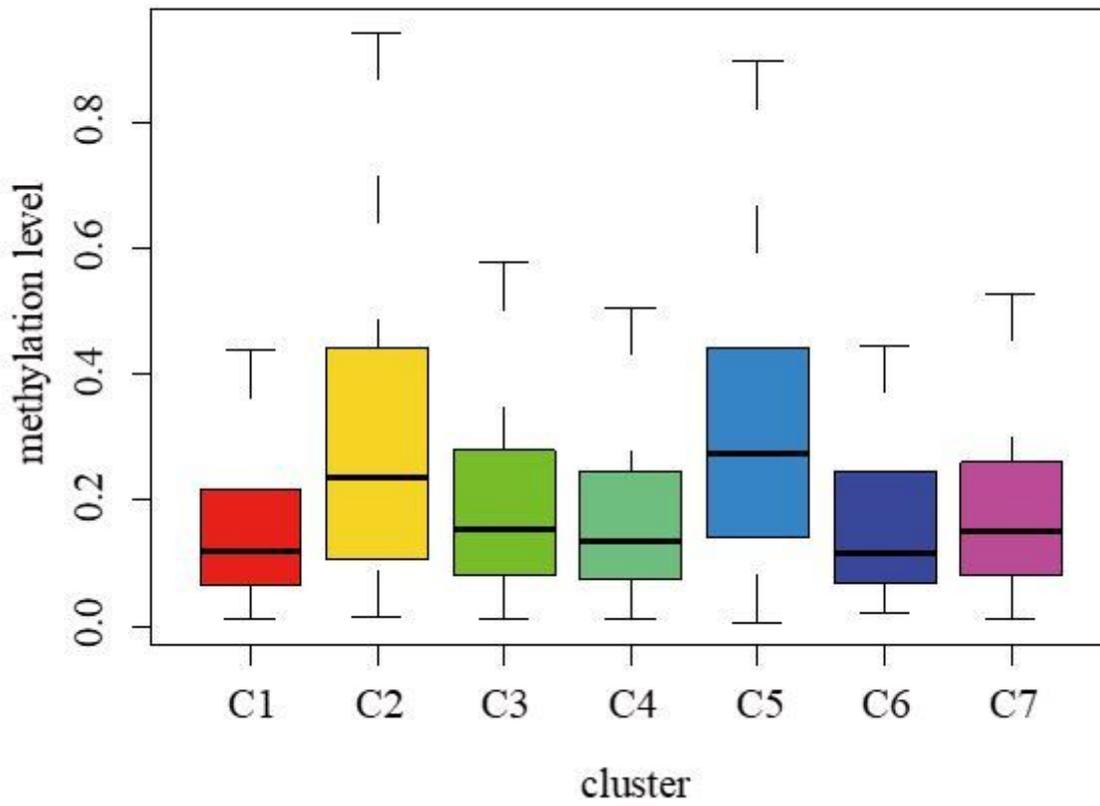


Figure 6

Box plot of the methylation level (z-score) of CpG sites in the seven clusters (compared to other clusters, cluster 5 has the highest methylation level).

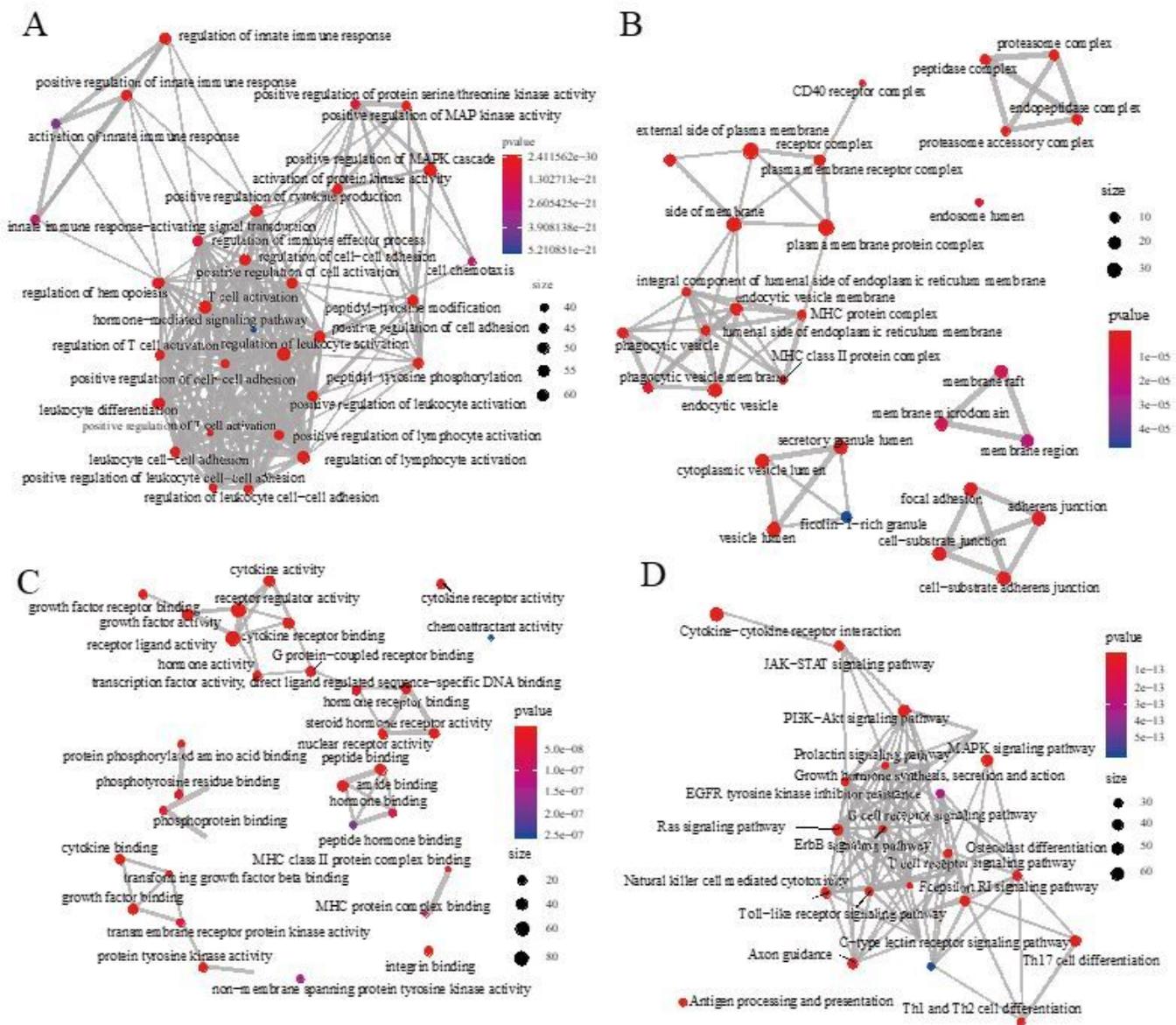


Figure 7

Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis of 840 annotated genes associated with prognosis methylation sites (the size of the dots is proportional to the number of genes enriched; the color of the dot is the degree of significance, the degree of significance gradually increases from red to blue, and the line indicates that there is a correlation between the two entries). (A) The first 30 items enriched by the biological process in GO analysis. (B) The first 30 items enriched by cellular component in GO analysis. (C) The first 30 items enriched in molecular function in GO analysis. (D) The first 20 items enriched in KEGG analysis.

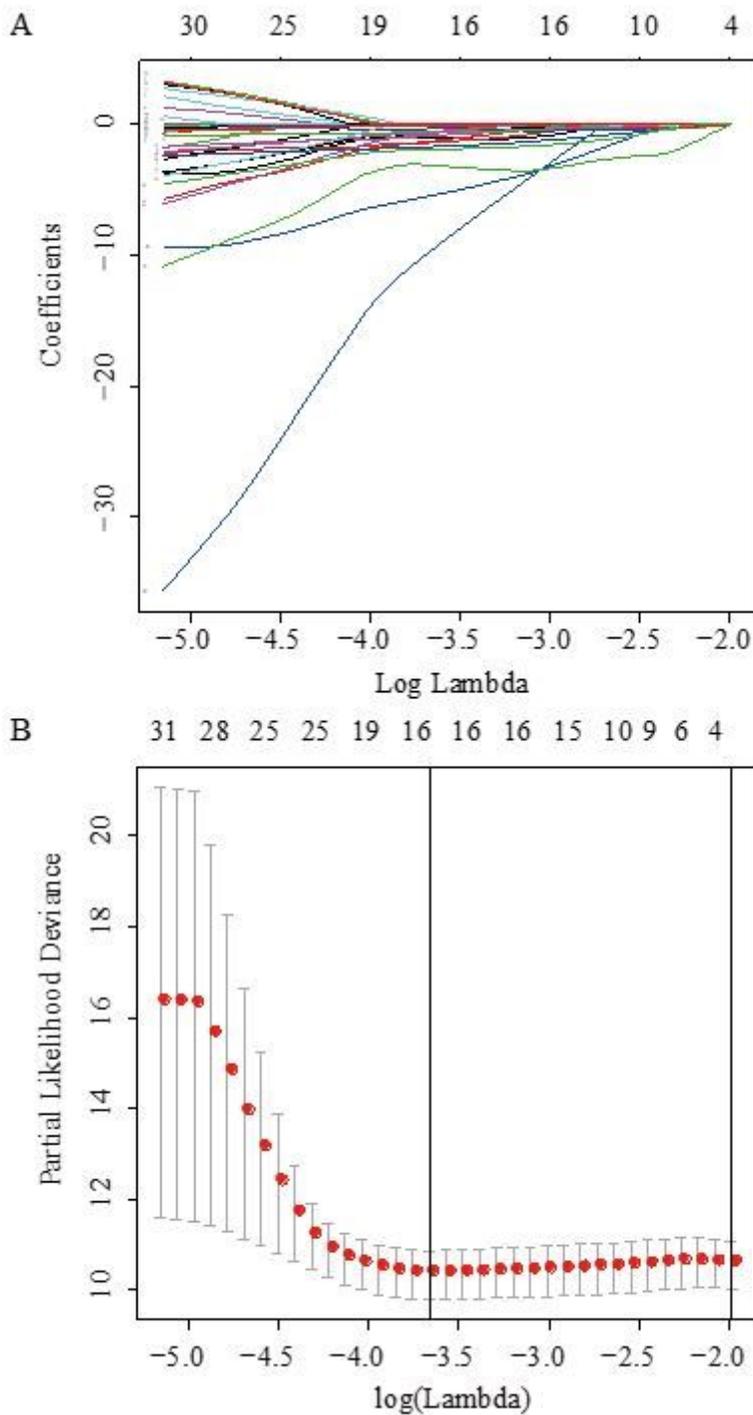


Figure 8

The least absolute shrinkage and selection operator regression constructs a prognostic model of DNA methylation sites associated with immune genes in hepatocellular carcinoma. (A) Change track of each independent variable. The horizontal axis represents the logarithmic value of the independent variable λ , and the vertical axis represents the coefficient of the independent variable. (B) The confidence interval of each λ ; the x-axis represents the range of λ values, and the y-axis represents the partial likelihood

deviance; when the y-axis takes the minimum values, the λ value and the number above the x-axis are the optimal numbers of prognostic models.

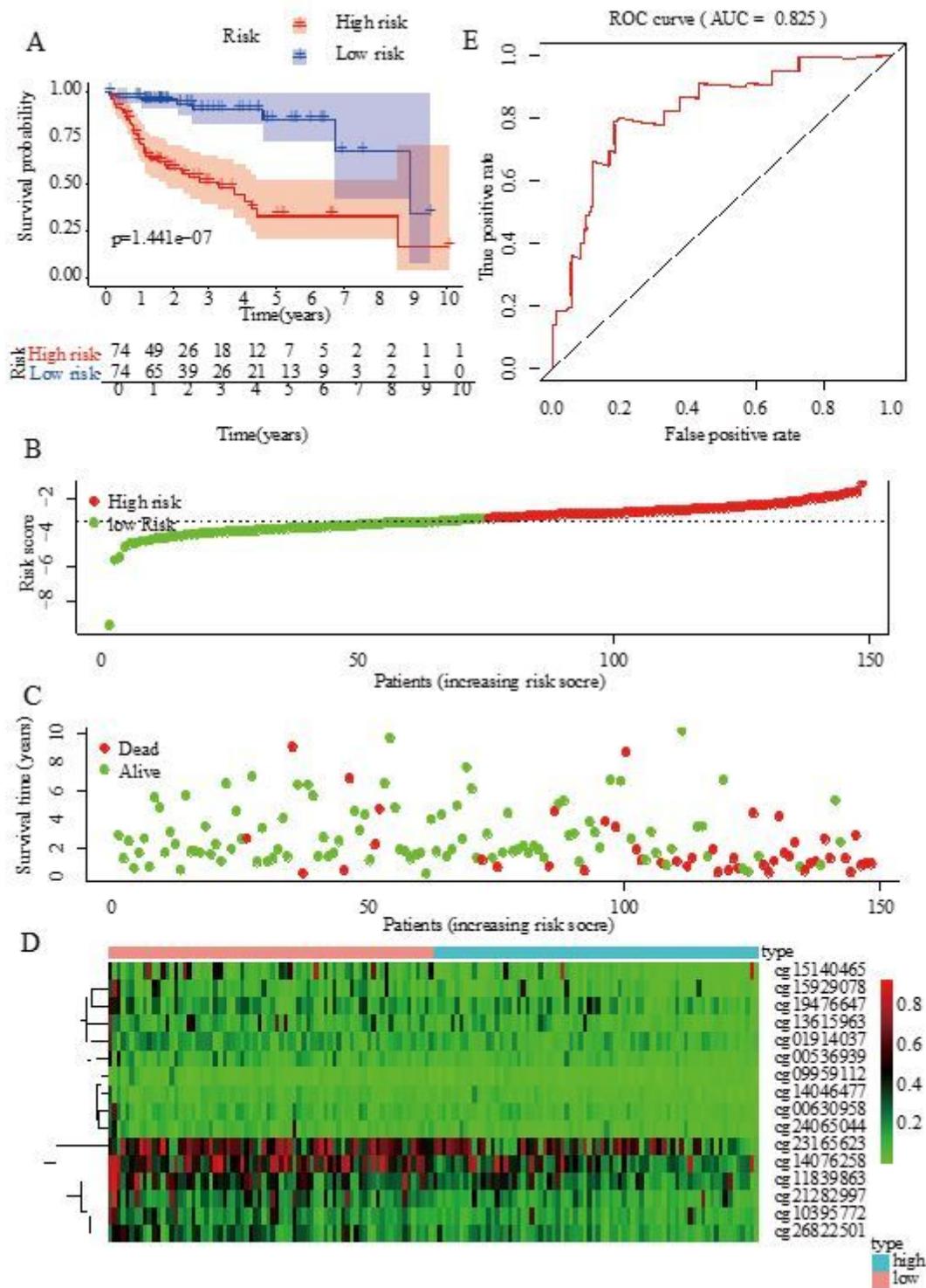


Figure 9

The verification of the stability and reliability of the prognosis prediction model for patients with hepatocellular carcinoma in the training sets. (A) The prognostic differences between high- and low-risk groups in the training set. (B) The distribution of risk scores in high- and low-risk groups in the training

set. (C) The relationship between risk score and survival status in the training set (as the risk score increases, the number of deaths gradually increases). (D) The distribution of methylation sites between high- and low-risk groups in the prognostic model (from the low-risk group to the high-risk group, the methylation level of each methylation site gradually decreases). (E) The receiver operating characteristic (ROC) curve in the training set. AUC: area under the curve.

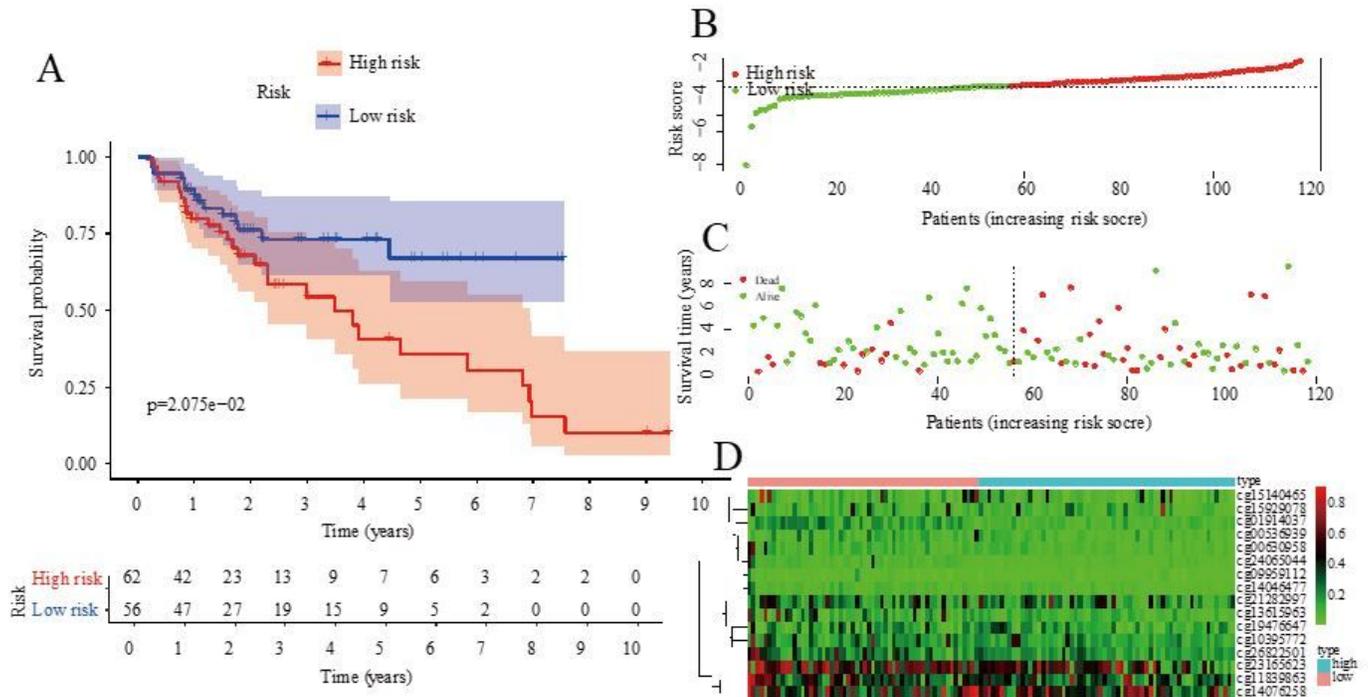


Figure 10

The verification of the stability and reliability of the prognosis prediction model for patients with hepatocellular carcinoma in the test set. (A) The prognostic differences between high- and low-risk groups in the test set. (B) The distribution of risk scores in high- and low-risk groups in the test set. (C) The relationship between risk score and survival status in the test set (as the risk score increases, the number of deaths gradually increases). (D) The distribution of methylation sites between high- and low-risk groups in the prognostic model (from the low-risk group to the high-risk group, the methylation level of each methylation site gradually decreases).

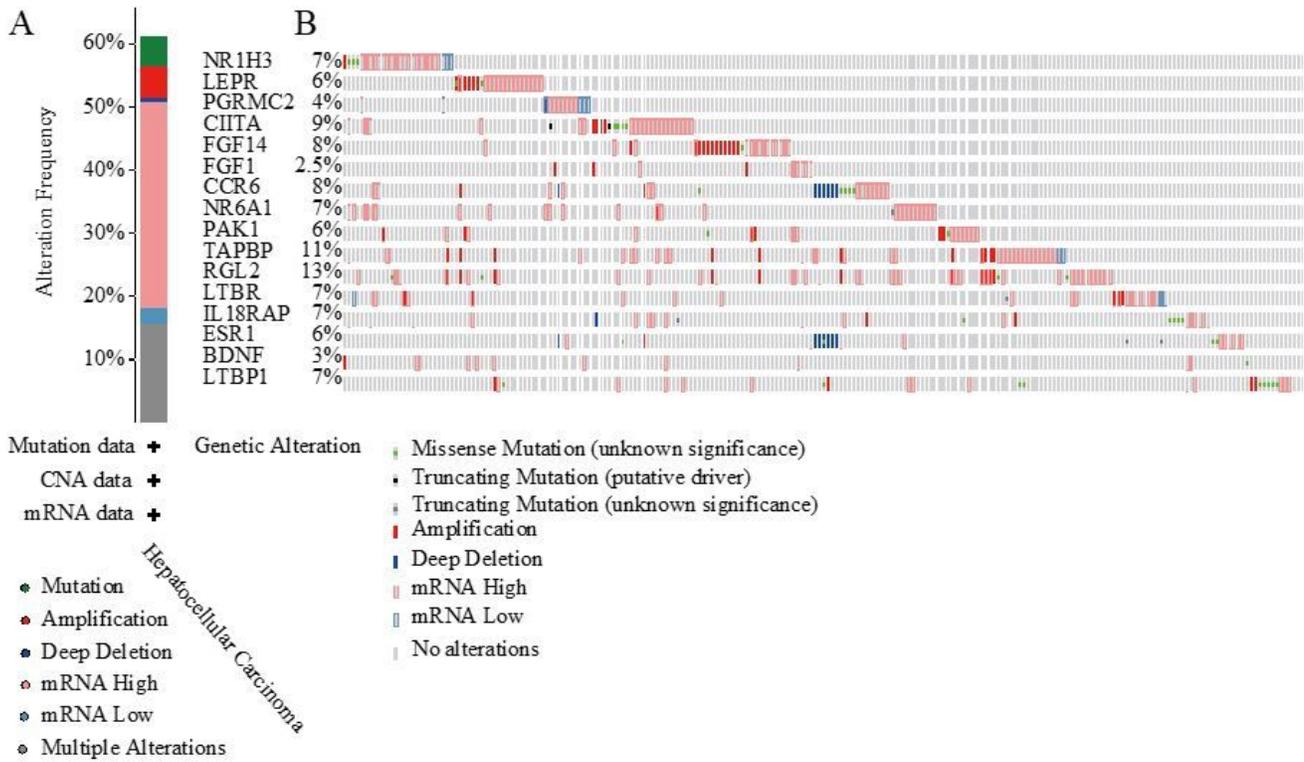


Figure 11

Genetic alterations of immune-related genes driven by 16 CpG sites in the model.

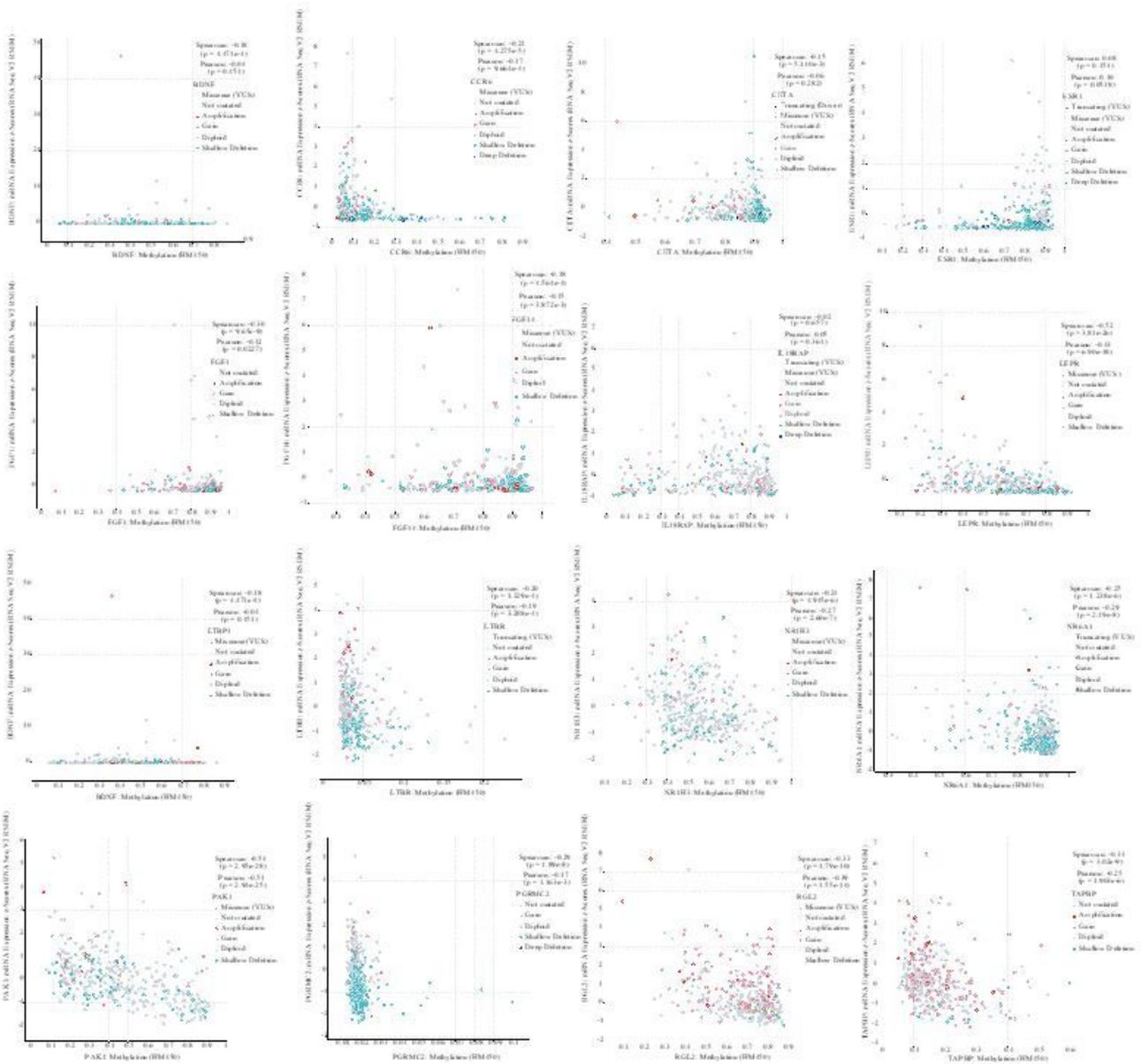


Figure 12

Pearson analysis between the DNA methylation levels of 16 CpG sites in the model and the expression of immune-related genes driven by DNA methylation sites.

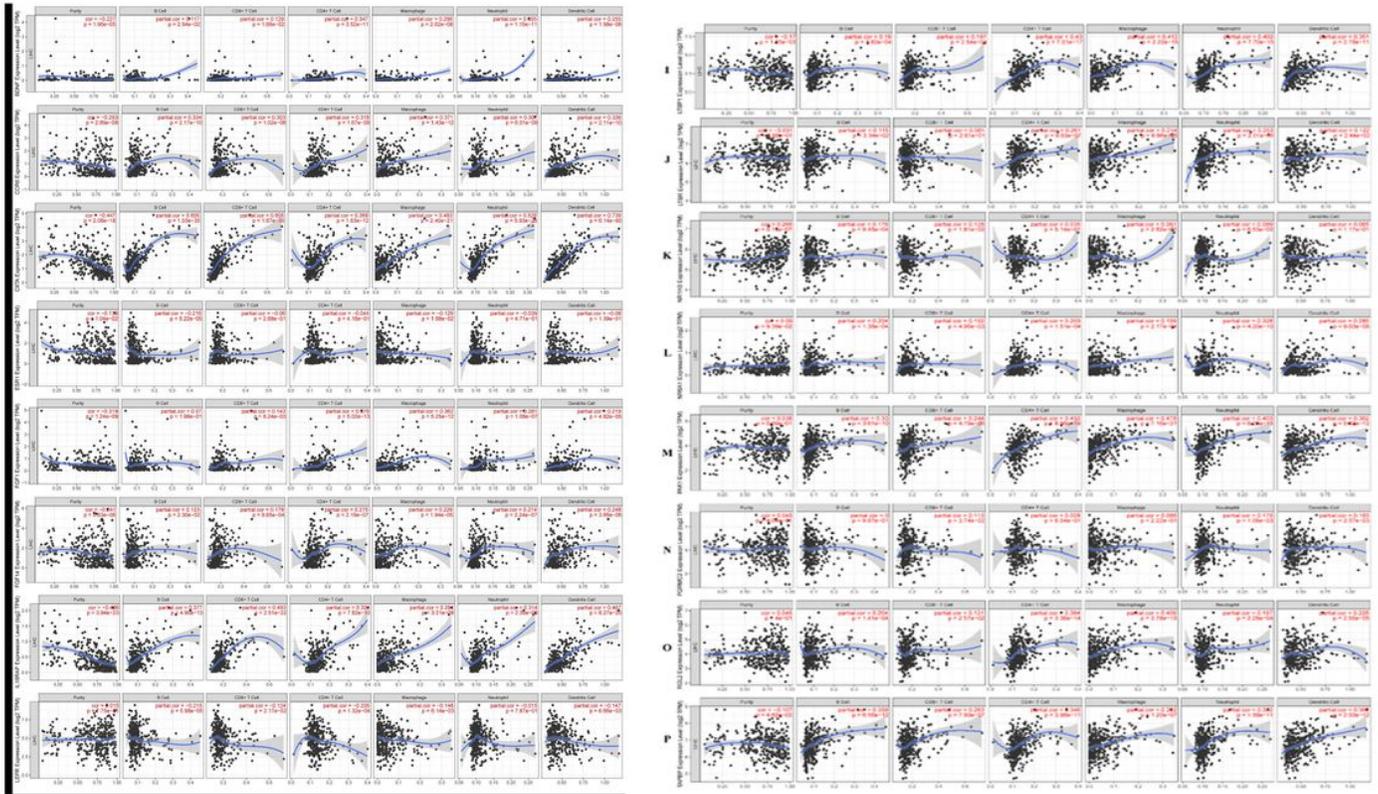


Figure 13

Correlation of 16 immune-related genes expression with immune infiltration level in LIHC (From top to bottom: BDNF, CCR6, CIITA, ESR1, FGF1, FGF14, 1L18RAP, LEPR, LTBP1, LTBR, NR1H3, NR6A1, PKA1, PGRMC2, RGL2 and TAPBP).

Supplementary Files

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

- [SupplementaryTable1.xlsx](#)
- [SupplementaryTable2.xlsx](#)
- [SupplementaryTable3.xlsx](#)
- [SupplementaryTable4.xlsx](#)