Multiomics Study of Key Differentially Expressed Methylation Genes in HCC Based on Epigenome and Transcriptome Analyses

Yu Wang
First Affiliated Hospital of Dalian Medical University

Fei Qi
First Affiliated Hospital of Dalian Medical University

Xun Wang
First Affiliated Hospital of Dalian Medical University

Shi Jin (jinshi@dmu.edu.cn)
First Affiliated Hospital of Dalian Medical University

Hui He
First Affiliated Hospital of Dalian Medical University

Primary research

Keywords: methylation, hepatocellular carcinoma, CHST4, CRHBP, IGFBP3

DOI: https://doi.org/10.21203/rs.3.rs-439115/v1

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

**Objective:** To screen out key differentially expressed methylation genes in hepatocellular carcinoma (HCC) by extracting data from the NCBI-GEO and TCGA databases, discuss the interaction of key genes in HCC and explore the influence of these genes on tumor progression.

**Methods:** The clinical information and sequence data of HCC patients and healthy controls were extracted from the NCBI-GEO and TCGA databases. The module that presented the highest correlation with the tumor phenotypes was selected by the WGCNA network and TOM analysis. GO and KEGG analyses were used for signaling pathway analysis. Key differentially expressed genes were screened out from the module. KM plot analysis was performed to determine the impact of key genes on patient survival. The relationship between core genes and tumor immune infiltration was also discussed. Drug sensitivity analysis was performed to observe the sensitivity of key genes methylated by chemotherapeutic drugs and thus determine the associated effects on patient prognosis. Finally, the molecular mechanisms of the genes involved in tumor progression were analyzed by GSVA, and the interactions between key genes were revealed using Gene MANIA analysis.

**Results:** A total of 373 differentially expressed genes were screened, including 88 upregulated genes and 285 downregulated genes. Three of these genes, CHST4, CRHBP, and IGFBP3, were found to be abnormally methylated in HCC patients and significantly associated with expected survival. The relationship between these key genes and tumor immune invasion was confirmed, and these genes play a protumor or antitumor role through immune cell mediation. In addition, we found that CHST4 and IGFBP3 expression can significantly reduce the sensitivity of several commonly used chemotherapy drugs to tumor cells, while CRHBP expression plays a synergistic role with chemotherapy drugs to enhance mutual effects. The participation of these genes in specific signaling pathways involved in liver metabolism and tumor progression has also been identified. Finally, complex interactions between the three core genes and other related genes were presented.

**Conclusion:** The CHST4, CRHBP and IGFBP3 genes present significant methylation and differential expression in HCC, are involved in tumor immune infiltration, and affect chemotherapy drug sensitivity. As potential biomarkers and therapeutic targets, they may be beneficial to the diagnosis and prognosis of HCC patients in the future.

Introduction

HCC is an extremely dangerous and common malignant tumor originating from hepatocytes and accounts for approximately 80% of all primary liver cancers. In 2018, HCC presented the sixth highest incidence among cancers and was the fourth leading cause of cancer death in the world, thus contributing to 841,000 new cases and 782,000 deaths annually[1]. HCC is generally considered to present a lengthy development process caused by various complex factors, including chronic hepatitis B virus (HBV) and hepatitis C virus (HCV) infections, aflatoxin B1, alcoholism, smoking, water pollution,
metabolic syndrome, heredity factors and medication[2, 3]. Significant regional differences are observed in the distribution of HCC. The majority of adult cases (80%) occur in South Africa and East Asia, where people have long suffered from hepatitis and aflatoxin B1. However, HCV and excessive alcohol consumption are major risk factors in the USA and Europe. In addition, surgical difficulty and poor prognosis are commonly observed for HCC due to its late diagnosis and high recurrence rate[2, 4, 5]. Currently, the combination of liver ultrasound and alpha-fetoprotein (AFP) is still the main method of identifying HCC at the very early stage. Unfortunately, the reported sensitivities were only 63% when imaging and serology were adopted simultaneously. Even HCC nodules detected early carry up to a 70% risk of five-year recurrence after radical resection[6]. Therefore, more accurate detection and more effective treatment protocols are urgently needed.

The occurrence of HCC is considered a negative outcome of chronic interactions involving multiple pathways. Although the specific pathogenesis of HCC remains a mystery, genetic and epigenetic abnormalities have been confirmed to exist in most tumor cells, and they include genome deletion, amplification, insertion, and abnormal DNA methylation[7-9]. DNA methylation in eukaryotes refers to the formation of 5-methylcytosine via the addition of a methyl group to the carbon-5 position of cytosine in the CpG dinucleotide sequence. This modification can inhibit downstream gene expression without changing the nucleotide order and structure. DNA methylation is an important factor in cell differentiation and proliferation. Abnormal methylation of tumor suppressor genes and demethylation of oncogenes are key events in tumorigenesis and progression, including HCC. Abnormal DNA methylation is triggered when risk factors, such as alcohol, HBV or HCV, impact the liver to initiate the inflammatory response. Once abnormalities accumulate to a certain extent, the normal cells of the liver will become tumorigenic, thereby accelerating the development of HCC[10]. DNA methylation abnormalities usually occur at the embryonic stage of HCC, even before the appearance of HCC[11, 12], and they are potentially reversible[10]. Therefore, it is of great clinical significance to identify differentially expressed methylation genes and investigate the regulatory mechanism of tumor cells at the epigenetic level because such information could help identify new biomarkers and effective targeted therapies, which may lead to early screening and improved long-term survival.

In this article, we integrated the data collected from the National Center for Biotechnology Information Gene Expression Omnibus (NCBI-GEO) and The Cancer Genome Atlas (TCGA). The key differentially expressed methylation genes were screened by a standard methylation process analysis and weighted gene coexpression network analysis (WGCNA). Then, verification experiments were carried out to confirm the conclusions. In summary, our study will provide insights for future breakthroughs in targeted therapies for HCC.

Material And Methods

1. Data download
NCBI-GEO (https://www.ncbi.nlm.nih.gov/geo/) is a gene expression database that stores chip and high-throughput sequencing data created and maintained by the NCBI. We downloaded the series matrix file of the GSE57957 dataset from the NCBI-GEO public database, and it contained transcriptome data from 78 groups, including healthy controls (n=39) and HCC patients (n=39). A difference analysis was performed, and the selection criteria were |LogFC|>1 and P-value <0.05. Similarly, methylation data for the series matrix file of the GSE113017 dataset, including healthy controls (n=30) and HCC patients (n=30), were downloaded from NCBI-GEO. The CHAMP function package was used for a 450K differential methylation site analysis. The screening required that one group had a β-value less than 0.2 and the other group had a β-value greater than 0.3. An adjusted P-value <0.05 was adopted as the cutoff criterion.

2. GO and KEGG pathway analysis

Gene Ontology (GO) is known for its ability to define and describe the function of genes in molecular function, biological process and cellular components categories. Kyoto Encyclopedia of Genes and Genomes (KEGG) is a comprehensive database that integrates genomic, chemical and systematic information, with an emphasis on storing information on gene pathways in different species. The Metascape database (www.metascape.org) was utilized to annotate and visualize the biological functions and signaling pathways involved in the development of HCC. Then, GO and KEGG pathway enrichment analyses were performed for the selected specific genes. Min overlap values ≥3 and P-values ≤0.01 were considered statistically significant.

3. WGCNA

WGCNA is one type of analysis that is performed to identify coexpression gene modules and explore the location of core genes and the association between genes and phenotypes by constructing a weighted gene coexpression network. The WGCNA-R package was used to construct the coexpression network for all the genes contained in the dataset. Then, we screened the top 5,000 genes presenting variance with this algorithm for further analysis. The soft threshold was set as 6. For the estimation of network connectivity, a weighted adjacency matrix was transformed into a topological overlap matrix (TOM). Moreover, a hierarchical clustering method was utilized to construct the clustering tree structure of the TOM. The different modules were colored differently to distinguish the different gene modules on the branches of the cluster tree. Based on the weighted correlation coefficient, the screened genes were classified according to the expression pattern. Genes with similar patterns can be grouped into a single module; therefore, 5000 genes were divided into multiple regular modules through gene expression patterns.

4. Data collection from TCGA

TCGA (https://portal.gdc.cancer.gov/) is the largest known cancer gene information database and stores large amounts of data, including data on gene expression, miRNA expression, copy number variation, DNA methylation, SNP, etc. We downloaded the processed original data of mRNA expression in HCC from
the TCGA database. A total of 369 patient specimens with complete expression profiles and clinical information were collected for follow-up analysis.

5. **Drug sensitivity analysis**

Genomics of Drug Sensitivity in Cancer (GDSC, https://www.cancerrxgene.org/) is the largest database of pharmacogenomics, and it was used in conjunction with the R software package "pRRophetic" to predict the chemotherapy sensitivity of each tumor sample. The estimated IC$_{50}$ for each specific chemotherapeutic agent was obtained by regression analysis. Regression and prediction accuracy were guaranteed by 10 cross-validation tests with the GDSC training set. Default values were selected for all parameters, including ComBat with removing batch effects and average repeated gene expression.

6. **GSVA**

Gene set variation analysis (GSVA) is a nonparametric and unsupervised method of evaluating the enrichment of transcriptome genomes, and it converts changes from the gene level to the pathway level to facilitate the determination of the biological function in the sample by comprehensively scoring gene sets of interest. In this study, gene datasets were downloaded from the Molecular Signatures Database (Version 7.0). GSVA was adopted to comprehensively score each gene dataset for the evaluation of potential biological function changes among different samples.

7. **GeneMANIA analysis**

GeneMANIA (http://www.genemania.org) is a flexible and user-friendly PPI network-building database for visualizing gene functional networks and analyzing functions and interactions among related genes. The website can set up the data source of gene nodes. Additionally, it includes a variety of bioinformatics analysis methods, such as physical interaction, gene coexpression, gene colocation, gene enrichment analysis and website prediction. In this search, the interaction network of core genes formed by GeneMANIA was used to explore and explain the possible mechanism impacting the core genes in HCC patients.

8. **Statistical analysis**

All statistical analyses were conducted using R (Version 3.6). All statistical tests were bilateral. A P-value <0.05 was considered statistically significant.

**Results**

1. **Module selection**

First, we downloaded transcriptome data for 78 patient groups in the GSE57957 dataset from the NCBI-GEO public database, including healthy controls (n=39) and HCC patients (n=39). The WGCNA network was constructed to explore and screen the key regulatory genes in HCC according to the clinical
characteristics of patients. For comparison, HCC patient groups were marked as orange while control groups were marked as white (Fig. 1A). The soft threshold $\beta$-value was determined by sft$PowerEstimate. We selected the most accurate value greater than the baseline of 0.9 for scale independence and close to 0 for mean connectivity. Therefore, the soft threshold was set as 6 (Fig. 1B). Then, 10 gene modules were detected based on the TOM analysis, including black (225), blue (1261), brown (763), green (241), gray (199), magenta (98), pink (200), red (234), turquoise (1316) and yellow (463) modules. We further analyzed the relationship between modules and traits and found that the magenta module had the highest module-trait relationship ($\text{cor}=-0.74, p=7e^{-15}$) (Fig. 2C). Moreover, the relationship between module membership and gene significance was significant in the magenta module ($\text{cor}=0.78, p=3e^{-21}$) (Fig. 2D). An additional file shows this in more detail (see Additional file 1). Therefore, the magenta module was chosen for subsequent validation.

2. **Signaling pathways by enrichment analysis**

Once the magenta module was selected, all genes in the module were analyzed by GO and KEGG. Our results showed that the genes of the magenta module were mainly enriched in the following mechanisms closely related to the occurrence and progression of HCC: collagen-containing extracellular matrix, complement activation, lectin pathway, blood vessel development, cytokine receptor binding and other signaling mechanisms (Fig. 2.1.A-B). This outcome was consistent with our expectation, suggesting that the mechanism of the magenta module that affects HCC progression is intimately related to the pathways mentioned above. The interaction between genes in the module is shown in the PPI diagram (Fig. 2.2).

3. **Differentially methylated genes in HCC**

We downloaded the GSE113017 methylation dataset (450K chip) with 60 groups of volunteers from the NCBI-GEO public database, including healthy control groups (n=30) and HCC patient groups (n=30). Subsequently, the standard methylation process analysis was performed using the CHAMP package. The results showed that a total of 4614 differentially methylated sites were screened, and they included 3838 upregulated sites and 776 downregulated sites (Fig. 3). Additional files show these in more detail (see Additional file 2-5). Additionally, the differentially expressed genes of HCC patients in the GSE57957 dataset were analyzed by the LIMMA packet. After comparison, a total of 373 differentially expressed genes were selected, including 88 upregulated genes and 285 downregulated genes (Fig. 4). Additional files show these in more detail (see Additional file 6-8). The cutoff criterion for gene screening was $|\text{LogFC}|>1$ with a P-value <0.05. Based on these data, a combined analysis was performed between methylation groups and transcription groups. The results showed that there were 12 genes with upregulated methylation and downregulated expression and 6 genes with downregulated methylation and upregulated expression existing in both datasets concurrently. A further WGCNA revealed that there were 4 overlapping genes: carbohydrate sulfotransferase 4 (CHST4), corticotropin releasing hormone binding protein (CRHBP), insulin like growth factor binding protein 3 (IGFBP3) and parathyroid hormone 1 receptor (PTH1R).
4. Genes affecting patient survival

We downloaded the mRNA expression data of HCC patients from the TCGA database. A total of 369 HCC patients with complete survival information and expression profiles were included. The survival status of patients with HCC under the influence of each key gene is shown in Figure 6. A KM plot analysis revealed that CHST4 (p=1.534e$^{-2}$), CRHBP (p=9.793e$^{-3}$) and IGFBP3 (p=1.373e$^{-2}$) significantly affected the survival of patients, while no significant association was found between PTH1R and patient survival (p=1.463e$^{-1}$). Next, we focused on exploring the potential mechanism of action among these key genes in HCC patients and describing the molecular map of HCC-related multiomics, which will contribute to the treatment and management of HCC and improve the prognosis of patients.

5. Key genes influencing tumor immune invasion

The tumor microenvironment is a functional whole that is mainly composed of tumor-related fibroblasts, immune cells, extracellular matrix, various growth factors, inflammatory factors, special physical and chemical characteristics and tumor cells themselves. The occurrence, growth and metastasis of tumor cells are closely related to their internal and external environment, which significantly affects the diagnosis, survival outcome and clinical treatment sensitivity of tumors. Therefore, we analyzed the relationship between the core genes obtained from the TCGA database that are involved in tumor immune invasion to further explore the potential molecular mechanism underlying the influence of the core genes on the progression of HCC. The results revealed that three key genes were significantly correlated with the content of immune cells (Fig. 7). Among them, CHST4 was positively correlated with neutrophils, M0 macrophages and resting dendritic cells but negatively correlated with activated NK cells, CD8 T cells and gamma delta T cells; CRHBP was positively correlated with naive B cells, plasma cells, and resting memory CD4 T cells, along with a significant negative correlation with M0 macrophages, memory B cells and memory CD4 T cells; and IGFBP3 was significantly positively correlated with follicular helper T cells, neutrophils and M0 macrophages but negatively correlated with monocytes, M1 macrophages and M2 macrophages.

6. Key genes affect chemotherapeutic drug sensitivity

Our study further explored the relationship between the core genes and commonly used chemotherapy drugs through chemotherapy sensitivity data from the GDSC database. The regression relationship between each chemotherapeutic drug and each core gene is shown in Figure 8. The results showed that high expression of the CHST4 gene could significantly increase the estimated IC$_{50}$ of erlotinib, AMG.706, CCT007093 and MG.132. In contrast, a general decrease in the estimated IC$_{50}$ of all six drugs was observed along with high expression of CRHBP. In addition, the high expression of IGFBP3 significantly improved the estimated IC$_{50}$ of the five chemotherapeutic drugs except CCT007093. This finding suggested that core genes may dominate the prognosis of HCC patients by affecting sensitivity to chemotherapy drugs.
7. Mutations and signaling pathways involved in key genes

We selected the Pan Cancer dataset from the TCGA database and studied the mutations of core genes by the cBioPortal database. As a result, core gene mutations were present in the tumor cells of 32 HCC patients (9%). The specific mutations of each core gene are shown in Figure 9. The mutation rates of the 3 key genes were 2.9% (CHST4), 2.0% (CRHBP) and 5.0% (IGFBP3). Only high mRNA expression was found in the CHST4 gene, while missense mutations, amplification and high mRNA expression occurred in the IGFBP3 gene simultaneously. Next, we studied the specific signaling pathways involved in the three core genes to explore the potential molecular mechanisms by which key genes influence tumor progression. GVSA showed that high expression of the CHST4 and IGFBP3 genes was associated in various signaling pathways, such as bile acid metabolism, fatty acid metabolism and DNA repair, while high expression of the CRHBP gene is associated with the PI3K-AKT-mTOR signaling, oxidative phosphorylation and glycolysis pathways, which are closely related to the occurrence and progression of HCC (Fig. 10). The GeneMANIA results suggested a close association among CHST4, ICRHBP, and IGFBP3 and related genes in various aspects involved in tumor progression, including the regulation of the insulin-like growth factor receptor signaling pathway, positive regulation of activated T cell proliferation, growth factor binding, smooth muscle cell (SMC) migration and its regulation.

Discussion

The development of HCC is a complex biological process characterized by the accumulation of genetic variation and epigenetic modification abnormalities. DNA methylation is an important part of the epigenetic regulatory mechanism in the human genome and responsible for regulating the expression and silencing of target genes by adding or removing a methyl group to the promoter regions under the catalysis of DNA methyltransferase[10]. Previous studies have confirmed that abnormal DNA methylation is common in human tumor cells, including HCC[2, 11, 13, 14]. Oncogene demethylation and tumor suppressor gene methylation were found to be the main initiators and promoters in the progression of malignant tumors[15]. Currently, there is little information on the mapping of HCC methylation genes; therefore, our goal was to screen and validate the key differentially expressed methylation genes in HCC patients.

In this study, we extracted 78 groups of transcriptional data from the NCBI GEO public database. The WGCNA network and TOM analyses were performed, and the module with the highest correlation to tumor phenotypes was selected. GO and KEGG analyses suggested that the module was associated with several signaling pathways that have been verified to be closely related to the progression of HCC, including collagen-containing extracellular matrix, complement activation, lectin pathway, blood vessel development and cytokine receptor binding[16, 17]. A total of 373 differentially expressed genes were screened from the module, including 88 upregulated genes and 285 downregulated genes. After a combined analysis, we finally identified four overlapping genes: CHST4, CRHBP, IGFBP3 and PTH1R. A KM plot analysis showed that three of the four genes (except PTH1R) significantly influenced the predicted survival of HCC patients, which were collectively referred to as key genes. It is well known that
multiple immune-mediated pathways are involved in the development of tumor microenvironments. The tumor microenvironment contains various components that can inhibit the growth and proliferation of tumor cells, including immune effector molecules and cells, and it also has components that promote the growth and metastasis of tumor cells, such as tumor-associated macrophages and immunosuppressive molecules. Therefore, the relationship between the three key genes and tumor immune invasion was also analyzed. The results indicated that each gene had a strong correlation with immune cells, including M0 macrophages, neutrophils, and naïve B cells. A drug sensitivity analysis suggested that gene overexpression could significantly affect the lethality and sensitivity of commonly used chemotherapy drugs to tumor cells, thus further producing a chain reaction on the prognosis of HCC patients. Moreover, we found that high expression of the CRHBP gene can enhance the targeting sensitivity of chemotherapy drugs, which is contrary to the effect of CHST4 and IGFBP3. In addition, a total of 32 HCC patients (9%) were confirmed to have at least one key gene mutation. GSVA suggested that the three core genes participated in multiple specific signaling pathways involved in liver metabolism and tumor progression, such as bile acid metabolism, DNA repair, oxidative phosphorylation and glycolysis[18, 19]. Their involvement in T cell proliferation, growth factor binding and SMC migration has also been revealed by GeneMANIA. These results demonstrated the roles of CHST4, CRHBP and IGFBP3 as key differentially expressed methylation genes in HCC.

CHST4 is responsible for encoding N-acetylglucosamine-6-O-sulfotransferase 2 (GlcNAc6ST-2), which is composed of 386 amino acids. GlcNac6ST-2 can participate in the synthesis of L-selectin ligand and regulate the capture and rolling of white blood cells on vascular endothelial cells, thereby mediating the initial interaction between lymphocytes and high endothelial venules (HEVs). HEV is a special type of postcapillary venule characterized by lymphocyte homing. In normal human tissues, CHST4 expression can be detected in the intestinal tract, pancreas, and liver[20]. When inflammation is present, HEV-like blood vessels and GLCNA6ST-2 expression will appear at the source, such as the lungs of asthmatic sheep and the synovium of patients with rheumatoid arthritis[21, 22]. The expression and function of CHST4 in tumor cells have also been found and confirmed in previous studies. Among all tumors, cholangiocarcinoma has the highest frequency of CHST4 gene change[20]. In colorectal adenocarcinoma, the transcriptional expression of the CHST4 gene was significantly higher than that in normal tissue[20]. Accelerated leukocyte infiltration occurred in CHST4 knockout mice, thus confirming the barrier protective role of CHST4 in colonic mucosa[23]. The overexpression of GlcNAC6ST-2 was also found in early mucinous ovarian adenocarcinoma and clear cell carcinoma, and it was not observed in benign ovarian tumors. The mutation and high expression of CHST4 were also found in gastric cancer tissues and gliomas, suggesting a poor prognosis[24, 25]. In HCC, low mRNA expression and high protein expression of the CHST4 gene have been detected compared with normal tissues in previous studies[20]. Notably, higher CHST4 expression in HCC patients indicated poor overall survival (OS) and disease-free survival (DFS), which is consistent with our findings, whereas lower CHST4 expression in HBV-induced HCC patients indicated poor OS and DFS. These trends may be the result of low CHST4 expression promoting HBV expression and replication in liver cells in the absence of immune system defense[26]. Moreover, CHST4 overexpression has been proven to impede the migration ability of HCC cells[20].
CHST4 gene downregulation disrupts the immune response, leading to neutrophil infiltration and microvascular angiogenesis, which increases the risk of extrahepatic metastasis[20]. Similarly, the expression of CHST4 was significantly positively correlated with the degree of infiltration of immune cells, including B cells, CD4 cells, and macrophages, which was confirmed in our study but had no correlation with age, sex, TNM stage or serum AFP[20]. Racial differences in CHST4 gene expression have also been observed, which may explain the bias in gene expression from different samples. Although the association between CHST4 gene expression and HCC progression has been partially studied, there are few studies on CHST4 methylation in HCC.

Together with urocortin, urocortin 3 (UCN3), and corticotropin releasing hormone (CRH) receptors 1 and 2 (CRHR1 and CRHR2), CRHBP is classified as a member of the CRH system, and it is primarily responsible for regulating the expression balance between CRH and CRHR1/CRHR2 by binding to the CRH complex. The mechanism of CRHBP gene methylation in tumor cells has been explored in several previous studies, especially in urinary tumors. A significant reduction in mRNA levels caused by CRHBP gene methylation in renal cell carcinoma (RCC) has been demonstrated in several studies[27, 28], and it has also been observed in bladder and prostate cancers[28, 29]. Tezval et al. observed an approximately 80-fold increase in the CRHBP CpG island sequence leading to hypermethylation in over 50% of RCC tissue samples[28]. He et al. reported similar data in another article and showed that CRHBP mRNA levels in clear cell carcinoma were 33-fold reduced compared with those in normal tissues[27]. In addition, the use of methyltransferase inhibitors increased the mRNA reexpression level by approximately 100 times and reduced the aggressiveness of RCC, suggesting the reversibility of methylation and the possibility of tumor therapy[28]. At the same time, the degree of gene methylation was significantly positively correlated with distant metastasis and advanced disease. Lower mRNA expression of CRHBP predicted shorter survival and poorer prognosis. The mechanism may be that low expression of CRHBP leads to decreased activation and loss of function of CRHR2. CRHR2 mediates tumor cell apoptosis and inhibits vascular endothelial growth factor (VEGF) expression by competitively binding CRH with CRHR1, thus leading to reduced angiogenesis and a lower risk of tumor metastasis and dissemination[30]. To some extent, the decrease in CRHR2 permeability in tumor microvascular endothelial cells and the increase in CRHBP overexpression in RCC apoptosis also confirmed this mechanism. Furthermore, indicators of low CRHBP expression have been found in ovarian and breast cancer cells. Other members of the CRH system have also been reported to be involved in tumor progression. For example, CRH was detected to promote the growth and proliferation of colon cancer cells by activating CRHR1 and upregulating VEGF expression[31]. Tabrizi et al. also reported that RCC samples presented low levels of UCN3 mRNA expression, which inhibits angiogenesis by activating CRHR2[30]. In the field of HCC, there are few reports on the methylation and mechanism of the CRHBP gene. Xia et al. confirmed that the protein expression of the CRHBP gene was significantly reduced in HCC tissues compared with adjacent normal tissues[32]. Additionally, low expression of the CRHBP gene was significantly associated with the Edmonson grade, high serum AFP and tumor size. Notably, there is a close association between hypermethylation of the CRHBP gene and long-term stimulation of HBV[32], although the mechanism is unclear. Finally, CRHBP is considered a tumor suppressor gene because of its ability to activate p53-mediated tumor apoptosis and
promote tumor immunity by the NF-kB signaling pathway[33]. A breakthrough in the methylation of the CRHBP gene will be a new focus for the treatment of cancer in the future.

The insulin-like growth family consists of insulin-like growth factors (IGF1 and IGF2), IGF receptors (IGFRs) and six IGFBPs (IGFBP1-6). IGFBP3 is the most widely and deeply studied IGFBP species in vivo. It is produced mainly by Kupffer cells in the liver but is also found in other local tissues, including the breast, lung, and digestive tract. It has been reported that IGFBP3 can reduce the pro-proliferation and anti-apoptotic effects of IGF/IGFR in cells and control the bioavailability of IGFs by binding to IGF with a high affinity[34]. Moreover, the inhibitory effect of IGFBP3 on angiogenesis was also demonstrated. When any link in the axis is damaged, the mechanisms involved in tumor development, chronic inflammation and diabetes are activated[35]. The methylation of IBFBP3 in human tumor cells has been widely reported over several years. An association between high expression of IGFBP3 and poor prognosis in breast cancer patients was found by some researchers[36]. Zielinska et al. observed that impaired glucose-regulated protein 78 (GRP78) synthesis in breast cancer resulted in the inability of IGFBP3 to enter cells, where IGFBP3 plays a pro-cell proliferative role by binding to cell surface factors[37]. This finding is consistent with the worst OS in GRP78-negative/IGFBP3-positive patients. The hypermethylated expression of IGFBP3 in gastric cancer and colorectal cancer has been proposed and verified by several reports[38]. In addition, Adachi et al. pointed out that IGFBP3 overexpression combined with adjuvant chemotherapy could inhibit tumor proliferation mediated by the NF-kB pathway, which would be beneficial for the prognosis of patients with gastric cancer[38]. The combined diagnosis of CA19-9, IGFBP2 and IGFBP3 has also been advocated to improve the detection rate of early pancreatic cancer[39]. High levels of IGFBP3 methylation in lung, ovarian, endometrial, and prostate cancers were mentioned to varying degrees[35, 40]. In contrast, overexpression of IGFBP3 in nasopharyngeal carcinoma was reported to have precisely the opposite effect, thus promoting proliferation, invasion, and metastasis of cancer cells. Similar to most other tumors, IGFBP3 expression was observed to be significantly lower in HCC than in adjacent nontumor tissues[34]. This reduction is partially associated with insufficient secretion from damaged liver cells, degradation of proteases, and hypermethylation of genes. The CpG promoter region of IGFBP3 has been proven to have binding sites for the tumor suppressor gene p53[34]. Methylation of this region inhibits p53 gene-mediated apoptosis. At the same time, reduced IGFBP3 secretion will encourage free IGF to bind to IGFR, which plays a role in promoting tumor cell proliferation. However, Bai et al. reported the opposite conclusion: high IGFBP3 expression was associated with poor prognosis[41]. In some studies, activation of IGFBP3 has been found to reduce cell proliferation and promote apoptosis[34], which may provide a direction for new targeted therapies in the future.

In conclusion, this study proposed a method of screening for key differentially expressed methylation genes in HCC patients. By discussing and verifying the preselected genes, we finally determined the core roles of CHST4, CRHBP and IGFBP3 in HCC gene methylation. Our work helps to clarify the role of gene methylation in the progression of HCC and shows that it is involved immune infiltration, molecular mechanisms and drug sensitivity. These results may lead to breakthroughs and possibilities for early accurate diagnosis and effective targeted therapy for HCC in the future.
Declarations

Acknowledgments
Not applicable.

Authors’ contributions
Conception and design: YW, HH and SJ
Acquisition of data: YW, FQ and HH
Analysis and interpretation of data: YW, XW and HH
Writing, review, and/or revision of the manuscript: HH, FQ and SJ
Study supervision: HH and SJ
All authors read and approved the final manuscript.

Funding
This study was supported by the National Natural Science Foundation of China (NSFC No. 81903964 to HH)

Availability of data and materials
The dataset used and/or analyzed during the current study is available from the corresponding author on reasonable request.

Ethics approval and consent to participate
This study was approved by the Ethics Committee of the Dalian Medical University of the 1st Affiliated Hospital.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

References
1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A: Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: a cancer journal for clinicians 2018, 68(6):394-424.

2. Umeda S, Kanda M, Kodera Y: Recent advances in molecular biomarkers for patients with hepatocellular carcinoma. Expert Rev Mol Diagn 2019, 19(8):725-738.

3. Ghouri YA, Mian I, Rowe JH: Review of hepatocellular carcinoma: Epidemiology, etiology, and carcinogenesis. J Carcinog 2017, 16:1.

4. Chen S, Cao Q, Wen W, Wang H: Targeted therapy for hepatocellular carcinoma: Challenges and opportunities. Cancer Lett 2019, 460:1-9.

5. Singal AG, Pillai A, Tiro J: Early detection, curative treatment, and survival rates for hepatocellular carcinoma surveillance in patients with cirrhosis: a meta-analysis. PLoS Med 2014, 11(4):e1001624.

6. Roayaie S, Obeidat K, Sposito C, Mariani L, Bhoori S, Pellegrinelli A, Labow D, Llovet JM, Schwartz M, Mazzaferro V: Resection of hepatocellular cancer \( \leq 2 \) cm: results from two Western centers. Hepatology 2013, 57(4):1426-1435.

7. Li X, Xu W, Kang W, Wong SH, Wang M, Zhou Y, Fang X, Zhang X, Yang H, Wong CH et al: Genomic analysis of liver cancer unveils novel driver genes and distinct prognostic features. Theranostics 2018, 8(6):1740-1751.

8. Wilting RH, Dannenberg JH: Epigenetic mechanisms in tumorigenesis, tumor cell heterogeneity and drug resistance. Drug Resist Updat 2012, 15(1-2):21-38.

9. Baylin SB, Esteller M, Rountree MR, Bachman KE, Schuebel K, Herman JG: Aberrant patterns of DNA methylation, chromatin formation and gene expression in cancer. Hum Mol Genet 2001, 10(7):687-692.

10. Long J, Chen P, Lin J, Bai Y, Yang X, Bian J, Lin Y, Wang D, Yang X, Zheng Y et al: DNA methylation-driven genes for constructing diagnostic, prognostic, and recurrence models for hepatocellular carcinoma. Theranostics 2019, 9(24):7251-7267.

11. Lu CY, Chen SY, Peng HL, Kan PY, Chang WC, Yen CJ: Cell-free methylation markers with diagnostic and prognostic potential in hepatocellular carcinoma. Oncotarget 2017, 8(4):6406-6418.

12. Kondo Y, Kanai Y, Sakamoto M, Mizokami M, Ueda R, Hirohashi S: Genetic instability and aberrant DNA methylation in chronic hepatitis and cirrhosis--A comprehensive study of loss of heterozygosity and microsatellite instability at 39 loci and DNA hypermethylation on 8 CpG islands in microdissected specimens from patients with hepatocellular carcinoma. Hepatology 2000, 32(5):970-979.

13. Irizarry RA, Ladd-Acosta C, Wen B, Wu Z, Montano C, Onyango P, Cui H, Gabo K, Rongione M, Webster M et al: The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. Nat Genet 2009, 41(2):178-186.

14. 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. Clin Epigenetics 2018, 10:8.
15. Kunej T, Godnic I, Ferdin J, Horvat S, Dovc P, Calin GA: Epigenetic regulation of microRNAs in cancer: an integrated review of literature. *Mutar Res* 2011, **717**(1-2):77-84.

16. Maehara N, Arai S, Mori M, Iwamura Y, Kurokawa J, Kai T, Kusunoki S, Taniguchi K, Ikeda K, Ohara O et al: Circulating AIM prevents hepatocellular carcinoma through complement activation. *Cell Rep* 2014, **9**(1):61-74.

17. Renne SL, Woo HY, Allegra S, Rudini N, Yano H, Donadon M, Vigano L, Akiba J, Lee HS, Rhee H et al: Vessels Encapsulating Tumor Clusters (VETC) Is a Powerful Predictor of Aggressive Hepatocellular Carcinoma. *Hepatology* 2020, **71**(1):183-195.

18. Oshi M, Kim TH, Tokumaru Y, Yan L, Matsuyama R, Endo I, Cherkassky L, Takabe K: Enhanced DNA Repair Pathway Is Associated with Cell Proliferation and Worse Survival in Hepatocellular Carcinoma (HCC). *Cancers (Basel)* 2021, **13**(2).

19. Feng J, Li J, Wu L, Yu Q, Ji J, Wu J, Dai W, Guo C: Emerging roles and the regulation of aerobic glycolysis in hepatocellular carcinoma. *J Exp Clin Cancer Res* 2020, **39**(1):126.

20. Zhang L, Fan Y, Wang X, Yang M, Wu X, Huang W, Lan J, Liao L, Huang W, Yuan L et al: Carbohydrate Sulfotransferase 4 Inhibits the Progression of Hepatitis B Virus-Related Hepatocellular Carcinoma and Is a Potential Prognostic Marker in Several Tumors. *Front Oncol* 2020, **10**:554331.

21. Rosen SD, Tsay D, Singer MS, Hemmerich S, Abraham WM: Therapeutic targeting of endothelial ligands for L-selectin (PNAd) in a sheep model of asthma. *Am J Pathol* 2005, **166**(3):935-944.

22. Pablos JL, Santiago B, Tsay D, Singer MS, Palao G, Galindo M, Rosen SD: A HEV-restricted sulfotransferase is expressed in rheumatoid arthritis synovium and is induced by lymphotoxin-alpha/beta and TNF-alpha in cultured endothelial cells. *BMC Immunol* 2005, **6**:6.

23. Tobisawa Y, Imai Y, Fukuda M, Kawashima H: Sulfation of colonic mucins by N-acetylglucosamine 6-O-sulfotransferase-2 and its protective function in experimental colitis in mice. *J Biol Chem* 2010, **285**(9):6750-6760.

24. Okayama H, Kumamoto K, Saitou K, Hayase S, Kofunato Y, Sato Y, Miyamoto K, Nakamura I, Ohki S, Koyama Y et al: Ectopic expression of MECA-79 as a novel prognostic indicator in gastric cancer. *Cancer Sci* 2011, **102**(5):1088-1094.

25. Subbarayan K, Seliger B: Tumor-dependent Effects of Proteoglycans and Various Glycosaminoglycan Synthesizing Enzymes and Sulfotransferases on Patients' Outcome. *Curr Cancer Drug Targets* 2019, **19**(3):210-221.

26. Amaddeo G, Cao Q, Ladeiro Y, Imbeaud S, Nault JC, Jaoui D, Gaston Mathe Y, Laurent C, Laurent A, Bioulac-Sage P et al: Integration of tumour and viral genomic characterizations in HBV-related hepatocellular carcinomas. *Gut* 2015, **64**(5):820-829.

27. Tezval H, Atschekzei F, Peters I, Waalkes S, Hennenlotter J, Stenzl A, Becker JU, Merseburger AS, Kuczyk MA, Serth J: Reduced mRNA expression level of corticotropin-releasing hormone-binding protein is associated with aggressive human kidney cancer. *BMC Cancer* 2013, **13**:199.

28. Tezval H, Dubrowinskaja N, Peters I, Reese C, Serth K, Atschekzei F, Hennenlotter J, Stenzl A, Kuczyk MA, Serth J: Tumor Specific Epigenetic Silencing of Corticotropin Releasing Hormone-Binding
Protein in Renal Cell Carcinoma: Association of Hypermethylation and Metastasis. PLoS One 2016, 11(10):e0163873.

29. Jin L, Zhang Q, Guo R, Wang L, Wang J, Wan R, Zhang R, Xu Y, Li S: Different effects of corticotropin-releasing factor and urocortin 2 on apoptosis of prostate cancer cells in vitro. J Mol Endocrinol 2011, 47(2):219-227.

30. Faraj Tabrizi P, Mohebbi Tafrechi A, Peters I, Atschekzei F, Kuczyk MA, Serth J, Tezval H: Cancer-Specific Loss of Urocortin 3 in Human Renal Cancer. Adv Ther 2020, 37(1):288-299.

31. Fang X, Hong Y, Dai L, Qian Y, Zhu C, Wu B, Li S: CRH promotes human colon cancer cell proliferation via IL-6/JAK2/STAT3 signaling pathway and VEGF-induced tumor angiogenesis. Mol Carcinog 2017, 56(11):2434-2445.

32. Xia HB, Wang HJ, Fu LQ, Wang SB, Li L, Ru GQ, He XL, Tong XM, Mou XZ, Huang DS: Decreased CRHBP expression is predictive of poor prognosis in patients with hepatocellular carcinoma. Oncol Lett 2018, 16(3):3681-3689.

33. Yang K, Xiao Y, Xu T, Yu W, Ruan Y, Luo P, Cheng F: Integrative analysis reveals CRHBP inhibits renal cell carcinoma progression by regulating inflammation and apoptosis. Cancer Gene Ther 2020, 27(7-8):607-618.

34. Han JJ, Xue DW, Han QR, Li XH, Xie L, Li S, Wu HY, Song B: Induction of apoptosis by IGFBP3 overexpression in hepatocellular carcinoma cells. Asian Pac J Cancer Prev 2014, 15(23):10085-10089.

35. Cai Q, Dozmorov M, Oh Y: IGFBP-3/IGFBP-3 Receptor System as an Anti-Tumor and Anti-Metastatic Signaling in Cancer. Cells 2020, 9(5).

36. Monson KR, Goldberg M, Wu HC, Santella RM, Chung WK, Terry MB: Circulating growth factor concentrations and breast cancer risk: a nested case-control study of IGF-1, IGFBP-3, and breast cancer in a family-based cohort. Breast Cancer Res 2020, 22(1):109.

37. Zielinska HA, Daly CS, Alghamdi A, Bahl A, Sohail M, White P, Dean SR, Holly JMP, Perks CM: Interaction between GRP78 and IGFBP-3 Affects Tumourigenesis and Prognosis in Breast Cancer Patients. Cancers (Basel) 2020, 12(12).

38. Adachi Y, Nojima M, Mori M, Kubo T, Yamano HO, Lin Y, Wakai K, Tamakoshi A, for JS: Circulating insulin-like growth factor binding protein-3 and risk of gastrointestinal malignant tumors. J Gastroenterol Hepatol 2019, 34(12):2104-2111.

39. Yoneyama T, Ohtsuki S, Honda K, Kobayashi M, Iwasaki M, Uchida Y, Okusaka T, Nakamori S, Shimahara M, Ueno T et al: Identification of IGFBP2 and IGFBP3 As Compensatory Biomarkers for CA19-9 in Early-Stage Pancreatic Cancer Using a Combination of Antibody-Based and LC-MS/MS-Based Proteomics. PLoS One 2016, 11(8):e0161009.

40. Wiley A, Katsaros D, Fracchioli S, Yu H: Methylation of the insulin-like growth factor binding protein-3 gene and prognosis of epithelial ovarian cancer. Int J Gynecol Cancer 2006, 16(1):210-218.
Weighted Gene Co-expression Network Analysis and Differential Gene Expression Analysis. *Front Genet* 2020, **11**:615308.

**Figures**

**Figure 1**

Identification of HCC-related genes based on a WGCNA coexpression network analysis. A: Cluster tree diagram and trait index of samples. B: Scale-free exponential fitting analysis of soft thresholds (1-20). C: Joint analysis of module and sample features. D: Correlation analysis of modules and sample characteristics (magenta).
Figure 2

2.1. Pathway analysis diagram of the magenta module. A: GO and KEGG analysis of module (magenta) enrichment. B: Interactions between GO and KEGG enrichment pathways. PPI diagram of the interaction between genes in the magenta module.
Figure 3

Heat map of methylation differences between the normal group and HCC group.
Figure 4

Volcano map, with upregulated genes shown in red and downregulated genes shown in green.
Figure 5

Survival curve, with high expression in red and low expression in blue. Kaplan-Meier survival curves of 4 hub genes associated with overall survival in HCC. CHST4 (A), CRHBP (B), IGFBP3 (C) and PTH1R (D). A P-value <0.05 was considered statistically significant.

Figure 6

A: Correlation between CHST4 and infiltrating immune cells. B: Correlation between CRHBP and infiltrating immune cells. C: Correlation between IGFBP3 and infiltrating immune cells. The size of the points represents the strength of the association between genes and immune cells. The larger the point, the stronger the correlation. The smaller the point, the weaker the correlation. A P-value <0.05 was considered statistically significant.

Figure 7

Sensitivity analysis of key genes to common antitumor drugs. Relationship between key genes CHST4 (A), CRHBP (B), and IGFBP3 (C) and the common antitumor drug IC50.
Figure 8

Panorama of mutations in key genes.

Figure 9

GSVA pathway analysis of key genes. Specific signaling pathways associated with key genes CHST4 (A), CRHBP (B) and IGFBP3 (C).

Figure 10
Correlation between the core genes analyzed by the GeneMANIA database.

**Supplementary Files**

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

- Additionalfile1.xls
- Additionalfile2.xls
- Additionalfile3.xls
- Additionalfile4.xlsx
- Additionalfile5.xlsx
- Additionalfile6.xls
- Additionalfile7.xls
- Additionalfile8.xls