Comprehensive Analysis of the Expression of Cell Adhesion Molecules Genes in Hepatocellular Carcinoma and their Prognosis, and Biological Significance

Jia Wang¹,†, Yuting Hu¹,†, Kunpeng Zhao¹, Jian Fan², Jian Zhu², Qingya Li³, Qilong Chen¹, Yiyu Lu¹,*

¹Institute of Interdisciplinary Integrative Medicine Research, Shanghai University of Traditional Chinese Medicine, 201203 Shanghai, China
²Qidong Liver Cancer Institute, Qidong People’s Hospital, the affiliated Qidong Hospital of Nantong University, 226200 Nantong, Jiangsu, China
³Department of Pathogenic Organism Biology, Henan University of Chinese Medicine, 450046 Zhengzhou, Henan, China
*Correspondence: yiyulu@shutcm.edu.cn (Yiyu Lu)
†These authors contributed equally.

Abstract

Background: Collagen-related cell adhesion molecules (CAMs) are a major component of the extracellular matrix (ECM) and often accumulate in the liver during chronic liver disease or hepatocellular carcinoma (HCC). In this study we identified several promising collagens related to CAMs that may be of clinical use for the diagnosis and prognosis of HCC. Methods: We obtained multi-omics data including RNA sequencing (RNA-Seq) data, microarray data, proteomic data from the TCGA, GEO databases, GTEx, and NODE. Bioinformatics analyses were then performed to investigate correlations between the expression patterns of significant genes and HCC. Tumor tissue and para-cancerous tissue samples from HCC patients were also used to validate the results using RT-PCR. Results: A literature research and LASSO-COX analysis identified three significant collagen-related CAM genes: SERPINH1, DCN, and ITGB1. Immunohistochemistry images in the Human Protein Atlas Project database showed that SERPINH1 and ITGB1 proteins were moderately or highly expressed in HCC tumor tissues compared to para-cancerous tissue, whereas DCN expression was lower in HCC tumor tissue. These results were validated by RT-PCR. Low- and high-risk groups of HCC patients were distinguished by the logistic panel in the TCGA database. These showed significantly different prognosis, clinicopathological features, and immune cell infiltration. Logistic regression was used to construct predictive models based on the individual expression levels of DCN, SERPINH1, and ITGB1. These showed highly accurate diagnostic ability (AUC = 0.987). Conclusions: The current findings suggest that the collagen-related CAMs DCN, SERPINH1, and ITGB1 may be potential therapeutic targets in HCC. Logistic panels of DCN, SERPINH1 and ITGB1 could serve as non-invasive and effective diagnostic biomarkers for HCC. Clinical Trial registration: Identifier: NCT03189992. Registered on June 4, 2017. Retrospectively registered (https://clinicaltrials.gov/).

Keywords: hepatocellular carcinoma; diagnosis; prognosis; cell adhesion molecules; bioinformatics analysis

1. Introduction

Liver cancer was the fourth leading cause of cancer-related mortality in 2018 and 2019 [1], and the sixth most commonly diagnosed cancer worldwide [2]. Approximately 85%–90% of primary liver malignancies are hepatocellular carcinoma (HCC) [3].

Hepatic fibrosis is one of the major causes of HCC. Chronic liver disease often results in excessive buildup of extracellular matrix (ECM), a condition that is closely associated with the accumulation of collagen. The ECM is a three-dimensional structure secreted by cells and enveloping the extracellular space within tissues. It consists of a non-cellular meshwork of collagen, proteoglycans/glycosaminoglycans, elastin, fibronectin, and laminin, with the major protein being collagen. In healthy tissue, the production and arrangement of collagen is tightly controlled through a precise equilibrium between matrix metalloproteinases (MMPs) and their inhibitors, together with enzymes such as lysyl oxidases [4]. As tumors grow and advance, cancer cells release substantial quantities of MMPs, which subsequently alter and break down the basement membrane within ECM. This remodeling triggers a breakdown of the normal ECM structure, and a complex interplay of pro- and anti-tumor signals from the degradation products [5]. Tumor cell proliferation is known to alter collagen functions, which are dependent on the individual collagen levels [6,7]. Cancer-associated fibroblasts (CAFs) disrupt the normal regulation of collagen turnover during tumorigenesis, thus resulting in tumor fibrosis, also known as desmoplasia. This manifests as an over-abundance of collagen deposition in the vicinity of the tumor [8] leading to stiffening of the tissue. The hardened ECM is associated with increased tumor aggressiveness and correlates with an increased propensity for metastasis and worse patient outcome [9]. Tumor fibrosis thus affects the surround-
ing tumor cells by causing changes in cell proliferation, differentiation, migration, and metastasis. In this way, excess collagen deposition directly affects some of the hallmarks of cancer [10]. Emerging evidence also indicates that fibroblast-derived stromal collagens are strongly correlated with poorer prognosis of cancer patients [11–15].

There is increasing interest in cancer biomarkers such as cytokines, CAFs, collagens and cell adhesion molecules (CAMs). Together, these create a unique ECM composition in different tissues. Despite numerous studies on possible biomarkers for HCC, little is still known regarding the diagnostic or prognostic value of collagen-related CAMs. The function of CAMs in HCC and their possible clinical application therefore require further investigation. Interactions between cancer cells and the tumor microenvironment (TME) promote tumorigenesis through CAMs. These are a group of cell surface molecules that promote intercellular communication and intercellular matrix binding. CAMs have an adhesive function, but also initiate intracellular signaling pathways that impact cell survival, proliferation, metastasis, and epithelial mesenchymal transition (EMT), while potentially also impacting drug resistance in tumor cells [16]. CAMs are expressed at increased levels in many solid tumor types. These include ITGA9 in various cancer types [17], PTGFRN in glioblastoma [18], and ALCAM in epithelial-derived cancers [19]. Our previous study focused on non-invasive biomarkers of HCC [20]. We built an edge panel (edge-based biomarker) as an accurate and robust predictive model for HCC [21]. In this model, COL5A1 was suggested to have an oncogenic function since it can stimulate cell proliferation and invasion, as well as enhancing viability. Greater knowledge of collagen and of related CAM genes should eventually lead to better cancer diagnosis, inhibition of fibrosis, and reduced tumor drug resistance [22,23]. As a highly expressed molecular group, CAMs have good potential as cancer biomarkers. Therefore, in the present study we systematically analyzed the expression level of several collagen-associated CAMs in HCC, together with their diagnostic and prognostic values.

2. Materials and Methods

2.1 Gene Screening, Literature Search, and Study Selection

To date, 28 different collagens have been identified. A literature search was conducted to identify the most extensively researched collagen genes in the field of cancer studies. Search queries for each known “collagen gene” were applied in conjunction with “cancer” to obtain studies published in Pubmed by June 30, 2022. “Pubmed.mineR” (version 1.0.19) packages in R (Ross Ihaka, Auckland, New Zealand) was applied to identify the collagen genes most frequently associated with cancer.

2.2 Network Construction

The College genes’ protein-protein interaction (PPI) network was visualized, and the central hub gene determined using Cytoscape 3.8.2 software (UC San Diego, San Diego, CA, USA). Transcription factor (TF) regulation relationship data was also downloaded from the RegNetwork database (https://regnetworkweb.org/).

2.3 LASSO-COX Analysis

Cox regression analysis was performed to assess the prognostic value of candidate genes, and to examine their association with overall survival (OS) in the TCGA database. Diagnostic markers were screened using the least absolute shrinkage and selection operator (LASSO)-Cox regression model implemented in the “glmnet” package. The penalty parameter (λ) was chosen based on the minimum standard. Subsequently, a prognostic signature panel model was constructed using logistic regression. The risk score for each HCC patient within the TCGA and ICGC cohorts was computed, after which they were categorized into high-risk group and low-risk groups according to the median risk score. The “survival” and “survminer” packages in R were applied to calculate the OS of patients in the high- and low-risk subgroups, with a p-value of <0.05 considered statistically significant.

2.4 Gene Mutation Analysis

The mutation analysis landscape of the three signature genes was conducted for high and low-risk HCC patients in the TCGA using the “Maftools” package in R version 4.1.0.

2.5 Analysis of Tumor-Infiltrating Immune Cells

Differences in the immune cell microenvironment between high- and low-risk groups were assessed using CIBERSORT software (Stanford University, Stanford, CA, USA). Twenty-two types of tumor-infiltrating immune cells were evaluated. Each risk score group’s composition of 22 immune infiltrates was calculated.

2.6 GEPIA, HPA, and TIMER Datasets

GEPIA 2 (available at http://gepia.cancer-pku.cn/index.html) was utilized to validate gene expression [15,16]. This web-based tool offers key interactive and customizable functions derived from the TCGA and GTEx (Genotype-Tissue Expression) datasets. Gene expression profiles were analyzed from the immunohistochemistry results available in the Human Protein Atlas Project (HPAP) database (HPA, https://www.proteinatlas.org/) [24]. TIMER2.0, accessible at https://cistrome.shinyapps.io/timer/ [25], was employed to investigate correlations between signature gene expression and immune cell types within the TME. Spearman’s rho correlation values depicting the relationship between the expression of specific genes and different immune cells were visualized in a heatmap.
2.7 Kaplan-Meier Plotter

The Kaplan-Meier Plotter (www.kmplot.com) was utilized to investigate the prognostic values of 8 collagen genes (https://kmplot.com/analysis/index.php?p=service&cancer=liver_rnaseq). To assess associations with OS, HCC patient samples were first divided into two groups according to the median expression value (i.e., high- and low-expression). Kaplan-Meier survival analysis was performed to calculate the hazard ratio (HR), 95% confidence interval (CI), and log-rank $p$-value.

2.8 Patient Clinical Information and Collection of Tissue And Serum

After obtaining written informed consent, the tumor tissue and adjacent normal tissue were collected prospectively from 31 HCC patients who attended the Shuguang Hospital, Shanghai University of TCM. Comprehensive clinical information for these participants is presented in Table 1. Tissue samples were collected before any treatment or surgery. Tissue samples underwent centrifugation at 3000 g for 15 min at 4 °C. All samples were stored at –80 °C before subsequent processing.

Table 1. Clinical characteristic of HCC patients in independent cohort.

<table>
<thead>
<tr>
<th>Group</th>
<th>HCC (n = 31)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>55.28 ± 6.95</td>
</tr>
<tr>
<td>Gender (M/F)</td>
<td>19/2</td>
</tr>
<tr>
<td>TBIL (µmol/L)</td>
<td>22.37 ± 11.37</td>
</tr>
<tr>
<td>ALB (g/L)</td>
<td>39.19 ± 3.52</td>
</tr>
<tr>
<td>ALT (IU/L)</td>
<td>37.71 ± 25.64</td>
</tr>
<tr>
<td>AST (IU/L)</td>
<td>33.9 ± 15.81</td>
</tr>
<tr>
<td>BUN (umol/L)</td>
<td>4.83 ± 0.97</td>
</tr>
<tr>
<td>UA (umol/L)</td>
<td>286.28 ± 56.92</td>
</tr>
<tr>
<td>GLOB (g/L)</td>
<td>17.86 ± 2.81</td>
</tr>
<tr>
<td>CEA (ng/mL)</td>
<td>3.45 ± 1.94</td>
</tr>
<tr>
<td>CA199 (U/mL)</td>
<td>17.15 ± 14.16</td>
</tr>
<tr>
<td>AFP (µg/L) (n)</td>
<td></td>
</tr>
<tr>
<td>&lt;20</td>
<td>9</td>
</tr>
<tr>
<td>20–400</td>
<td>10</td>
</tr>
<tr>
<td>&gt;400</td>
<td>12</td>
</tr>
<tr>
<td>Clinical stage(n)</td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>6</td>
</tr>
<tr>
<td>II</td>
<td>18</td>
</tr>
<tr>
<td>III</td>
<td>7</td>
</tr>
</tbody>
</table>

TBIL, total bilirubin; ALB, albumin; ALT, alanine amino transferase; AST, aspartate transaminase; BUN, blood urea nitrogen; UA, uric acid; GLOB, Globulin; CEA, carcinoembryonic antigen; CA199, carbohydrate antigen 199; AFP, alpha-feto protein.

2.9 Functional Analysis

The ClusterProfiler package in R [26,27] was used to conduct the functional analysis of selected genes, including Gene Ontology (GO) analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis.

2.10 RNA Extraction and RT-qPCR

Extraction of total RNA and RT-qPCR were carried out as previously described [21]. Primer sequences are presented in Supplementary Table 1. Gene expression levels were quantified compared to the expression of ACTB.

2.11 Transriptome and Proteome Data Acquisition

Transcriptome data was obtained from GEO databases. RNA-seq data for HCC was obtained by downloading from the TCGA website (https://portal.gdc.cancer.gov/). Microarray data (GSE36376, GSE112790) [14,28] was obtained from the NCBI GEO database. In addition to the validation data from GEO, proteomics data was also acquired from the study by Gao et al. [29]. This information is accessible for viewing on NODE (https://www.biosino.org/node) by entering the accession code (OEP000321), or via the following URL: https://www.biosino.org/node/project/detail/OEP000321. The investigation involves a study of hepatitis B virus (HBV)-related HCC using paired tumor and adjacent liver tissues from 159 patients.

2.12 Statistical Analysis

Differences in gene expression levels between groups were compared using the Mann-Whitney U test. The diagnostic marker was developed using a stepwise logistic regression model. A $p$-value $< 0.05$ was considered statistically significant.

3. Results

3.1 Literature Screening for the Collagen Gene Family

To date, 28 collagen types encoded by 44 genes have been identified. To identify the most well-studied of these in cancer research, a literature search was conducted as described in the Methods (Supplementary Fig. 1A). A total of 2175 records were retrieved from Pubmed (https://pubmed.ncbi.nlm.nih.gov/) by scanning the title and abstract, with 1053 relevant articles retained for further evaluation after the exclusion of duplicate articles. Of the 28 known collagen genes, the most frequently studied in cancer research were COL1A1, COL1A2, COL3A1, COL4A1, COL4A2, COL5A1, COL6A1, and COL6A3 (Supplementary Fig. 1B).

To investigate the role of frequently altered neighboring genes associated with these 8 collagen genes, we constructed a collagen regulation network. This merged the PPI and TF networks and identified the hub genes. The collagen regulation network was found to consist of 252
nodes and 681 edges (Fig. 1A). First-neighbor genes related to the 8 collagen genes were extracted to construct a hub gene network (Fig. 1B). Fibril-organization-related genes, including SPARC, SERPINH1, DCN, PCOLCE, LUM, and ITGB1, were found to be closely associated with the 8 collagen genes. The GO and KEGG databases were used to conduct functional analyses of the 8 collagen genes in the regulation network. Most of the identified pathways were associated with CAMs, while some were also related to ECM-receptor interaction (Fig. 1C,D). These pathways are closely involved with the tumorigenesis and pathogenesis of HCC [11,30].

3.2 Construction of Prognostic Models Based on Least Absolute Shrinkage and Selection Operator (LASSO) Model and Survival Analysis

Fourteen genes (8 collagen genes and 6 neighbor genes) underwent LASSO-Cox regression analysis. The model included 10-time cross-validation for tuning parameter selection, and resulted in the identification of three CAM genes (Fig. 2A,B). The prognostic signature for these genes (DCN, SERPINH1, ITGB1) was constructed using logistic regression. The resulting risk score panel was: -4.647 – 0.391 × \(\text{DCN} \) + 0.989 × \(\text{SERPINH1} \) + 0.137 × \(\text{ITGB1} \). Kaplan–Meier survival curves for HCC patients within the TCGA database revealed significantly longer OS for patients with a low-risk score compared to those with a high-risk score (Fig. 2C, \( p < 0.001 \)). The Kaplan-Meier and log-rank test analyses also indicated significant correlations between elevated levels of \(\text{SERPINH1} \) and \(\text{ITGB1} \) mRNA and lower survival rates (\( p = 0.013 \) and 0.022, respectively, Fig. 2D,E). Decreased levels of \(\text{DCN} \) mRNA were associated with significantly longer OS (\( p = 0.02 \), Fig. 2F).

3.3 Correlation between High- and Low- Risk Score Groups and Tumor Mutation Burden

Genetic variations were compared between the high- and low-risk score groups defined by the three signature genes. The high-risk score group showed significantly more frequent mutations of the 20 top mutated genes (Fig. 3). The “maftool” package was utilized to compute somatic mutation rates and to show the top 20 mutated driver genes. Gene mutations occurred more frequently in the high-risk score group. The most significant difference between the two groups was for \(\text{TP53} \) mutations (\( p = 3.7 \times 10^{-11} \)). This could indicate that patients with a high-risk score had a greater likelihood of amplifications and more DNA replication errors. However, the \(\text{TP53} \) mutation frequency (40%) was slightly higher in the low-risk score group (Fig. 3).
3.4 Differences in Immune Infiltration Level between High- and Low-Risk Groups, and TIMER Analysis

The infiltration of memory B cells \( (p < 0.001) \), plasma cells \( (p < 0.05) \), T cells follicular helper \( (p < 0.01) \), T regulatory cells (Tregs) \( (p < 0.01) \), macrophages M0 \( (p < 0.001) \) and neutrophils \( (p < 0.05) \) was significantly higher in the high-risk group from the TCGA dataset compared with the low-risk group. In contrast, the infiltration of T cells CD8 \( (p < 0.001) \), T cells CD4 memory naive \( (p < 0.05) \), T cell gamma delta \( (p < 0.05) \), NK cell resting \( (p < 0.001) \), monocytes \( (p < 0.001) \), macrophages M1 \( (p < 0.05) \), and resting mast cells \( (p < 0.01) \) was significantly lower (Fig. 4A).

We next employed the TIMER tool to investigate potential associations between the three signature genes and the inflammatory response. The three genes linked to changes in ECM and collagen formation showed negative correlations with the abundance of T CD8+ cells, Treg cells, mast cells, NK cell activated cells, and NK resting cells (Fig. 4B). However, significant positive correlations were found with the presence of B cells, T CD4+ cells, macrophages, neutrophils, and dendritic cells. Interestingly, DCN was strongly correlated with CAFs \( (R = 0.67, p < 0.001) \), which is a predictor of poor prognosis in HCC.

3.5 Gene and Protein Expression Levels of the Signature Genes in HCC from the TCGA Dataset and HPA (Human Protein Atlas)

The mRNA expression levels of the three signature genes were compared between tumor tissue and adjacent normal tissue in HCC from the GEPIA dataset. Of the 8 collagen genes, the mRNA levels for SERPINH1 and ITGB1 were notably elevated in HCC tumor compared to normal tissue (Fig. 5A,B). In contrast, the mRNA level of DCN in HCC tumor tissue was notably lower than in normal tissue.
Hepatic expression of the signature genes (SERPINH1, DCN, and ITGB1) was further examined using immunohistochemistry results from the Human Protein Atlas Project (HPAP) dataset. This analysis confirmed that SERPINH1 and ITGB1 showed moderate to high expression levels in tumor tissue compared to normal tissue. In contrast, DCN showed moderately higher expression levels in normal tissue compared to tumor tissue, consistent with the mRNA result (Fig. 5C).

3.6 Validation of Hepatic Expression of Signature Genes in Transcriptome and Proteome Datasets

Fig. 6A,B shows the expression levels of SERPINH1, DCN, and ITGB1 in tumor and normal tissues from the GSE36376 and GSE112790 cohorts. Similar patterns of expression for the three signature genes were observed in both datasets (Fig. 6A,B).

In addition, we validated the results and tested the prognostic value of the signature gene score panel using an independent proteomics dataset generated by Gao et al. [29] and deposited in the National Omics Data Encyclopedia (NODE) database. The study by Gao et al. [29] focuses on HBV-related HCC and includes paired tumor and non-tumor liver tissues from 159 patients. The signature genes were able to separate the non-tumor and HCC groups (Fig. 6C,D). Receiver operator characteristic (ROC) analysis further revealed the high efficacy of the signature gene panel in differentiating the non-tumor and HCC groups (Fig. 6C). We also developed and tested predictive models based on SERPINH1, DCN, and ITGB1. The logistic panel showed excellent diagnostic performance for differentiating between non-tumor and HCC samples, with an area under the curve (AUC) of 0.987. This result suggests the logistic panel identified here may have broad application in the clinical diagnosis of HCC (Fig. 6D). The three signature genes are therefore promising independent diagnostic markers for patients with HCC.

3.7 RT-PCR Validation of Collagen and Hub Genes

To explore the robustness of the above signatures for HCC diagnosis, we examined the mRNA expression profiles of core genes using in samples from an independent cohort. This comprised 31 HCC patients from which tumor tissue and adjacent normal tissues were collected and evaluated (Table 1). The results showed significant up-regulation of SERPINH1 expression ($p = 0.035$) and ITGB1 expression ($p = 0.006$) in HCC tumor tissue. In contrast, DCN expression was notably decreased ($p = 0.025$) in HCC tumor tissues compared to normal tissues (Fig. 7).

4. Discussion

TME formation involves interactions between host and cancer cells that are mediated through CAMs [30]. Alterations in the collagen content within the TME are intricately associated with tumor onset and progression, mainly through changes in the level of collagen expression and its density, direction, length, and cross-linking [31]. These
changes can have major effects on the invasive and metastasis properties of tumors. A comprehensive understanding of the relationship between collagen and tumors is therefore required to improve the prevention and treatment of cancer [32].

Collagen fibrils form the core of ECM molecular organization and the cellular microenvironment. Changes in collagen fibrils affect the adhesion and migration of cancer cells [33]. Alterations in the composition of ECM components, such as collagens, alter their interaction with CAMs and affect various cell functions such as growth, migration, and gene expression [34]. Both collagen and CAMs therefore play crucial roles in the development and progression of HCC. However, the diagnostic and prognostic significance of collagen-related CAMs in HCC still requires extensive research.

Collagen plays an important role in HCC, and the 8 collagen genes identified in the current literature search are

---

**Fig. 4. Different immune cell infiltration levels in high- and low-risk groups from the TCGA dataset.** (A) Differences in immune infiltration between the high- and low-risk groups. (B) Spearman’s rho analysis shows correlations between the three signature genes and 11 primary immune cell types within the tumor microenvironment (TME). ns, \( p > 0.05; \) * \( p < 0.05; \) ** \( p < 0.01; \) *** \( p < 0.001. \)
Fig. 5. Expression of the signature genes in HCC from the TCGA dataset (mRNA, left panel) and from the HPA dataset (protein analysis of tumor and normal tissue by immunohistochemistry, right panel) (Human Protein Atlas, http://www.proteinatlas.org/).

(A) Expression of SERPINH1 in TCGA and HPA. (B) Expression of ITGB1 in TCGA and HPA. (C) Expression of DCN in TCGA and HPA. The images from HPA database are available from version 23.0.proteinatlas.org. TPM, transcripts per million; LIHC, liver hepatocellular carcinoma.

known to be strongly associated with HCC. Type I collagen is the most prevalent collagen type within the body [35] and has been shown to affect the invasive behavior of tumor cells, leading to metastasis [36,37]. COL1A1 and COL1A2 are the main components of ECM and are involved in ECM remodeling, tumor cell adhesion, cell migration, and vascular development [38]. Similar to type I collagen, high expression of COL3A1 has been reported in lung [39] and ovarian cancers [40]. Moreover, COL3A1 participates in the invasion and metastasis of glioblastoma cells [41]. Type V collagen is expressed along with types I and III collagen, but is a less abundant fibrillar collagen. Huang et al. [42] reported that ablation of α3(V) in a mouse mammary tumor model (Col5a3−/−) impedes cancer progression by reducing the proliferative ability of tumor cells. In our previous study, COL5A1 was a member of the edge panel of biomarkers for HCC [21]. COL5A1 expression was found to be elevated in cirrhosis compared to chronic hepatitis B, suggesting it may be important during the onset and activation of liver fibrosis. The main function of type IV collagen is the formation of networks. COL4A1 and COL4A2 encode the alpha-1 and alpha-2 chains of collagen IV, respectively,
Fig. 6. Validation of collagen expression with proteomic data, and diagnostic assessment of the collagen and hub genes. (A) Expression of the three signature genes in the GSE36376 database. (B) Expression of the three signature genes in the GSE112790 database. (C) Principal component analysis (PCA) of the HCC (n = 159) and non-tumor liver tissues (n = 159) using the logistic panel in the NODE dataset. (D) ROC curve analysis of each of the signature genes and of the logistic panel in the NODE dataset. DCN: AUC = 0.875 (95% CI, 0.833 to 0.916); SERPINH1: AUC = 0.827 (95% CI, 0.781 to 0.874); ITGB1: AUC = 0.530 (95% CI, 0.465 to 0.596); logistic panel of three genes: AUC = 0.987 (95% CI, 0.974 to 1.000). ROC, receiver operating characteristic; AUC, area under curve; NODE, National Omics Data Encyclope.

which are subsequently secreted into the basement membrane of ECM [43]. COL4A1 is highly expressed in gastric [44], colon [44] and breast cancers [45], and is also strongly associated with the proliferation, differentiation, and migration of cancer cells [43]. Wang et al. [46] reported that increased expression of COL4A1 facilitated the proliferation and metastasis of HCC cells.

Collagen-related CAMs are also important since they function with collagen to promote the formation of pre-cancerous liver lesions, or even cancer. Through network analysis, we found that LUM, SERPINH1, DCN, SPARC, PCOLCE and ITGB1 were closely connected to the collagen gene regulation network. The degrees for these 6 genes in the network were all >20 (Fig. 1A,B).
The current results, deep RNA sequencing found that toxicity of CD8+ T cells function to suppress IFN-gamma secretion and the cyto-
group of HCC. Treg cells are often enriched in HCC and this panel.

CIBERSORT analysis indicated an elevated proportion of infiltrating Treg cells in the high-risk score subgroup in HCC. Treg cells are often enriched in HCC and function to suppress IFN-gamma secretion and the cytotoxicity of CD8+ T cells [47]. The transformation of M0 macrophages into the M2 subtype is commonly observed in the TME during cancer cell invasion [48]. Cytokines such as IL-10 and transforming growth factor-β were secreted by the M2 subtype, thereby promoting inflammation in tumor [49]. The increased proportion of M0 macrophages within the tumor immune microenvironment might therefore significantly contribute to liver carcinogenesis. Resting NK cells could convert to activated NK cells and target tumor cells [50]. In the present study, the fraction of resting NK cells was lower in the high-risk score subgroup, but the proportion of activated NK cells showed no difference between the high- and low-risk groups. Collectively, these findings indicate that high-risk scores may correlate with immunosuppression in HCC.

**DCN** belongs to a small, leucine-rich proteoglycans family that suppresses tumor growth [51]. Initially, it was identified as an efficient collagen-binding partner crucial for fibrillogenesis, and was therefore named decorin (DCN). DCN has since been reported to influence various biological processes such as cell growth, proliferation, adhesion, spread, and migration. Additionally, DCN plays a regulatory role in inflammation and fibrillogenesis [52]. A lack of DCN facilitates tumor development, and hence dysregulated DCN expression is observed in several cancer types including pancreatic and breast [52]. Consistent with the current results, deep RNA sequencing found that DCN expression levels were significantly decreased in HCC samples [53]. In summary, DCN could be an ideal target for treating solid malignancies. Interestingly, TIME analysis also revealed a strong association between DCN expression and CAFs in this study (Fig. 4B). CAFs are recognized as key cells in tumor development and invasion through their secretion of cytokines and growth factors. CAFs also promote tumor cell proliferation and can cause immunotherapy failure. The significant correlation observed between DCN and CAFs suggests an important role for DCN in HCC.

Our LASSO-Cox analysis led to the construction of a prognostic signature consisting of three CAM genes (DCN, SERPINH1, and ITGB1). Based on this risk score panel, the high- and low-risk HCC groups derived from the TCGA dataset showed significant differences in gene mutation status (Fig. 3) and immune cell infiltration (Fig. 4), suggesting the possibility of stratified treatment for HCC patients using this panel.

Our LASSO-Cox analysis led to the construction of a prognostic signature consisting of three CAM genes (DCN, SERPINH1, and ITGB1). Based on this risk score panel, the high- and low-risk HCC groups derived from the TCGA dataset showed significant differences in gene mutation status (Fig. 3) and immune cell infiltration (Fig. 4), suggesting the possibility of stratified treatment for HCC patients using this panel.

**ITGB1** is one of the most important members of the integrin family and has been linked to tumor cell adhesion, tumor immunity, and metabolism [19]. ITGB1 is a tumor promoting factor that can induce the proliferation, migration, and invasion of cancer cells. ITGB1 also has the ability to bind to EpCAM, thereby regulating cancer cell adhesion [55]. Previous research demonstrated that breast cancer [56], colon cancer [57] and other solid tumor types expressed high levels of ITGB1. There are also some reports on the mechanism of action of ITGB1 in HCC. For example, Shi et al. [58] reported that Integrin Alpha 5 and ITGB1 cause resistance to Sorafenib by inducing the formation of vasculogenic mimicry in HCC.

**Fig. 7. RT-PCR validation of DCN, SERPINH1 and ITGB1 expression in an independent HCC cohort.** (A–C) The vertical axis depicts the relative mRNA expression values of genes normalized to ACTB, together with the corresponding variance calculated by the Mann-Whitney U test. Non-tumor: normal adjacent tissue samples (n = 31); tumor: tumor tissue samples (n = 31). RT-PCR, reverse transcription polymerase chain reaction; mRNA, messenger RNA; ACTB, actin-β.
Finally, by integrating the expression of hub genes from the RT-PCR and transcriptome data, we selected DCN, SERPINH1, and ITGB1 for additional diagnostic analyses. The combination of these three biomarkers showed very strong diagnostic accuracy for HCC (AUC = 0.987).

5. Conclusions

In summary, we carried out a systematic analysis of the expression and diagnostic value of three collagen-related CAM genes in HCC. Our results showed that the expression of DCN, SERPINH1, and ITGB1 were significantly altered in HCC. Moreover, we developed a combined logistic panel that proved to be an effective biomarker for HCC diagnosis. Our findings also suggest that DCN, SERPINH1, and ITGB1 are potential therapeutic targets for HCC. A logistic panel comprising these three genes could serve as a future non-invasive and effective diagnostic biomarker for HCC.

Availability of Data and Materials

The raw transcriptomic data of the present study are downloaded from the GEO data portal (https://www.ncbi.nlm.nih.gov/geo/; Accession number: GSE36376, GSE112790). The raw proteomic data was obtained from NODE (https://www.biosino.org/node) by pasting the accession (OEP000321). These are publicly available databases.

Author Contributions

JW and YTH carried out the main analysis. JW and YTH drafted the manuscript. YTH and KPZ carried out the RT-PCR experiment. KPZ and QLC drafted the RT-PCR part in manuscript. QYL, JZ, and JF collected the samples from clinic and analyzed the clinical data. QYL, JZ, and JF also helped to revise the manuscript. QYL, QLC and YYL performed the statistical analysis. YYL participated in the design of the study. YYL conceived of the study and coordination and helped to draft the manuscript. All authors read and approved the final manuscript. All authors have participated sufficiently in the work and agreed to be accountable for all aspects of the work.

Ethics Approval and Consent to Participate

This study was approved by the Official Ethics Committee of the Shanghai University of TCM (2014-345-41-01) and written informed consent was obtained from all participants.

Acknowledgment

Not applicable.

Funding

This research was funded by the National Natural Science Foundation of China (82274183, 81503478); Shanghai Municipal Health Commission’s special clinical research project in the health industry (202240243); the Shanghai Science and Technology Committee (STCSM) Science and Technology Innovation Program (No.20ZR1453700); Henan Natural Science Foundation of China (162300410189).

Conflict of Interest

The authors declare no conflict of interest.

Supplementary Material

Supplementary material associated with this article can be found, in the online version, at https://doi.org/10.31083/j.fbl2902076.

References

[13] Boyd NF, Guo H, Martin LJ, Sun L, Stone J, FisheII E, et al. Mammographic density and the risk and detection of breast can-


