Development of a Novel CTCs/CTMs Related Prognostic Signature for Hepatocellular Carcinoma

Xiangyu Li

Huazhong University of Science and Technology

Yuanxin Shi

Huazhong University of Science and Technology

Kai Zhao

Huazhong University of Science and Technology

Yun Lu

Huazhong University of Science and Technology

Peng Qiu

Huazhong University of Science and Technology

Zhengdong Deng

Huazhong University of Science and Technology

Wei Yao

Huazhong University of Science and Technology

Jianming Wang (✉ wjm18jgm@tjh.tjmu.edu.cn)

Huazhong University of Science and Technology <https://orcid.org/0000-0003-3143-4568>

Primary research

Keywords: Circulating tumor cells, CDCA8, Hepatocellular carcinoma, Prognostic signature, TCGA

Posted Date: September 9th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-842218/v1>

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

[Read Full License](#)

Abstract

Background: Hepatocellular carcinoma (HCC) is a common malignancy and the third most deadly cancer worldwide. Previous studies have demonstrated that circulating tumor cells are involved in the occurrence and development of various cancers, including HCC. For this study, we aimed to comprehensively analyze data related to HCC to develop a prognostic model based on CTCs/CTMs related genes (CRGs).

Methods: Data were obtained from TCGA, ICGC, and GEO. Firstly, we screened the differentially expressed CRGs and constructed a signature in the TCGA cohort by Lasso-penalized Cox regression analysis and the multivariate cox regression analysis. Then, the prognostic model was validated in the ICGC dataset and GES14520 dataset with survival analysis and receiver operating characteristic analysis. Moreover, we investigated the clinical significance of prognostic signature, including the correlations with clinical characteristics, immune cell infiltration, and immune checkpoints. Next, we also established the nomogram and to better predict the prognosis of patients. We identified five potential small molecule drugs by Connectivity Map (CMap) and validated them using the Comparative Toxicogenomics Database (CTD). Besides, we further explored the biological role of CDCA8 in hepatocellular carcinoma cells.

Results: The prognostic signature exhibited good predictive power and clinical application. Besides, the signature was associated with immune checkpoints (PD-1, PD-L1, and CTLA4), implying that high-risk patients might benefit more from immunotherapy. Additionally, In vitro experiments showed that CDCA8 could promote the proliferation, invasion, and metastasis of hepatocellular carcinoma cells, and silencing CDCA8 could lead to cell cycle arrest and increased apoptosis.

Conclusion: We developed a multi-gene classifier that can effectively help the HCC patients benefit from target therapy or immune therapy. And CDCA8 may be the next therapeutic target for hepatocellular carcinoma.

Introduction

Primary liver cancer is the sixth most prevalent and the third most deadly cancer worldwide, with hepatocellular carcinoma (HCC) being the most common type of liver cancer, accounting for about 90% of cases. (1–3). Cirrhosis of any etiology is the most important factor for HCC. The main risk factors for HCC include chronic alcohol consumption, diabetes or obesity-related non-alcoholic steatohepatitis (NASH), and HBV or HCV infection. Other less common risk factors include primary biliary cirrhosis, hemochromatosis, and alpha1-antitrypsin deficiency (4, 5). Currently, the main treatment options for patients with liver cancer remain surgical liver resection and liver transplantation. However, for patients with advanced, recurrent HCC or those who have lost the chance of surgery, the prognosis is still not promising. Despite some advances in the detection, diagnosis, and treatment of HCC in recent years, its prognosis remains poor due to the high rates of recurrence, vascular invasion, or metastasis (6). Therefore, it is urgent to explore both effective and representative biomarkers and new predictive tools.

Circulating tumor cells (CTCs) are those cells that depart from original or metastatic tumors and shed them into the bloodstream, while rare in healthy individuals (7–9). As an important member of liquid biopsy technology, CTCs play an essential role in cancer diagnosis and treatment, carrying heterogeneous information about the primary tumor and serving as an effective biomarker and modeling tool. Many studies have shown that CTCs are closely associated with cancer metastasis. Therefore, the isolation and identification of CTCs with non-invasive biopsy can be widely applied for early diagnosis, real-time efficacy monitoring, and prognosis assessment of tumors (10–14). In recent years, an increasingly popular view shows that the higher level of CTCs, the worse prognosis of tumor patients. In the CTC-positive cohort with HCC, survival time was significantly shorter and was also correlated with poor clinical features (15, 16). Similarly, sun et al. found that the risk of tumor recurrence was increasing when HCC patients had a preoperative CTC (7.5ml) of ≥ 2 , especially at AFP levels ≤ 400 ng / mL (17). With advances in technology, genomic, transcriptomic, and proteomic analysis of CTCs at the single-cell level, and refinement of CTC in vitro models, we can better understand the critical role of CTCs in cancer (18, 19). Nevertheless, our understanding of the biological functions of CTC in tumors at the molecular level is still not sufficient. Therefore, this research aimed to identify the CTCs/CTMs related genes (CRGs) and explore their clinical significance in HCC.

In this study, we comprehensively analyzed the transcriptomic data and clinical information of HCC in TCGA and ICGC databases. By combining the mRNA data associated with HCC in ctcRbase, we identified 258 CTCs/CTMs related genes (CRGs). Subsequently, we utilized the least absolute shrinkage and selection operator (LASSO) analysis and Cox regression analysis to establish a five CRGs prognostic signature. In addition, the prognostic model was validated in GSE14520 dataset and the correlation was further investigated between the signature genes and immune cells, immune checkpoints. Through the analysis and validation of CDCA8, we identified it as a potential target for HCC. And we conducted some experiments to investigate the biological behaviors of CDCA8 in hepatocellular carcinoma such as proliferation, invasion, metastasis. Meanwhile, flow cytometric assays were used to assess whether CDCA8 affects the cell cycle and apoptosis in HCC.

Materials And Methods

Acquisition of gene expression and clinical data

The training gene expression of transcriptome profiles and corresponding clinical trait data of LIHC patients were downloaded from The Cancer Genome Atlas database (TCGA, <https://portal.gdc.cancer.gov/>); while the validation dataset of mRNA expression and clinical trait data (the Liver Cancer LIRI-JP) were downloaded from the International Cancer Genome Consortium database (ICGC, <https://icgc.org/>). There are 12518 CTCs/CTMs genes were downloaded from the ctcRbase(<http://www.origin-gene.cn/database/ctcRbase/>) (7).

Identification of differentially expressed genes (DEGs)

To obtain the differentially expressed genes (DEGs) associated with CTCs/CTMs, we utilized the “limma” R package for processing the RNA-seq data and the cutoff point was set as false discovery rate (FDR) < 0.05 and $|\log_2(\text{fold change})| > 2$. Then we applied the Venn diagram (<http://bioinformatics.psb.ugent.be/webtools/Venn/>) to illustrate the intersection among TCGA-DEGs, ICGC-DEGs, and CTCs/CTMs genes. And these intersected genes we called CTCs/CTMs related genes (CRGs).

GO Enrichment Analysis, KEGG Pathway, GSEA, and Comprehensive Analysis of PPI Network

Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis were applied to explore functional annotation and pathway analysis of the CRGs with the “clusterProfiler” R package (20). The cut-offs were set as: $p < 0.05$, and $\text{FDR} < 0.05$. Gene Set Enrichment Analysis (GSEA) was employed to explore the common biological pathways (21) and `cp.kegg.v7.1.symbols.gmt` served as a reference gene set, with a threshold of $p < 0.05$, to screen for dominant enrichment pathways in different risk groups. Besides, to analyze the interaction between the CRGs, we employed the STRING database (<https://string-db.org/>) to set a protein-protein interaction (PPI) network. And an interaction score > 0.7 was set as significant. Moreover, we used the Cytoscape software to visually converting the PPI network. Plug-in Molecular Complex Detection (MCODE) was utilized in Cytoscape to select the central modules of PPI network with degree cut-off = 2, node score cut-off = 0.2, k-core = 2, and max. Depth = 100 as the criterion (22). In addition, we adopted a novel Cytoscape plugin, cytoHubba, for ranking the nodes in the network by their network functionality (23).

Construction and Validation of the prognostic gene signature

Univariate Cox regression analysis was performed to identify the prognostic CRGs and the CRGs associated with survival time ($p < 0.01$) were considered to be statistically significant. Then, the least absolute shrinkage and selection operator (LASSO) penalized Cox regression analysis was applied to further filter prognostic CRGs for overall survival in patients with HCC (24). In addition, the risk signature was developed based on multivariate Cox proportional hazards regression analysis. As previously reported, prognostic gene signatures were constructed based on linear combinations of regression coefficients derived by multiplying the LASSO Cox regression model coefficients by their mRNA expression levels (25). The risk score = $\sum \beta \text{mRNA}_n \times \text{expression level of mRNAs}$. ROC curves and Kaplan-Meier curves were used to evaluate the predictive value of the TCGA cohort in the prognostic model. Meanwhile, the ICGC database was used as external validation data to test the predictive capability of the model.

Establishing the predictive nomogram

Researchers frequently construct nomograms to predict the prognosis of cancer (26). We established a simplified nomogram of HCC by employing the package “rms” in R. Besides, the calibration curves of the nomogram were performed to estimate the predictive capability of this model.

Bioinformatics analysis of the prognostic signature

We explored the correlations between different risk groups and clinical characteristics using chi-square tests, and the results were represented as heat maps. Moreover, we utilized the Spearman analysis to research the correlation between risk signature and immune checkpoint genes such as cytotoxic T-lymphocyte protein 4 (CTLA4), programmed cell death protein 1 (PDCD1), programmed cell death 1 ligand 1 (CD274). We also studied the association between the signature genes and immune cell infiltration in TIMER (<https://cistrome.shinyapps.io/timer/>).

Screening potential therapeutic small molecule drugs for HCC

To better identify small molecule drugs relevant to HCC treatment, we uploaded the differentially expressed CRGs to cMap (<http://portals.broadinstitute.org/cmap/>) for further analysis (27). Besides, we employed the Comparative Toxicogenomics Database (CTD; <http://ctdbase.org/>) to validate the potential drugs (28). Positive scores indicate the promotive effect of the drug and negative scores indicate the inhibitory effect of the drug. While, the set threshold was $p < 0.001$, $n \geq 4$ and $|\text{mean}| > 0.5$.

Patients and tissue specimens

HCC tissues and nonneoplastic tissues were collected from the Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology (Wuhan, China) between 2018 and 2020. The study methods met the criteria set on the Declaration of Helsinki and ethical approval was secured from the Ethics Committee of Tongji Hospital, Huazhong University of Science and Technology. Informed consent was also obtained from all patients.

Cell culture and transfection

HCC cell lines were acquired from the Hepatic Surgery Center, Tongji Hospital. All cells were grown in high-glucose Dulbecco's modified Eagle's medium (Gibco, Carlsbad, CA, USA) supplemented with 10% fetal bovine serum (Gibco, USA) at 37°C in 5% CO₂. CDCA8 siRNA and siRNA-NC were provided by DesignGene Biotechnology (Shanghai, China). All the reagents were transfected into HCC cells via Lipofectamine 2000 reagent (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's instructions.

Western blot analysis

Cells were lysed on ice with RIPA buffer containing protease inhibitor cocktail (MedChemExpress, USA) for 20 min. Cell lysates were quantified using a BCA protein assay kit (Beyotime, Jiangsu, China), and protein was separated by SDS-PAGE (Boster Biological Technology, Wuhan, China) and transferred to PVDF membranes (Millipore, MA, USA). The membranes were blocked with 5% fat-free milk and incubated with the respective primary and secondary antibodies. In the end, the bands were detected and analyzed with ChemiDoc™ XRS + by Image Lab™ software (Bio-Rad, USA). The antibodies of CDCA8 (#A15463, 1:1000) and GAPDH (#AC001, 1:3000) purchased from ABclonal (Wuhan, China).

Quantitative real-time PCR

Total RNA was extracted from cells using TRIzol reagent (VAZYME, Nanjing, China) and synthesized into cDNA using PrimeScript™ RT Master Mix (Takara), as recommended by the supplier's instruction. Quantitative real-time PCR (qRT-PCR) was carried with CFX96 Real-Time PCR System (Bio-Rad) using SYBR Green PCR kit (Thermo Fisher Scientific) according to the standard protocol. Sequences of primer pairs were as follows: (CDCA8 forward, 5'- TGTCTGTTCCACCCTCCCATCCC-3'; CDCA8 reverse, 5'- GTCAGTCTGCTGCCAACTGTCCTG-3'; GAPDH forward, 5'-GGGAGCCAAAAGGGTCATCATCTC-3'; GAPDH reverse, 5'-CCATGCCAGTGAGCTTCCCGTTC-3') synthesized by Sangon Biotech (Shanghai, China). The relative expression calculated using the formula $2^{-\Delta\Delta Ct}$.

Cell proliferation assay

Cell Counting Kit 8 assay (ABclonal, Wuhan, China) was utilized to assess the cell proliferation ability. 3000 cells were placed in each well of the 96-well plate and incubated overnight at 37°C to allow adhesion. At the indicated times 100µl of 10% CCK8 solution was added to each well and cultured in a cell incubator for 2h and absorbance was measured at 450 nm by a microplate reader (Thermo Scientific, China). All experiments were repeated 3 times.

Transwell migration and invasion assays

According to the manufacturer's protocol, cell migration assays were detected with transwell inserts (24-well inserts, 8-µm pore size; Corning Inc., Corning, NY, USA). For the Matrigel invasion assay, the transwell inserts were pre-coated with Matrigel (BD Biosciences, NJ, USA) and were serum-starved for 4h in a serum-free DMEM mixture. In general, for invasion and migration assays, 5×10^4 cells in 100µl serum-free medium were placed into the upper chamber of the transwell unit and the lower chamber was fixed with DMEM containing 10% FBS. After incubated for 24h, the cells were removed from the upper membrane surface, while the lower cells were fixed with 4% paraformaldehyde for 15 min and stained with 0.1% crystal violet for 20 min. Subsequently, the cells were then observed under a light microscope and counted in five random fields. All experiments were repeated 3 times.

Colony formation assay

Cells were inoculated into 6-well plates at a concentration of 1000 cells per well. The HCC cells were cultured in the medium at 37°C with 5% CO₂ with fresh medium and the cells were allowed to grow for 14 days, then dyed with crystalline violet (0.5% wt/vol) and photographed to quantify the colonies formed. All experiments were repeated 3 times.

Wound-healing assay

The HCC cells were cultured in 6-well plates and grown to 90% confluence. The cell layers were lightly scratched with a sterile 200ul micropipette tip and cultured in 10% FBS medium for 24h. Wound closures were imaged on a microscope. All experiments were repeated 3 times.

Cell cycle analysis and Cell apoptosis assay

Cell cycle analysis kit (Beyotime Biotechnology, China) was applied to analyze the cell cycle distribution. Cells were transferred and then cultured in the medium for 24h, fixed overnight in 75% ice ethanol. Then, the cells were incubated in RNase A and PI for 30 minutes and analyzed by flow cytometry. The Annexin V-FITC/PI Apoptosis Detection Kit (Beyotime Biotechnology, China) was used to detect apoptosis in HCC cells according to the manufacturer's instructions. Briefly, cells were transferred, collected, and resuspended in 1x binding buffer containing 2.5ul Annexin V-FITC and 5ul propidium iodide (PI) in the dark for approximately 15 min. Finally, the cell suspensions were detected by flow cytometry. All experiments were repeated 3 times.

Statistical analysis

R language (Version 3.6.2) was performed to conduct public data statistical analysis and plotting. For evaluating the prognostic value, we employed the Kaplan-Meier method and log-rank test. The Student's t-test (two-tailed) was applied to compare the differences between groups. Data analysis was performed using GraphPad Prism 7.0 (GraphPad, San Diego, CA, USA) or SPSS v23.0 (IBM Corp., Armonk, NY, USA). All statistical tests were two-sided. A $P < 0.05$ was considered statistically significant.

Results

Differentially expressed CRGs

We downloaded the LIHC gene expression profiles from the TCGA portal and ICGC portal and screened out 1622 and 628 differentially expressed genes (DEGs) respectively using "limma" R package. The DEGs of the TCGA database and ICGC database volcano plots were shown in Fig. 1A and Fig. 1B. A Venn diagram was used to identify the differentially expressed CRGs (Fig. 1C). By taking the intersection of the DEGs of TCGA and ICGC database with the 12518 CTCs/CTMs related genes (CRGs) from ctcRbase, we evaluated a total of 258 differentially expressed CRGs.

PPI network construction

To investigate the interrelationship of the differentially expressed CRGs and acquire hub genes, we performed PPI, module analysis and obtained the co-expression networks. Firstly, the 258 differentially expressed CRGs were analyzed in STRING, and the minimum required interaction score was set to 0.7, which indicates a strong interaction between CRGs. Then, the STRING interactions were analyzed in Cytoscape and the co-expression network was displayed in Fig. 2A, which contained 155 nodes and 2731deg. Besides, we identified modules with genes > 50 using the MCODE plugin and characterized 10 hub genes (TOP2A, CCNB2, CDCA8, BIRC5, AURKB, CCNB1, BUB1, BUB1B, KIF20A, TTK) in that module using the cytoHubba plugin (Fig. 2B). Meanwhile, this module includes 57 nodes and 1497 edges. These potential hub genes may play a critical role in the progress of hepatocellular carcinoma.

GO and KEGG enrichment analyses

To explore the biological category and biological processes of differentially expressed CRGs, GO and KEGG enrichment analyses were performed using R software, and the enrichment results for each term were shown in bubble charts (Fig. 2C-D). GO enrichment analysis showed that the differentially expressed CRGs were principally enriched in biological process (BP) terms nuclear division, organelle fission, chromosome segregation, mitotic nuclear division, etc. Besides, the cellular component (CC) terms mainly included chromosomal region, centromeric region, spindle, kinetochore, etc. Moreover, the molecular function (MF) terms showed that the CRGs were associated with tubulin binding, microtubule-binding, ATPase activity, moto activity, etc. (Fig. 2C). Regarding the KEGG analysis, the main terms were showed in Fig. 2D, and these differentially expressed CRGs were particularly enriched in cell cycle, microRNAs in cancer, p53 signaling pathway, and cellular senescence.

Construction of prognostic model and validation of the model in ICGC cohort

A total of 88 CRGs were found to be strongly associated with survival in HCC patients by univariate Cox regression analysis ($p < 0.01$), of which all of the 88 CRGs were prognostic risk factors (Fig. 3A). Then, the 88 CRGs were regression penalized using LASSO Cox regression to exclude relatively insignificant parameters (Fig. 3B-C). Besides, the stepwise multivariate Cox regression was employed to construct a predictive signature for HCC patients in the TCGA cohort (Fig. 3D). The five genes characterized were cell division cycle associated 8 (CDCA8), tumor protein p53 inducible protein 3 (TP53I3), hepatitis A virus cellular receptor 1 (HAVCR1), MYCN proto-oncogene (MYCN), thioredoxin reductase 1 (TXNRD1). The risk score = $(0.0826 \times \text{expression level of CDCA8}) + (0.0112 \times \text{expression level of TP53I3}) + (0.0824 \times \text{expression level of MYCN}) + (0.0376 \times \text{expression level of HAVCR1}) + (0.0120 \times \text{expression level of TXNRD1})$. Patients in the TCGA cohort were divided into low-risk and high-risk groups with median predictive index as the cut-off point. As Fig. 4A shown, we founded that low-risk HCC patients were significantly associated with better clinical outcomes. Moreover, to better judge the predictive ability of this prognostic signature, we conducted the ROC analysis with the risk score. The data showed that the area under the curve (AUC) predicted using the ROC curve for 1-year, 3-year, and 5-year overall survival was 0.804, 0.736, and 0.707, respectively (Fig. 4B). According to Fig. 4C, the upper panel shows the expression heat map of 5 prognostic model genes in the high- and low subgroup, the middle panel reveals that the risk of HCC patients increases with the increasing risk score, the lower panel demonstrates the poor survival time of the patients in the high-risk group compared to those in the low-risk group. For validating the predictive power of the signature, we employed the same formula to analyze the risk score of each patient in the ICGC data portal (LIRI-JP) as an independent external validation. Similarly, the Kaplan–Meier curves show a poor prognosis for the high-risk subgroup of HCC patients (Fig. 4D). In addition, the ROC curve also showed the good ability of the signature to predict the prognosis of the patients, and the AUCs for risk score predicting OS at 1-year, 3-year, 5-year was 0.714, 0.730, 0.726 (Fig. 4E). Likewise, the expression of the 5 CRGs and mortality in HCC patients increased in parallel with the rise in risk scores (Fig. 4F).

In dependent prognostic role of the gene signature

Furthermore, to investigate whether the CTCs/CTMs related 5-gene signature could be an independent prognostic factor in HCC patients, we compared the prognostic value of this signature with several clinicopathological factors such as age, gender, grade and AJCC stage for both cohorts by using univariate and multivariate Cox regression analyses. For the TCGA cohort, univariate Cox analysis of risk score ($p < 0.001$, HR = 1.258, 95%CI = 1.183–1.338), AJCC stage ($p < 0.001$, HR = 1.658, 95%CI = 1.340–2.053) and T status ($p < 0.001$, HR = 1.633, 95%CI = 1.332–2.003) were candidate factors (Fig. 4G). Further multivariate Cox regression analysis emphasized that risk score was an independent factor for HCC patients ($p < 0.001$, HR = 1.227 95%CI = 1.142–1.318) (Fig. 6B). For the ICGC cohort, univariate Cox regression analysis demonstrated that risk score ($p = 0.010$, HR = 1.076 95%CI = 1.017–1.137) and tumor stage ($p < 0.001$, HR = 2.252, 95%CI = 1.543–3.287) were potential risk factors (Fig. 4H). Multivariate Cox regression analysis also indicated that risk score ($p = 0.030$, HR = 1.070 95%CI = 1.006–1.138) and tumor stage ($p < 0.001$, HR = 2.330, 95%CI = 1.601–3.391) were independent predictor for ICGC patients (Fig. 6D). In conclusion, risk score was capable of being an independent risk factor for patients with hepatocellular carcinoma in the TCGA and ICGC cohorts.

Validation of the signature in GEO cohort

To further validate the prognostic signature, we employed the GES14520 dataset to evaluate the predictive power of this model. By using the same formula in the GEO cohort, we divided the HCC patients into high-risk and low-risk score groups. As the results illustrated in Fig. 5A, patients in the low-risk subgroup had better survival outcomes compared to the high-risk subgroup. In the GEO cohort, the AUC for 5-year OS reached 0.684 (Fig. 5B). The distribution of life status, time, and expression of model genes for each patient in the validation cohort is shown in Fig. 5C. As indicated by the results of univariate and multivariate Cox regression analyses, risk score ($p < 0.001$, HR = 1.280 95%CI = 1.145–1.432) and AJCC stage ($p < 0.001$, HR = 2.017 95%CI = 1.518–2.681) (Fig. 5D-E) were also considered to be independent prognosis factors. From the analysis of the above results, we found that the prognostic model is promising for predictor signature.

The relationship of the signature and the clinicopathological characteristics in HCC

To investigate the clinical utility of the CRGs signature, we plotted a heat map to reflect the potential correlation between the signature and clinicopathological features in the high-risk subgroup and low-risk subgroup (Fig. 6A). The expression of CDCA8 and TXNRD1 was higher in the high-risk group compared to the low-risk group. In addition, we found that T stage ($p < 0.001$), grade ($p < 0.001$), tumor stage ($p < 0.001$) and patient survival status ($p < 0.001$) were significantly associated with higher risk score (Fig. 6A-B).

Analysis of immune cell infiltration, immune checkpoints and TMB

We investigated the association between abundant tumor immune infiltration (CD4 + T cells, CD8 + T cells, B cells, neutrophils, macrophages and dendritic cells) and the expression of five signature genes (Fig. 7). Meanwhile, correlation coefficients > 0.3 and $p < 0.05$ were considered significant. As the result shown in the figure, we identified that CDCA8 was positive associated with the six immune cells, B cells (partial.cor = 0.441, $p = 9.08e-18$), CD8 + T cells (partial.cor = 0.303, $p = 1.03e-08$), CD4 + T cells (partial.cor = 0.359, $p = 6.74e-12$), macrophages (partial.cor = 0.439, $p = 1.70e-17$), neutrophils (partial.cor = 0.368, $p = 1.63e-12$), dendritic cells (partial.cor = 0.465, $p = 1.22e-19$). Similarly, HAVCR1 had positive correlations with B cells (partial.cor = 0.302, $p = 1.14e-08$), macrophages (partial.cor = 0.302, $p = 1.34e-08$), neutrophils (partial.cor = 0.392, $p = 4.00e-14$), dendritic cells (partial.cor = 0.317, $p = 2.18e-09$). And TXNRD1 was also positive associated with neutrophils (partial.cor = 0.322, $p = 8.67e-10$). Due to the important role of immune checkpoints in immune regulation and the potential application of immune checkpoint inhibitors in cancer therapy. Therefore, we evaluated the association between the signature genes and expression of immune checkpoints (including PD-1, PD-L1, CTLA4). The results in the Fig. 8A indicated that the expression of CDCA8 was positive associated with the three immune checkpoints, PD-1 ($R = 0.3$, $p = 4.9e-09$), PD-L1 ($R = 0.32$, $p = 3.8e-10$), CTLA4 ($R = 0.3$, $p = 4.9e-09$). Meanwhile, the expression of HAVCR1 was positively correlated with CTLA4 ($R = 0.32$, $p = 1.3e-10$). In addition, we evaluated the expression of immune checkpoints expression in high-risk and low-risk group. The expression of PD-1, PD-L1 and CTLA4 were remarkably higher in the high-risk group than in the low-risk HCC group (Fig. 8B-D). This suggests that the poor prognosis manifested by high-risk HCC patients is at least partially caused by the immunosuppressive microenvironment. Tumour mutational burden (TMB) is increasingly recognized as a potential biomarker of response to immune checkpoint inhibitors (ICIs) (29). Therefore, we assessed the correlation between TMB in high- and low-risk groups and then noticed that TMB levels were higher in the high-risk group compared to the low-risk group (Fig. 8E).

Pathway Analysis by GSEA

To further investigate the potential molecular mechanism for the signature genes and prognostic module, we utilized the Gens Set Enrichment Analysis (GSEA) in the HCC cohort. Figure 9A-E reveals the KEGG pathways of five signature genes (CDCA8, HAVCR1, MYCN, TP53I3, TXNRD1), which contains the five up-regulated and down-regulated pathways, respectively. The signature genes are mainly enriched in pathways including the KEGG cell cycle, KEGG p53 signaling pathway, KEGG complement, and coagulation cascades, KEGG drug metabolism cytochrome p450, etc. In addition, GSEA compared high and low-risk groups according to risk scores. KEGG pathways enriched in the high-risk group involved KEGG cell cycle, KEGG WNT signaling pathway, KEGG ubiquitin-mediated proteolysis, KEGG snare interactions in vesicular transport, and KEGG p53 signaling pathway. KEGG pathways enriched in the low-risk group included KEGG complement and coagulation cascades, KEGG fatty acid metabolism, KEGG primary bile acid biosynthesis, KEGG tryptophan metabolism, and KEGG valine leucine and isoleucine degradation (Fig. 9F).

Identification of related small molecule drugs

We utilized cMap database and Comparative Toxicogenomics Database (CTD) to further explore potential small molecule drugs for HCC. We obtained the 13 potential drugs based on the cMap and screened five chemicals (apigenin, sulfadimethoxine, guanabenz, nadolol, and ginkgolide A) for potential therapy of HCC by CTD (Table 1). In addition, we employed the PubChem (<https://pubchem.ncbi.nlm.nih.gov/>) to display the interactive chemical structure model of five potential small molecule drugs (Fig. 10A).

Table 1
The significant small molecule drugs for HCC after CMap and CTD analysis.

rank	cmap name	mean	n	enrichment	p	specificity	% non-null	CTD Inference Score
1	apigenin	-0.816	4	-0.956	0.00000	0.0000	100	50
2	Prestwick-692	0.672	4	0.917	0.00004	0.0061	100	none
3	adiphenine	0.709	5	0.907	0.00004	0.0000	100	none
4	0175029-0000	-0.744	6	-0.830	0.00006	0.0352	100	none
5	sulfadimethoxine	0.571	5	0.870	0.00008	0.0000	100	3
6	felbinac	0.669	4	0.882	0.00024	0.0059	100	none
7	3-acetamidocoumarin	0.608	4	0.879	0.00030	0.0000	100	none
8	guanabenz	0.560	5	0.807	0.00062	0.0000	100	3
9	timolol	0.553	4	0.852	0.00072	0.0056	100	none
10	podophyllotoxin	0.658	4	0.851	0.00072	0.0286	100	none
11	nadolol	0.612	4	0.848	0.00076	0.0110	100	1
12	ginkgolide A	-0.763	4	-0.850	0.00095	0.0000	100	7
13	pheneticillin	0.633	4	0.844	0.00097	0.0063	100	none
CTD, Comparative Toxicogenomics Database, none, the small-molecule drugs for HCC not validated in CTD								

Establishing and validating a predictive nomogram

To better predict the survivability of HCC patients, we established a nomogram to predict the probability of 1-, 3- and 5-years OS in the TCGA cohort. Meanwhile, the predictors of the nomogram consisted of age, gender, stage, risk score, etc. In addition, calibration plots were developed to assess the predictive power of the nomogram (Fig. 10B-C). Similarly, we also constructed two nomograms in ICGC and GSE14520 cohorts (Figure S1). All of these findings indicate the good predictive power of the prediction model.

CDCA8 served as a diagnostic biomarker and upregulated in HCC patients

To further analyze the signature genes, we verified the expression of the five genes in the six GEO databases. As shown in Figure S2A, we found that CDCA8 and TP53I3 were markedly increased in the tumor tissues. Besides, the Kaplan-Meier survival analysis indicated that CDCA8, HAVCR1, and MYCN were related to prognosis in the ICGC dataset (Figure S2B). In conclusion, from all the results we can assume that CDCA8 can be used as a potential target in HCC. So, we selected the CDCA8 for further analysis. Firstly, based on the TCGA data, the expression level of CDCA8 was significantly higher in HCC tissues compared to normal tissues (Fig. 11A). And Fig. 11B illustrates the shorter survival time of HCC patients with higher CDCA8 expression. By analyzing the correlation between CDCA8 and clinical features, we identified that the expression of CDCA8 was significantly associated with the HCC tumor grade, tumor stage, and T stage (Fig. 11C-E). In addition, the upregulation of CDCA8 on the protein was further confirmed in a representative set of eight paired HCC tissues and normal tissues (Fig. 11F). We also explored the translational expression levels of CDCA8 using the Human Protein Atlas (HPA) database (Fig. 11G), and indicating CDCA8 was increased in the tumor tissues compared to normal tissues.

Downregulation of CDCA8 inhibits HCC cell proliferation, invasion and migration

To elucidate whether CDCA8 affects the biological development of HCC, we first assessed the expression of CDCA8 in cell lines. We found that CDCA8 was expressed in most cell lines, notably in HepG2 and Huh7 (Fig. 12A-B). To investigate further, we knocked down CDCA8 in HepG2 and Huh7 cells using siRNA, and the transfection efficiency was validated by western blot analysis and qRT-PCR (Fig. 12C-D). Therefore, we selected the siRNA1-CDCA8 and siRNA2-CDCA8 for subsequent experiments in vitro. To investigate the effect of CDCA8 on the proliferation of HCC cells in vitro, we performed CCK-8 and clone formation assays. As shown in the figure, the proliferation ability of HCC cells was significantly inhibited after CDCA8 depletion (Fig. 12E-F). To determine the ability of CDCA8 to affect HCC metastasis, we assessed HCC cell migration and invasion ability in vitro using wound healing assays as well as transwell migration and invasion assays. As the results demonstrated, depletion of CDCA8 significantly inhibited the migration and invasion ability of HepG2 and Huh7 cells compared to negative controls (Fig. 12G-H). Overall, these results indicate that CDCA8 promotes proliferation, migration, and invasion of HCC.

Knockdown of CDCA8 induces apoptosis and cycle arrest in HCC cells

To assess whether CDCA8 affects the growth cycle and apoptosis of HCC cells, we conducted flow cytometry assays to investigate the impacts of CDCA8 knockdown on apoptosis and cell cycle regulation in HCC cells. Cell cycle experiments revealed that downregulation of CDCA8 decreased the proportion of S phase and increased the proportion of G2/M phase in HepG2 and Huh7 cells, suggesting that inhibition of

CDCA8 gene expression may lead to G2/M phase growth arrest in HCC cells (Fig. 13A). In addition, the results emphasized that the silencing of CDCA8 significantly increased the apoptosis rate of HCC cells (Fig. 13B).

Discussion

Hepatocellular carcinoma remains a significant challenge to human health, with a high rate of recurrence even after surgical resection. In recent years, many studies have demonstrated that CTCs are closely related to the metastasis, epithelial-mesenchymal transition (EMT), and recurrence of malignant tumors, including HCC (30). Therefore, it is critical to screen molecules associated with CTCs to identify validated biomarkers for predicting HCC. Here, we constructed a reliable prognostic signature based on CTCs related genes and explored its clinical utilization in HCC patients. Moreover, we initially investigated the biological role of CDCA8 in hepatocellular carcinoma and identified CDCA8 as a potential therapeutic target.

Firstly, we identified 258 CRGs by analyzing the DEGs in TCGA and ICGC databases and CTC expression profiles of hepatocellular carcinoma. Then, these genes were screened to construct a five-CRGs signature in the TCGA cohort. Next, we utilized the KM survival analysis and ROC analysis to confirm the prognostic value of the signature. In addition, similar results were also validated in the ICGC cohort and GEO cohort. Univariate and multifactorial Cox analyses confirmed that signature was an independent prognostic factor. Moreover, the nomogram for all three cohorts demonstrated a good predictive power of the model.

To better understand the interactions between these CRGs, we further analyzed them using the STRING website and visualized them with Cytoscape software. In addition, we identified ten genes that may play important roles in HCC development, including TOP2A, CCNB2, CDCA8, BIRC5, AURKB, CCNB1, BUB1, BUB1B, KIF20A, TTK. Many genes have been validated in HCC, such as TTK as a target for intrahepatic spread therapy in hepatocellular carcinoma; BUB1B activates mTORC1 to promote hepatocellular carcinoma progression; KIF20A may serve as an independent prognostic factor for overall and disease-free survival in hepatocellular carcinoma patients (31–33). As indicated by KEGG, CRGs promote development, metastasis, and recurrence of hepatocellular carcinoma probably through the cell cycle and p53 signaling pathway. GSEA analysis of five signature genes and high-risk groups in the prognostic model identified various oncogenesis-related features, including cell cycle, p53 signaling pathway, WNT signaling pathway, DNA replication, TGF-beta signaling pathway, and notch signaling pathway.

To evaluate the value of the signature in clinical terms, we explored correlations with clinical characteristics and found that the high-risk group was positively associated with tumor stage, tumor grade, and T stage. Immune cell infiltration analysis revealed that CDCA8 and HAVCR1 may be associated with several immune cell abundances, like B cell, CD8 + T cell, macrophage, neutrophil, and dendritic cell. In recent years, immune checkpoint inhibitors have shown a potential therapeutic role for advanced HCC (34). As mentioned previously, immune checkpoints such as PD-1, PD-L1, and CTLA4 and tumor mutation burden (TMB) were highly expressed in the high-risk subgroup, suggesting that the high-risk group patients may benefit more from immunotherapy. In addition, Pearson correlation analysis indicated that CDCA8

and HAVCR1 were also positively correlated with immune checkpoints, which further revealed the contributions of these two genes to the development of hepatocellular carcinoma in terms of immunity.

Notably, we screened five small molecule drugs that were highly relevant to CRGs by using cMAP, including apigenin, sulfadimethoxine, guanabenz, nadolol, and ginkgolide A. Apigenin, an edible plant flavonoid, is an antioxidant and can also be used as a therapeutic agent to overcome several diseases and even cancer. Previous studies found that apigenin inhibits EMT and metastasis in hepatocellular carcinoma through NF- κ B and snail signaling pathways and also enhances the cytotoxicity of sorafenib on HepG2 cells (35, 36). Several reports have shown that sulfadimethoxine has antitumor effects in in vitro cellular assays, but the biological contribution in hepatocellular carcinoma needs to be further investigated (37, 38). Guanabenz (GBZ), an α 2-adrenergic agonist, is an FDA-approved antihypertensive drug with a good safety profile. Kang et al. demonstrated the antitumor value of GBZ in human HCC cell lines (39). In addition, GBZ not only inhibits the invasion and migration of chondrosarcoma but also serves as a potential therapeutic agent for triple-negative breast cancer (40, 41). Ginkgolide A is a platelet-activating factor antagonist extracted from Ginkgo biloba, and a literature review mentioned its many biological effects, such as anti-inflammatory, anti-cancer, anti-anxiety, anti-atherosclerotic, neurological, and hepatoprotective effects (42). However, a recent report highlighted that another derivative from Ginkgo biloba, ginkgolide C, exhibited its anti-hepatocarcinoma effect by modulating c-Met phosphorylation (43). In conclusion, our analysis is undoubtedly helpful, but its application in future studies is necessary.

In the prognostic signature, we screened five CRGs, including HAVCR1, TP53I3, MYCN, TXNRD1, and CDCA8. HAVCR1 is highly expressed in a variety of tumors, including colorectal cancer, non-small-cell lung cancer, clear cell renal cell carcinoma, and hepatocellular carcinoma; and can be an independent prognostic factor (44–47). Moreover, Ye et al. found that TIM-1⁺ (HAVCR1⁺) regulatory B cell infiltration was significantly higher in tumor tissues of HCC patients compared with paraneoplastic tissues and could promote the immune escape of hepatocellular carcinoma cells, implying that it could be used as a new immune therapeutic target (47). The p53-inducible gene 3 (PIG3), also known as TP53I3, is involved in the process of apoptosis and DNA damage response. Previous studies have revealed that TP53I3 promotes invasion and metastasis of lung cancer cells and that silencing TP53I3 enhances the chemosensitivity of non-small cell lung cancer to docetaxel (48, 49). These findings may provide us with a new strategy for the treatment of liver cancer. MYCN, a member of the MYC proto-oncogene family, Qin et al. highlighted that MYCN may be a stem cell-like marker for HCC and could be used as a therapeutic target of acyclic retinoid (ACR) for HCC (50, 51). TXNRD1 is an antioxidant enzyme that has been reported to be overexpressed in hepatocellular carcinoma. Lee et al. observed that inhibition of TXNRD1 suppressed HCC cell proliferation, promoted apoptosis, and could induce oxidative stress, which implies that it could be used as a therapeutic target for HCC (52). In addition, CDCA8 is a protein associated with the human cell division cycle and has been reported to be involved in a variety of tumorigenesis. Recent results have suggested that overexpression of CDCA8 enhances the proliferation of breast cancer, bladder cancer, and cutaneous melanoma cells and leads to poor prognosis. Interestingly, our study found that downregulation of CDCA8 inhibited proliferation, invasion, and metastasis of HCC cells and could lead to cell cycle arrest and

increasing apoptosis. Therefore, targeting CDCA8 may be the next molecular strategy for the treatment and prevention of metastasis from HCC.

In our study, we focused on the analysis of CTC-related genes in HCC and constructed a signature of five CRGs. Through the validation of several databases, we found that this signature has a good predictive and clinical application profile. It is easier to apply in clinical practice as it reduces the necessity and medical expenses of whole-genome sequencing for patients. And, we explored the correlation with immune checkpoints and observed that high-risk patients may be better suited for immunotherapy. In addition, we established nomograms and screened candidate small molecule drugs, which may provide clinical benefits. However, our paper also inevitably has some limitations. Firstly, our study is still a prospective study, and future multicenter studies are needed to further validate the accuracy of the model. Secondly, since the database lacks tumor progression variables, such as vascular invasion status and tumor size, it may affect the statistical power of the nomogram and influence the predictive abilities. Finally, although we have initially explored the function of CDCA8 in HCC development, the specific molecular mechanisms are not clear; therefore, it is necessary to further elucidate the molecular mechanisms linking CDCA8 to circulating tumor cells.

Conclusions

In summary, our study provides further insight into the function of CTC in HCC and proposed a five CRGs signature that could be personalized to predict patient prognosis. In addition, we highlighted the important role of CDCA8 in hepatocellular carcinoma, while targeting CDCA8 could be the next potential therapeutic strategy.

Abbreviations

HCC, hepatocellular carcinoma; CTCs/CTMs, Circulating tumor cells/microemboli; CRGs, CTCs/CTMs related genes; TCGA, The Cancer Genome Atlas; ICGC, The International Cancer Genome Consortium database; TMB, tumor mutation burden; DEGs, differentially expressed genes; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; PPI, protein–protein interaction; ROC, receiver operating characteristic; GSEA, Gene Set Enrichment Analysis; LASSO, the least absolute shrinkage and selection operator;

Declarations

Disclosures: The authors disclose no conflicts.

Ethics approval and consent to participate

The study methods met the criteria set on the Declaration of Helsinki and ethical approval was secured from the Ethics Committee of Tongji Hospital, Huazhong University of Science and Technology. Informed consent was also obtained from all patients.

Consent for publication

Not applicable.

Availability of data and materials

All data generated or analyzed during this study are included in this published article and its supplementary information file.

Competing interests

The authors declare that there is no conflict of interest.

FUNDING

This study was funded by the National Natural Science Foundation of China (No. 81874062; No.82072730) and Youth Program of National Natural Science Foundation of China (N0.81902439).

AUTHOR CONTRIBUTIONS

X.L., Z.D. and W.Y. designed the study. X.L., Y.S. and P.Q. analyzed the data. Y.L., K.Z. and Z.D. completed the experiments. X.L., and Y.S. wrote the manuscript and were responsible for language revisions. W.Y. and J.W. supervised the project. All authors reviewed the manuscript.

Acknowledgements

We sincerely thank the public database for using their data.

References

1. Fitzmaurice C, Allen C, Barber RM, Barregard L, Bhutta ZA, Brenner H, et al. Global, Regional, and National Cancer Incidence, Mortality, Years of Life Lost, Years Lived With Disability, and Disability-Adjusted Life-years for 32 Cancer Groups, 1990 to 2015: A Systematic Analysis for the Global Burden of Disease Study. *JAMA oncology*. 2017;3(4):524-48.
2. Villanueva A. Hepatocellular Carcinoma. *The New England journal of medicine*. 2019;380(15):1450-62.
3. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: a cancer journal for clinicians*. 2021;71(3):209-49.
4. Llovet JM, Kelley RK, Villanueva A, Singal AG, Pikarsky E, Roayaie S, et al. Hepatocellular carcinoma. *Nature reviews Disease primers*. 2021;7(1):6.
5. Marrero JA, Kulik LM, Sirlin CB, Zhu AX, Finn RS, Abecassis MM, et al. Diagnosis, Staging, and Management of Hepatocellular Carcinoma: 2018 Practice Guidance by the American Association for

- the Study of Liver Diseases. *Hepatology* (Baltimore, Md). 2018;68(2):723-50.
6. Kluger MD, Salceda JA, Laurent A, Tayar C, Duvoux C, Decaens T, et al. Liver resection for hepatocellular carcinoma in 313 Western patients: tumor biology and underlying liver rather than tumor size drive prognosis. *Journal of hepatology*. 2015;62(5):1131-40.
 7. Zhao L, Wu X, Li T, Luo J, Dong D. ctcRbase: the gene expression database of circulating tumor cells and microemboli. *Database : the journal of biological databases and curation*. 2020;2020.
 8. Alix-Panabières C, Pantel K. Circulating tumor cells: liquid biopsy of cancer. *Clinical chemistry*. 2013;59(1):110-8.
 9. Allard WJ, Matera J, Miller MC, Repollet M, Connelly MC, Rao C, et al. Tumor cells circulate in the peripheral blood of all major carcinomas but not in healthy subjects or patients with nonmalignant diseases. *Clinical cancer research : an official journal of the American Association for Cancer Research*. 2004;10(20):6897-904.
 10. Cristofanilli M, Budd GT, Ellis MJ, Stopeck A, Matera J, Miller MC, et al. Circulating tumor cells, disease progression, and survival in metastatic breast cancer. *The New England journal of medicine*. 2004;351(8):781-91.
 11. Aceto N, Bardia A, Miyamoto DT, Donaldson MC, Wittner BS, Spencer JA, et al. Circulating tumor cell clusters are oligoclonal precursors of breast cancer metastasis. *Cell*. 2014;158(5):1110-22.
 12. Hou JM, Krebs MG, Lancashire L, Sloane R, Backen A, Swain RK, et al. Clinical significance and molecular characteristics of circulating tumor cells and circulating tumor microemboli in patients with small-cell lung cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2012;30(5):525-32.
 13. Ye Q, Ling S, Zheng S, Xu X. Liquid biopsy in hepatocellular carcinoma: circulating tumor cells and circulating tumor DNA. *Molecular cancer*. 2019;18(1):114.
 14. De Rubis G, Rajeev Krishnan S, Bebawy M. Liquid Biopsies in Cancer Diagnosis, Monitoring, and Prognosis. *Trends in pharmacological sciences*. 2019;40(3):172-86.
 15. Yu JJ, Xiao W, Dong SL, Liang HF, Zhang ZW, Zhang BX, et al. Effect of surgical liver resection on circulating tumor cells in patients with hepatocellular carcinoma. *BMC cancer*. 2018;18(1):835.
 16. Kelley RK, Magbanua MJ, Butler TM, Collisson EA, Hwang J, Sidiropoulos N, et al. Circulating tumor cells in hepatocellular carcinoma: a pilot study of detection, enumeration, and next-generation sequencing in cases and controls. *BMC cancer*. 2015;15:206.
 17. Sun YF, Xu Y, Yang XR, Guo W, Zhang X, Qiu SJ, et al. Circulating stem cell-like epithelial cell adhesion molecule-positive tumor cells indicate poor prognosis of hepatocellular carcinoma after curative resection. *Hepatology* (Baltimore, Md). 2013;57(4):1458-68.
 18. Wang Y, Navin NE. Advances and applications of single-cell sequencing technologies. *Molecular cell*. 2015;58(4):598-609.
 19. Abouleila Y, Onidani K, Ali A, Shoji H, Kawai T, Lim CT, et al. Live single cell mass spectrometry reveals cancer-specific metabolic profiles of circulating tumor cells. *Cancer science*. 2019;110(2):697-706.

20. 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):284-7.
21. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America*. 2005;102(43):15545-50.
22. Bader GD, Hogue CW. An automated method for finding molecular complexes in large protein interaction networks. *BMC bioinformatics*. 2003;4:2.
23. Chin CH, Chen SH, Wu HH, Ho CW, Ko MT, Lin CY. cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC systems biology*. 2014;8 Suppl 4(Suppl 4):S11.
24. Tibshirani R. The lasso method for variable selection in the Cox model. *Statistics in medicine*. 1997;16(4):385-95.
25. Lossos IS, Czerwinski DK, Alizadeh AA, Wechser MA, Tibshirani R, Botstein D, et al. Prediction of survival in diffuse large-B-cell lymphoma based on the expression of six genes. *The New England journal of medicine*. 2004;350(18):1828-37.
26. Iasonos A, Schrag D, Raj GV, Panageas KS. How to build and interpret a nomogram for cancer prognosis. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2008;26(8):1364-70.
27. Subramanian A, Narayan R, Corsello SM, Peck DD, Natoli TE, Lu X, et al. A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. *Cell*. 2017;171(6):1437-52.e17.
28. Davis AP, Grondin CJ, Johnson RJ, Sciaky D, McMorran R, Wieggers J, et al. The Comparative Toxicogenomics Database: update 2019. *Nucleic acids research*. 2019;47(D1):D948-d54.
29. Samstein RM, Lee CH, Shoushtari AN, Hellmann MD, Shen R, Janjigian YY, et al. Tumor mutational load predicts survival after immunotherapy across multiple cancer types. *Nature genetics*. 2019;51(2):202-6.
30. Qi LN, Xiang BD, Wu FX, Ye JZ, Zhong JH, Wang YY, et al. Circulating Tumor Cells Undergoing EMT Provide a Metric for Diagnosis and Prognosis of Patients with Hepatocellular Carcinoma. *Cancer Res*. 2018;78(16):4731-44.
31. Miao R, Wu Y, Zhang H, Zhou H, Sun X, Csizmadia E, et al. Utility of the dual-specificity protein kinase TTK as a therapeutic target for intrahepatic spread of liver cancer. *Scientific reports*. 2016;6:33121.
32. Qiu J, Zhang S, Wang P, Wang H, Sha B, Peng H, et al. BUB1B promotes hepatocellular carcinoma progression via activation of the mTORC1 signaling pathway. *Cancer medicine*. 2020;9(21):8159-72.
33. Lu M, Huang X, Chen Y, Fu Y, Xu C, Xiang W, et al. Aberrant KIF20A expression might independently predict poor overall survival and recurrence-free survival of hepatocellular carcinoma. *IUBMB life*. 2018;70(4):328-35.
34. Chang H, Jung W, Kim A, Kim HK, Kim WB, Kim JH, et al. Expression and prognostic significance of programmed death protein 1 and programmed death ligand-1, and cytotoxic T lymphocyte-associated molecule-4 in hepatocellular carcinoma. *APMIS : acta pathologica, microbiologica, et immunologica Scandinavica*. 2017;125(8):690-8.

35. Qin Y, Zhao D, Zhou HG, Wang XH, Zhong WL, Chen S, et al. Apigenin inhibits NF- κ B and snail signaling, EMT and metastasis in human hepatocellular carcinoma. *Oncotarget*. 2016;7(27):41421-31.
36. Şirin N, Elmas L, Seçme M, Dodurga Y. Investigation of possible effects of apigenin, sorafenib and combined applications on apoptosis and cell cycle in hepatocellular cancer cells. *Gene*. 2020;737:144428.
37. Khowdiary MM, Mostafa NS. Synthesis of sulfadimethoxine based surfactants and their evaluation as antitumor agents. *Journal of cancer research and therapeutics*. 2016;12(1):137-41.
38. Vila-Caballer M, Codolo G, Munari F, Malfanti A, Fassan M, Rugge M, et al. A pH-sensitive stearyl-PEG-poly(methacryloyl sulfadimethoxine)-decorated liposome system for protein delivery: An application for bladder cancer treatment. *Journal of controlled release : official journal of the Controlled Release Society*. 2016;238:31-42.
39. Kang HJ, Seol HS, Lee SE, Suh YA, Kim J, Jang SJ, et al. Guanabenz Acetate Induces Endoplasmic Reticulum Stress-Related Cell Death in Hepatocellular Carcinoma Cells. *Journal of pathology and translational medicine*. 2019;53(2):94-103.
40. Xu W, Wan Q, Na S, Yokota H, Yan JL, Hamamura K. Suppressed invasive and migratory behaviors of SW1353 chondrosarcoma cells through the regulation of Src, Rac1 GTPase, and MMP13. *Cellular signalling*. 2015;27(12):2332-42.
41. Haggag YA, Yasser M, Tambuwala MM, El Tokhy SS, Isreb M, Donia AA. Repurposing of Guanabenz acetate by encapsulation into long-circulating nanopolymersomes for treatment of triple-negative breast cancer. *Int J Pharm*. 2021;600:120532.
42. Sarkar C, Quispe C, Jamaddar S, Hossain R, Ray P, Mondal M, et al. Therapeutic promises of ginkgolide A: A literature-based review. *Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie*. 2020;132:110908.
43. Yang MH, Baek SH, Um JY, Ahn KS. Anti-neoplastic Effect of Ginkgolide C through Modulating c-Met Phosphorylation in Hepatocellular Carcinoma Cells. *International journal of molecular sciences*. 2020;21(21).
44. Wang Y, Martin TA, Jiang WG. HAVcR-1 expression in human colorectal cancer and its effects on colorectal cancer cells in vitro. *Anticancer research*. 2013;33(1):207-14.
45. Zheng X, Xu K, Chen L, Zhou Y, Jiang J. Prognostic value of TIM-1 expression in human non-small-cell lung cancer. *Journal of translational medicine*. 2019;17(1):178.
46. Cuadros T, Trilla E, Sarró E, Vilà MR, Vilardell J, de Torres I, et al. HAVCR/KIM-1 activates the IL-6/STAT-3 pathway in clear cell renal cell carcinoma and determines tumor progression and patient outcome. *Cancer Res*. 2014;74(5):1416-28.
47. Ye L, Zhang Q, Cheng Y, Chen X, Wang G, Shi M, et al. Tumor-derived exosomal HMGB1 fosters hepatocellular carcinoma immune evasion by promoting TIM-1(+) regulatory B cell expansion. *Journal for immunotherapy of cancer*. 2018;6(1):145.

48. Gu MM, Gao D, Yao PA, Yu L, Yang XD, Xing CG, et al. p53-inducible gene 3 promotes cell migration and invasion by activating the FAK/Src pathway in lung adenocarcinoma. *Cancer science*. 2018;109(12):3783-93.
49. Li M, Li S, Liu B, Gu MM, Zou S, Xiao BB, et al. PIG3 promotes NSCLC cell mitotic progression and is associated with poor prognosis of NSCLC patients. *Journal of experimental & clinical cancer research* : CR. 2017;36(1):39.
50. Qin XY, Suzuki H, Honda M, Okada H, Kaneko S, Inoue I, et al. Prevention of hepatocellular carcinoma by targeting MYCN-positive liver cancer stem cells with acyclic retinoid. *Proceedings of the National Academy of Sciences of the United States of America*. 2018;115(19):4969-74.
51. Qin XY, Su T, Yu W, Kojima S. Lipid desaturation-associated endoplasmic reticulum stress regulates MYCN gene expression in hepatocellular carcinoma cells. *Cell death & disease*. 2020;11(1):66.
52. Lee D, Xu IM, Chiu DK, Leibold J, Tse AP, Bao MH, et al. Induction of Oxidative Stress Through Inhibition of Thioredoxin Reductase 1 Is an Effective Therapeutic Approach for Hepatocellular Carcinoma. *Hepatology (Baltimore, Md)*. 2019;69(4):1768-86.

Figures

Figure 1

Identification of CTCs/CTMs related genes. (A-B) Volcano plots of the differentially expressed genes in TCGA dataset (A) and ICGC dataset (B). Green dots represent down-regulated genes and red dots represent up-regulated genes. (C) The Venn diagram shows the intersection of the three datasets (TCGA, ICGC, and ctcRbase), and a total of 258 CRGs were identified.

Figure 2

PPI networks and enrichment analysis of the CRGs. (A) PPI network comprises 155 genes, and interaction score >0.7 was set as significant. Yellow represented module 1, with genes >50 by MCODE. (B) module 1, blue represented the top 10 hub genes calculated by cytoHubba. (C) GO enrichment analysis including biological process, cellular component, and molecular function. Each item contains the top ten results. (D) Six pathways were enriched by KEGG.

Figure 3

Construction of the prognostic signature. (A) Univariate Cox regression analysis identified 88 prognosis-related CRGs. (B-C) The LASSO Cox regression analysis excluded relatively insignificant parameters. (D) Five signature genes were identified by multivariate Cox regression.

Figure 4

Validation the prognostic signature in HCC. (A) Kaplan-Meier survival analysis of high-risk and low-risk risk group in TCGA cohort. (B) ROC curves for 1-, 3- and 5-year survival in TCGA cohort. (C) The distribution of CRGs levels (upper), risk score (middle), and survival time (lower) in the TCGA cohort. (D-F) Survival analysis, time-dependent ROC analysis, and risk score model in ICGC cohort, respectively. Univariate (upper) and multivariate (lower) analysis of in-dependent prognostic factors for overall survival in TCGA cohort (G) and ICGC cohort (H).

Figure 5

Validation of the signature in GEO cohort. (A) Kaplan-Meier survival analysis, (B) time-dependent ROC analysis, and (C) risk score model in GSE14520 dataset. (D-E) Univariate and multivariate analysis of independent prognostic factors for overall survival in the GEO co-hort.

Figure 6

The relationship between the prognostic signature and clinicopathological characteristics. (A) The heatmap showing the correlation of prognostic signature with clinicopathological characteristics and five signature genes. (B) Comparison of the age, gender, stage, grade, T stage, N stage, M stage, and patient survival status in the low- and high-risk groups, respectively. ($p < 0.05$, *; $p < 0.01$, **; $p < 0.001$, ***; ns, not significant).

Figure 7

The correlation between the five CRGs and six immune cells in TIMER.

Figure 8

Analysis of immune correlation. (A) The correlation between signature genes and immune checkpoints (including PD-1, PD-L1, and CTLA4). (B-E) Comparison of the PD-1, PD-L1, CTLA4, and TMB in the low- and high-risk groups, respectively.

Figure 9

Gene set enrichment analysis (GSEA) of the CRGs and the prognostic signature. (A-E) GSEA analysis for CDCA8, HAVCR1, MYCN, TP53I3, and TXNRD1. Each project includes five up-regulated and five down-regulated pathways. (F) GSEA analysis between different risk groups in TCGA cohort.

Figure 10

Screening of potential small molecule drugs and construct the nomogram for the TCGA cohort. (A) The interactive chemical structure model of five potential small molecule drugs. (B) The nomogram was developed for predicting 1-, 3- and 5-year survival of HCC patients. (C) Calibration curves for 1-, 3- and 5-year survival prediction of the nomogram in TCGA cohort.

Figure 11

The expression and clinical significance of CDCA8 in HCC. (A) The expression of CDCA8 in TCGA database. (B) The Kaplan-Meier survival analysis of CDCA8 in TCGA dataset. (C-E) Comparison of the expression of CDCA8 in different clinical features (tumor grade, tumor stage, and T stage) in TCGA-LIHC. (F) CDCA8 protein expression in representative HCC samples (T) and pair-matched normal tissues (N)

evaluated by western blotting. (G) The representative protein expression of the CDCA8 in HCC and normal liver tissue. Data were from the Human Protein Atlas (<http://www.proteinatlas.org>) online database.

Figure 12

Silencing CDCA8 inhibited the proliferation, migration, and invasion of HCC cells in vivo. (A-B) Western blotting and qPCR results of CDCA8 expression in different HCC cell lines. (C-D) CDCA8 was stably knocked down in HCC HepG2 and Huh7 cells, and the protein levels of CDCA8 were detected by western

blotting and qPCR. (E-F) CCK-8 and clone formation assays were performed to analyze HCC cell proliferation after CDCA8 knockdown. (G) Transwell experiments were used to evaluate the cell migration and invasion in CDCA8 knockdown HCC cells. Scale bar, 100 magnification. (H) Wound healing assays showing the capacity of HCC cell migration. Scale bar, 40 magnification. (p < 0.05, *; p < 0.01, **; p < 0.001, ***; ns, not significant).

Figure 13

Knockdown of CDCA8 induces apoptosis and cycle arrest in HCC cells. (A) Effect of CDCA8 knockdown on HepG2 and Huh7 cell cycle. (B) Apoptosis rate after CDCA8 knockdown was detected by flow cytometry. (p < 0.05, *; p < 0.01, **; p < 0.001, ***; ns, not significant).

Supplementary Files

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

- [FigureS1.tif](#)
- [FigureS2.tif](#)