Definition of a novel vascular invasion-associated multi-gene signature for predicting survival in patients with hepatocellular carcinoma

BO YI, CAIXI TANG, YIN TAO and ZHIJIAN ZHAO

Department of Hepatobiliary Surgery, Zhu Zhou Central Hospital, Zhuzhou, Hunan 412007, P.R. China

Received January 5, 2019; Accepted September 11, 2019

DOI: 10.3892/ol.2019.11072

Abstract. The aim of the present study was to identify a vascular invasion-associated gene signature for predicting prognosis in patients with hepatocellular carcinoma (HCC). Using RNA-sequencing data of 292 HCC samples from The Cancer Genome Atlas (TCGA), the present study screened differentially expressed genes (DEGs) between patients with and without vascular invasion. Feature genes were selected from the DEGs by support vector machine (SVM)-based recursive feature elimination (RFE-SVM) algorithm to build a classifier. A multi-gene signature was selected by L1 penalized (LASSO) Cox proportional hazards (PH) regression model from the feature genes selected by the RFE-SVM to develop a prognostic scoring model. TCGA set was defined as the training set and was divided by the gene signature into a high-risk group and a low-risk group. Involvement of the DEGs between the two risk groups in pathways was also investigated. The presence and absence of vascular invasion between patients of training set was 175 DEGs. A classification model of 42 genes performed well in differentiating patients with and without vascular invasion on the training set and the validation set. A 14-gene prognostic model was built that could divide the training set or the validation set into two risk groups with significantly different survival outcomes. A total of 762 DEGs in the two risk groups of the training set were revealed to be significantly associated with a number of signaling pathways. The present study provided a 42-gene classifier for predicting vascular invasion, and identified a vascular invasion-associated 14-gene signature for predicting prognosis in patients with HCC. Several genes and pathways in HCC development are characterized and may be potential therapeutic targets for this type of cancer.

Introduction

Hepatocellular carcinoma (HCC) is a major type of primary liver cancer (1). The mortality rate is increasing, and patients with the tumor present with a poor prognosis (2). An increasing number of studies have demonstrated that vascular invasion is an adverse prognostic factor in HCC (3-5). Furthermore, vascular invasion is an independent predictive factor of long-term survival in patients with early-stage HCC, and is significantly associated with intrahepatic metastasis (6). Hence, it is extremely necessary to differentiate patients with HCC that present with vascular invasion from those patients with HCC that do not present with vascular invasion, so as to improve survival time.

A risk classification model of micro-vascular invasion based on histopathological features has been introduced for predicting the prognosis of patients with HCC (7). Differentially expressed genes (DEGs) in HCC tissue samples in the presence or absence of vascular invasion have been studied in order to extract multi-gene signatures for detecting vascular invasion (8,9). High-throughput technologies allow for the development of a classification model, wherein vascular invasion information can be derived from molecular features. The Cancer Genome Atlas (TCGA) provides comprehensive maps of genomic alterations in various types of cancer (https://portal.gdc.cancer.gov/). A recent study derived a 16-miRNA-based classifier from the analysis of micro (mi)RNA and mRNA expression data derived from TCGA, which could effectively identify vascular invasion and predict overall survival (OS) (10). These studies indicated the feasibility of these multi-gene signatures for prediction of cancer prognosis. Nevertheless, more efforts should be made in order to generate more reliable and accurate prognostic models based on feature genes of vascular invasion.

The present study analyzed HCC RNA-sequencing data from TCGA in order to identify feature genes using a recursive feature elimination (RFE) method (11), thus constructing a support vector machine (SVM) classifier for separating patients with vascular invasion from patients without vascular invasion. Furthermore, L1 penalized (LASSO) Cox proportional hazards (PH) regression model was used to determine prognostic genes from the identified feature genes of vascular invasion so as to develop a prognostic scoring model. The performance of the classifier and the prognostic model was...
tested on an independent set. In addition, a function analysis was performed in order to provide further insights into the molecular mechanisms underlying HCC.

Materials and methods

Data resource. The present study obtained the RNA-sequencing data of 373 HCC samples from TCGA portal based on Illumina HiSeq 2000 RNA Sequencing platform (Download date: 18th, October, 2018). Among these samples, 292 had clinical information of vascular invasion and survival information, including survival time and survival status, and were therefore selected as the training set (TCGA set). Furthermore, the GSE10141 (12) dataset was downloaded from Gene Expression Omnibus (GEO; https://www.ncbi.nlm.nih.gov/) at the National Center for Biotechnology Information (NCBI; https://www.ncbi.nlm.nih.gov/) based on the GPL5474 Human 6k Transcriptionally Informative Gene Panel platform, including the microarray gene expression data of 80 HCC tissue samples with survival information. Only 62 HCC samples had vascular invasion, and these were selected as the validation set.

The present study performed uni- and multivariate Cox regression analyses in order to analyze the associations between clinical factors and OS in the training set using survival package v2.44-1.1 (13) of R language (http://bioconductor.org/packages/survival/). The significant clinical factors (log-rank P<0.05) were selected as the cut-off to classify the training set.

Differential expression analysis. Data from the TCGA and GEO databases were normalized using R software (version 3.4.1; https://www.r-project.org/). Following data normalization, the present study performed a differential gene expression analysis using HCC samples with and without vascular invasion in the training set using the limma (14) package (version 3.34.7; https://bioconductor.org/packages/release/bioc/html/limma.html) of R software. The genes with false discovery rate (FDR) <0.05 and |log2 FC|>0.263 were selected and subsequently underwent a two-way hierarchical clustering analysis based on centered pearson correlation (15) algorithm using pheatmap package (16) (version 1.0.8) of R language (version 3.34.7). The results were presented in a heatmap.

Development of an SVM classifier. The present study initially performed a Cox regression analysis to investigate the associations between the identified DEGs and OS. From the significant DEGs with log-rank P<0.05, the present study then identified the optimal combination of feature genes using an RFE (17) algorithm in the caret (18) package (version 6.0-79; https://cran.r-project.org/web/packages/caret) of R language, which was then used to develop an SVM classifier using the SVM (19) function with a sigmoid kernel.

In both the training set and the validation set, the robustness of the established SVM classifier was evaluated using concordance index (C-index) (20), Brier score (21), log-rank P-value of cox-PH regression, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV) and area under receiver operating characteristic curve (AUROC). C-index and Brier score was calculated using the survcomp version 3.9 (22) package (http://www.bioconductor.org/packages/release/bioc/html/survcomp.html) of R language (version 3.4.1), which are two metrics for assessing accuracy. Kaplan-Meier estimate was applied to depict survival time using the survival package in R language. The Log-rank P-value for the difference in OS time between the two groups was calculated. AUROC ranged from 0.5 to 1, with a higher value implying better performance. Sensitivity, specificity, PPV and NPV of ROC curves were computed using pROC v1.15.3 (23) package of R language (https://cran.r-project.org/web/packages/pROC/index.html).

Development and validation of a prognostic scoring model. The present study further utilized the feature genes to fit a LASSO Cox-PH regression model (24) in order to determine the optimal panel of genes for prognosis using the penalized package (v0.9-51) of R language. Based on Cox-PH regression coefficients and expression levels of the identified optimal genes, a prognostic scoring model was built using the following formula:

\[ \text{Risk score} = \sum \text{coef}_{\text{DEGs}} \times \text{Exp}_{\text{DEGs}} \]

\( \text{coef}_{\text{DEGs}} \) represents Cox-PH regression coefficients of DEGs; \( \text{Exp}_{\text{DEGs}} \) represents expression levels of DEGs.

Risk score was calculated for each sample in the training set. Samples in the training set were then split into a high-risk group and a low-risk group according to median risk score (0.0663803). Kaplan-Meier survival curves were plotted for both risk groups using survival package (version 2.41-1) of R language, and OS of the two groups was compared by log-rank test. Similarly, samples in the validation set were divided into a high-risk group and a low-risk group using the median risk score of the validation set (0.132434) so as to test the prognostic ability of the prognostic scoring model in this set. The present study further validated the results by using SurvExpress, which is an online biomarker validation tool for cancer gene expression data (25). A total of four datasets, including GSE10143 (12), GSE10186 (26), TCGA-Liver-cancer and LIHC-TCGA-Liver HCC were included into SurvExpress.

![Figure 1. Kaplan-Meier curves of overall survival time of patients with or without vascular invasion of The Cancer Genome Atlas set. HR, hazard ratio.](image-url)
Stratified analysis. In both the high and low-risk groups of the training set, the present study investigated the associations between clinical factors and OS by performing a Cox regression analysis with the survival package in R language (version 2.41-1).

Table I. Uni-and multivariate Cox regression analysis of the training set.

<table>
<thead>
<tr>
<th>Clinical characteristics</th>
<th>TCGA (n=292)</th>
<th>Uni-variable cox</th>
<th>Multi-variable cox</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, years, mean ± SD</td>
<td>59.85±12.92</td>
<td>1.017 (0.999-1.035) 0.051</td>
<td>-</td>
</tr>
<tr>
<td>sex (male/female)</td>
<td>194/98</td>
<td>0.731 (0.474-1.127) 0.154</td>
<td>-</td>
</tr>
<tr>
<td>Pathological M (M0/M1/-) (28)</td>
<td>220/3/69</td>
<td>4.91 (1.523-15.84) 0.003</td>
<td>3.848 (1.089-13.588) 0.036</td>
</tr>
<tr>
<td>Pathological N (N0/N1/-)</td>
<td>210/2/80</td>
<td>1.602 (0.221-2.610) 0.638</td>
<td>-</td>
</tr>
<tr>
<td>Pathological T (T1/T2/T3/T4/-)</td>
<td>160/75/48/8/1</td>
<td>1.538 (1.23-1.923) &lt;0.001</td>
<td>0.607 (0.217-1.699) 0.342</td>
</tr>
<tr>
<td>Pathological stage (I/II/III/IV/-)</td>
<td>150/72/48/18</td>
<td>1.473 (1.153-1.881) 0.003</td>
<td>2.177 (0.797-5.944) 0.129</td>
</tr>
<tr>
<td>Histological grade (G1/G2/G3/G4/-)</td>
<td>36/141/101/12/2</td>
<td>1.19 (0.889-1.593) 0.243</td>
<td>-</td>
</tr>
<tr>
<td>Virus infection (HBV/HCV/Mixed/-)</td>
<td>50/10/35/197</td>
<td>1.167 (0.801-1.702) 0.420</td>
<td>-</td>
</tr>
<tr>
<td>Vascular invasion (yes/no)</td>
<td>102/190</td>
<td>1.353 (1.087-2.098) 0.009</td>
<td>1.678 (1.195-2.962) 0.037</td>
</tr>
<tr>
<td>Recurrence (yes/no/-)</td>
<td>119/156/17</td>
<td>1.343 (0.843-2.141) 0.213</td>
<td>-</td>
</tr>
<tr>
<td>Status (dead/alive)</td>
<td>87/205</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Overall survival time, months, mean ± SD</td>
<td>26.52±24.43</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

TCGA, The Cancer Genome Atlas; SD, standard deviation; M, metastasis; N, node; T, tumor; HBV, hepatitis B virus; HCV, hepatitis C virus; HR, hazard ratio; CI, confidence interval; -, information unavailable.

Figure 2. DEGs of patients with presence and absence of vascular invasion of the training set. (A) Volcano plot of 175 DEGs. Green spots represent DEGs; red horizontal dash line implies FDR <0.05; two red vertical dash lines indicate |logFC|>0.263. (B) Kernel density plot of log2 (FC) of 175 DEGs. (C) Heatmap for two-way hierarchical clustering of samples based on expression of DEGs. The red and green represent upregulated and downregulated genes, respectively. DEGs, differentially expressed genes; FDR, false discovery rate; FC, fold change.

Stratified analysis. In both the high and low-risk groups of the training set, the present study investigated the associations.
Functional analysis. The cases in the training set were divided into high- and low-risk groups according to the risk score of the gene signature. The present study then screened for DEGs in the two risk groups using a strict cut-off at FDR<0.05 and $|\log_{2} FC|>0.263$. The significant DEGs were selected for the pathway enrichment analysis using Gene Set Enrichment
Analysis (27) (GSEA, version 3.0; http://software.broadinstitute.org/gsea/index.jsp). P<0.05 was considered to indicate a statistically significant result.

Results

Vascular invasion is an independent predictor of prognosis. The present study performed uni- and multivariate Cox regression analyses in order to analyze the associations between clinical factors and OS in the training set using the survival package in R language. As presented in Table I, vascular invasion and pathological M stage (28) were identified as independent predictors of prognosis in the univariate and multivariate analyses (P<0.05). However, there were only three samples at pathological M1 stage, which was an insufficient amount for accurately assessing prognostic value of pathological M stage. Therefore, the present study classified all the samples of the training set into two groups according to vascular invasion. Patients without vascular invasion (n=190) had significantly better survival time compared with patients with vascular invasion (n=102; P=8.609x10⁻³; Fig. 1).

DEGs were screened between patients with and without vascular invasion. Following the removal of genes with a median expression level of 0, a total of 13,812 genes were inputted into the Limma package and among them, 175 significant DEGs in patients both with and without vascular invasion in the training set that satisfied the cut-off threshold (FDR<0.05 and |log2FC|>0.263) were identified (Table SI), consisting of 62 (35.43%) downregulated genes and 113 (64.57%) upregulated genes in the HCC samples with vascular invasion (Fig. 2A-C).

SVM analysis. Of the aforementioned 175 DEGs, 51 were significantly associated with OS (log-rank P<0.05) in the Cox regression analysis (Table III). For the purpose of obtaining the optimal feature genes for predicting vascular invasion in HCC, the present study utilized an SVM-RFE algorithm based on the 51 prognosis-associated genes. Maximal prediction accuracy (0.873) (Fig. 3A) and minimal root-mean-square error (0.1038) (Fig. 3B) were reached when using a 42-gene combination (Table II).

The SVM classifier was built with the 42-gene combination and its performance was assessed in both the training set and the validation set. A scatter plot and confusion matrix of the optimal feature genes for the training set or the validation set classified by the classifier are presented in Fig. 3C and D. Table III demonstrates that both sets generated high C-index scores (>0.75), low Brier scores (<0.1) and significant average log-rank P-values (2.97x10⁻⁸, 0.0264) in OS difference between the patients with and without vascular invasion (Fig. 4). AUROC of the two sets were 0.970 and 0.942, respectively (Table III; Fig. 4). The sensitivity, specificity, PPV and NPV values are presented in Table III. These results suggest that the SVM classifier was able to classify the samples effectively.

Prognostic model based on a 14-gene signature. The present study also used the 42 feature genes to create a LASSO Cox-PH regression model. When the maximal value of cross-validation likelihood (-498.517) was achieved, the optimal lambda value was 13.049, and the optimal panel of 14 genes was obtained (Table IV), including Wnt family member 1 (WNT1), crystallin α A (CRYAA), RAS like estrogen regulated growth inhibitor like (RERGL), hydroxysteroid 17-Beta dehydrogenase 13 (HSD17B13), scinderin (SCIN), premature ovarian failure (POF1B), erythropoietin (EPO), USH1 protein network component harmonin (USH1C), ADP-ribosyltransferase 5 (ART5).
Table III. Performances of the SVM classifier on the training and validation sets.

<table>
<thead>
<tr>
<th>Datasets</th>
<th>C-index</th>
<th>Brier score</th>
<th>Log-rank P-value</th>
<th>AUROC</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>PPV</th>
<th>NPV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training set (TCGA, N=292)</td>
<td>0.814</td>
<td>0.0394</td>
<td>&lt;0.0001</td>
<td>0.970</td>
<td>0.814</td>
<td>0.926</td>
<td>0.856</td>
<td>0.903</td>
</tr>
<tr>
<td>Validation set (GSE10141, N=62)</td>
<td>0.757</td>
<td>0.0884</td>
<td>0.0264</td>
<td>0.942</td>
<td>0.824</td>
<td>0.889</td>
<td>0.737</td>
<td>0.930</td>
</tr>
</tbody>
</table>

SVM, support vector machine; TCGA, The Cancer Genome Atlas; ROC, receiver operating characteristic curve; AUROC, area under receiver operating characteristic curve; PPV, positive predictive value; NPV, negative predictive value.
Table IV. Prognostic signature with 14 genes.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Coefficient</th>
<th>Hazard ratio (95% CI)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>WNT1</td>
<td>-0.2500</td>
<td>0.602 (0.459-0.789)</td>
<td>2.400x10^-04</td>
</tr>
<tr>
<td>CRYAA</td>
<td>-0.0002</td>
<td>0.108 (0.0092-0.493)</td>
<td>4.963x10^-02</td>
</tr>
<tr>
<td>RERGL</td>
<td>-0.0263</td>
<td>0.463 (0.244-0.854)</td>
<td>4.533x10^-02</td>
</tr>
<tr>
<td>HSD17B13</td>
<td>-0.0153</td>
<td>0.586 (0.176-0.906)</td>
<td>4.688x10^-02</td>
</tr>
<tr>
<td>SCIN</td>
<td>0.0852</td>
<td>1.115 (1.086-1.267)</td>
<td>4.939x10^-02</td>
</tr>
<tr>
<td>POIF1B</td>
<td>0.0756</td>
<td>1.085 (1.001-1.178)</td>
<td>1.513x10^-02</td>
</tr>
<tr>
<td>EPO</td>
<td>0.0616</td>
<td>1.068 (1.013-1.152)</td>
<td>4.897x10^-02</td>
</tr>
<tr>
<td>USH1C</td>
<td>0.0106</td>
<td>1.043 (1.001-1.171)</td>
<td>4.897x10^-02</td>
</tr>
<tr>
<td>ART5</td>
<td>0.0134</td>
<td>1.047 (1.035-1.171)</td>
<td>4.231x10^-02</td>
</tr>
<tr>
<td>MMP12</td>
<td>0.0236</td>
<td>1.051 (1.048-1.165)</td>
<td>3.410x10^-02</td>
</tr>
<tr>
<td>TRIM54</td>
<td>0.0454</td>
<td>1.059 (1.028-1.194)</td>
<td>2.392x10^-02</td>
</tr>
<tr>
<td>SLC35F3</td>
<td>0.0124</td>
<td>1.057 (1.029-1.203)</td>
<td>2.394x10^-02</td>
</tr>
<tr>
<td>HOXD10</td>
<td>0.0101</td>
<td>1.044 (1.127-1.924)</td>
<td>0.069x10^-03</td>
</tr>
<tr>
<td>PPP2R2C</td>
<td>0.0047</td>
<td>1.004 (1.002-1.085)</td>
<td>4.926x10^-01</td>
</tr>
</tbody>
</table>

CI, confidence interval.

Identification and pathway analysis of DEGs between the two risk groups in the training set. In the training set, 599 upregulated genes and 163 downregulated genes were identified in the high-risk group compared with the low-risk group. These genes were significantly involved in pathways of ‘retinol metabolism’, ‘drug metabolism other enzymes’, ‘drug metabolism cytochrome P450’, ‘peroxisome proliferator-activated receptor (PPAR) signaling pathway’, ‘primary bile acid biosynthesis’, ‘steroid hormone biosynthesis’ and ‘histidine metabolism pathways’ (Table VI).

Discussion

HCC is an aggressive malignancy characterized by high incidence rates of recurrence and metastasis (29). Vascular invasion is an unfavorable prognostic factor for patients with HCC (30). Therefore, unraveling the underlying molecular landscape of vascular invasion is of significance for the prognosis of HCC. In the present study, a total of 175 DEGs were identified between patients with the presence and absence of vascular invasion. An SVM classifier was built that consisted of 42 feature genes by implementing an RFE-SVM algorithm. In both the training and validation sets, the classifier had high C-index values, low Brier scores and significant log-rank P-values, indicating good performances in separating patients with vascular invasion from patients without vascular invasion. Furthermore, through using a LASSO Cox-PH model, a 14-gene prognostic signature was obtained and consequently, a prognostic scoring model was established. The 14-gene signature was able to predict those patients with HCC that would have a shorter survival time, as evidenced by the result that OS time was significantly different between the predicted high-risk patients and the predicted low-risk patients. T prognostic performance of the 14-gene signature was successfully confirmed in the validation set.

The 14-gene prognostic combination included WNT1, CRYAA, RERGL, HSD17B13, SCIN, POIF1B, EPO, USH1C, ART5, MMP12, TRIM54, SLC35F3, HOXD10 and PPP2R2C. Proto-oncogene protein Wnt-1 encoded by the WNT1 gene has been demonstrated as upregulated in HCC, acting as a direct target of miR-122 (31). RERGL is a member of the RAS superfamily of GTPases that participates in regulating several biological processes, such as cell proliferation, differentiation and apoptosis (32). There was one HSD17B13 protein, namely 17β-HSD type 13, that was downregulated in HCC (33). There is evidence to suggest that HSD17B13 suppresses HCC progression by delaying the G1/S phase transition of HCC cells (34). Furthermore, HSD17B13 is a novel liver-specific protein associated with lipid droplet, and may be a promising biomarker of liver cancer (35). SCIN encodes scinderin, which is an actin-severing protein of the gelsolin superfamily. It acts as a regulator of HCC cell apoptosis and growth, and has been identified as a transcriptional target of tumor suppressor factor breast cancer metastasis-suppressor 1 (36). It has long been established that the EPO/EPO-receptor plays an important role in angiogenesis and progression of HCC (37). EPO protein expression is positively correlated with vasculogenic mimicry in HCC, and has been identified as an independent predictor.
of prognosis in patients with HCC (38). Furthermore, *EPO* is upregulated in HCC and could promote HCC cell proliferation through translocation of its specific receptor induced by hypoxia (39). *MMP12* belongs to the MMP family implicated in the degradation of the extracellular matrix. It is upregulated in HCC and is an independent predictive factor for OS in patients with HCC (40,41). *TRIM54* is a member of the TRIM protein family. Several members in the TRIM family have been reported to be involved in biological processes, such as cell proliferation, differentiation and apoptosis, and may play a role in cancer initiation and progression (42). However, to the best of our knowledge *TRIM54* has not been reported previously. *HOXD10*, a member of the Abd-B homeobox family, exhibits decreased expression levels in HCC and serves as a tumor-suppressor gene through prohibiting extracellular signal-regulated kinase signaling (43). *PPP2R2C* encodes serine/threonine-protein phosphatase 2A 55 kDa regulatory subunit Bγ isoform, and has been identified as upregulated in HCC (44). To the best of our knowledge, there are little studies that focus on the function of *CRYAA*, *RERGL*, *POFI1B*, *POFI1B*, *USH1C*, *TRIM54* and *SLC35F3* in HCC. The results of the present study indicate that the 14 vascular invasion-associated genes may be prognostic biomarkers of HCC.

Another aim of the present study was identifying the potential roles of DEGs between the high- and low-risk groups of the training set. There were 762 DEGs between the two
risk groups, which were significantly involved in a number of signaling pathways, such as 'retinol metabolism', 'drug metabolism cytochrome P450', and 'PPAR signaling pathway'. The association between retinol metabolism and HCC has been demonstrated previously and a synthetic retinoid has been indicated to prevent HCC recurrence (45). Drug-metabolizing
cytochrome P450 enzyme activities are severely disrupted in HCC (46). The PPAR signaling pathway plays a part in tumorigenesis and tumor progression via different metabolic pathways: Glycolysis/gluconeogenesis, lipid, glycerolipid and glycerophospholipid metabolism, protein synthesis and degradation and purine metabolism (47). These findings reveal the critical roles of these pathways in HCC.

There are some limitations in the present study; though the 14-gene prognostic signature has been validated by an independent dataset, the expression levels of these 14 genes have not been confirmed by individual gene expression experiments.

In summary, using TCGA data, the present study defined a classifier of 42 feature genes for classification of patients with HCC with and without vascular invasion, and identified a vascular invasion-associated 14-gene prognostic signature for HCC. Several genes and pathways have been revealed to be critical for HCC. These results further the current knowledge on the molecular mechanisms underlying HCC and may aid in the development of personalized treatment for patients with HCC. Large-scale studies are required in order to further validate the results of the present study.

Acknowledgements
Not applicable.

Funding
No funding was received.

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

Table V. Results of Cox regression analysis for the high- and low-risk groups of The Cancer Genome Atlas set.

<table>
<thead>
<tr>
<th>Clinical characteristics</th>
<th>Low risk group</th>
<th></th>
<th>High risk group</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HR (95% CI)</td>
<td>P-value</td>
<td>HR (95% CI)</td>
<td>P-value</td>
</tr>
<tr>
<td>Age, years, mean ± SD</td>
<td>1.018 (0.986-1.052)</td>
<td>0.273</td>
<td>1.012 (0.992-1.032)</td>
<td>0.257</td>
</tr>
<tr>
<td>Sex (male/female)</td>
<td>0.568 (0.246-1.308)</td>
<td>0.178</td>
<td>1.129 (0.676-1.886)</td>
<td>0.643</td>
</tr>
<tr>
<td>Pathological M (M0/M1/-) (28)</td>
<td>8.721 (1.090-69.77)</td>
<td>0.138</td>
<td>3.227 (0.770-13.520)</td>
<td>0.090</td>
</tr>
<tr>
<td>Pathological N (N0/N1/-)</td>
<td>3.01 (1.052-33.22)</td>
<td>0.763</td>
<td>1.429 (0.195-10.490)</td>
<td>0.724</td>
</tr>
<tr>
<td>Pathological T (T1/T2/T3/T4/-)</td>
<td>1.527 (0.955-2.443)</td>
<td>0.072</td>
<td>1.240 (0.914-1.681)</td>
<td>0.165</td>
</tr>
<tr>
<td>Pathological stage (I/II/III/IV/-)</td>
<td>1.346 (0.805-2.252)</td>
<td>0.254</td>
<td>1.215 (0.885-1.668)</td>
<td>0.228</td>
</tr>
<tr>
<td>Histological grade (G1/G2/G3/G4/-)</td>
<td>1.115 (0.630-1.971)</td>
<td>0.709</td>
<td>0.916 (0.642-1.307)</td>
<td>0.629</td>
</tr>
<tr>
<td>Virus infection (HBV/HCV/Mixed/-)</td>
<td>2.333 (1.962-5.655)</td>
<td>0.038</td>
<td>0.932 (0.613-1.416)</td>
<td>0.741</td>
</tr>
<tr>
<td>Vascular invasion (yes/no)</td>
<td>2.478 (1.044-5.885)</td>
<td>0.034</td>
<td>3.446 (1.913-6.209)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Recurrence (yes/no/-)</td>
<td>1.569 (0.670-3.672)</td>
<td>0.296</td>
<td>0.924 (0.526-1.623)</td>
<td>0.783</td>
</tr>
</tbody>
</table>

M, metastasis; T, tumor; N, node; HBV, hepatitis B virus; HCV, hepatitis C virus; HR, hazard ratio; CI, confidence interval.

Table VI. Significant signaling pathways.

<table>
<thead>
<tr>
<th>Pathway</th>
<th>ES</th>
<th>NES</th>
<th>Normal</th>
<th>FDR</th>
<th>Count</th>
<th>Gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>Retinol metabolism</td>
<td>-0.7987</td>
<td>-2.3043</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>CYP4A22, CYP26A1, CYP3A43, CYP2A7, CYP2A6, CYP2A13</td>
</tr>
<tr>
<td>Drug metabolism other enzymes</td>
<td>-0.9022</td>
<td>-2.2834</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>CYP3A43, CYP2A7, CYP2A6, CYP2A13</td>
</tr>
<tr>
<td>Drug metabolism cytochrome P450</td>
<td>-0.9011</td>
<td>-2.0480</td>
<td>0</td>
<td>0.047</td>
<td>4</td>
<td>CYP3A43, CYP2A7, CYP2A6, CYP2A13</td>
</tr>
<tr>
<td>PPAR signaling pathway</td>
<td>-0.7106</td>
<td>-1.9354</td>
<td>0.0026</td>
<td>0.0121</td>
<td>3</td>
<td>CYP4A22, CYP8B1, ACADL</td>
</tr>
<tr>
<td>Primary bile acid biosynthesis</td>
<td>-0.9631</td>
<td>-1.9038</td>
<td>0</td>
<td>0.0124</td>
<td>3</td>
<td>CYP8B1, AKR1D1, CYP7A1</td>
</tr>
<tr>
<td>Steroid hormone biosynthesis</td>
<td>-0.5989</td>
<td>-1.8162</td>
<td>0.0084</td>
<td>0.0188</td>
<td>6</td>
<td>AKR1D1, CYP7A1, HSD3B2, HSD3B1, CYP11A1, CYP3A43</td>
</tr>
<tr>
<td>Histidine metabolism</td>
<td>-0.8709</td>
<td>-1.6875</td>
<td>0.01</td>
<td>0.0468</td>
<td>3</td>
<td>HDC, CNDP1, UROC1</td>
</tr>
</tbody>
</table>

ES, enrichment score; NES, normalized enrichment score; count of genes, the number of genes enriched in a pathway; FDR, false discovery rate.
Authors' contributions

BY performed the data analyses and wrote the manuscript. CT and YT contributed significantly towards the data analyses and revised the manuscript. ZZ conceived and designed the study. All authors read and approved the final published version of the manuscript.

Ethics approval and consent to participate

In the original article of the datasets, the trials were approved by the local institutional review boards of all participating centers, and informed consent was obtained from all patients.

Patient consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References


This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) License.