A novel 12-marker panel of cancer-associated fibroblasts involved in progression of hepatocellular carcinoma

Baojia Zou1,1,*
Xia lei Liu1,1,*
Yihang Gong1,1,*
Chaonong Cai1
Peiping Li1
Shan Xing2
Bibesh Pokhrel1
Baimeng Zhang1
Jian Li1

1Department of Hepatobiliary Surgery, The Fifth Affiliated Hospital of Sun Yat-sen University, Zhuhai 519000, China; 2Department of Clinical Laboratory, Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Guangzhou 510060, China

*These authors contributed equally to this work

Background/Aim: Cancer-associated fibroblasts (CAFs) are important factors in the progression of hepatocellular carcinoma (HCC). But the characterization of these cells remains incomplete. This study aims to identify a panel of markers for CAFs that are associated with HCC progression.

Materials and methods: The sequencing data and clinicopathological characteristics of 366 patients were obtained from the Cancer Genome Atlas (TCGA) database (366 HCC tissues and there were 50/366 cases with corresponding normal liver tissues). In vitro validation of the markers was performed by quantitative real-time PCR using the hepatic stellate cell line LX2 induced by the HCC cell line Huh7. The activation of LX2 was confirmed by α-smooth muscle actin and fibroblast activation protein, using quantitative real-time PCR and immunofluorescence staining. In vivo detections of the 12 markers were done in 40 tissue samples (30 HCC and 10 normal).

Results: We successfully identified 12 CAF markers from TCGA data: FGF5, CXCL5, IGFL2, MMP1, ADAM32, ADAM18, IGFL1, FGF8, FGF17, FGF19, FGF4, and FGF23. The 12-marker panel was associated with the pathological and clinical progressions of HCC. All 12 markers were upregulated in vitro. In vivo expressions of these markers were paralleled with those in TCGA data.

Conclusion: A 12-marker panel of CAFs in HCC is identified, which is associated with both pathological and clinical progressions of cancer.

Keywords: hepatocellular carcinoma, cancer-associated fibroblasts, CAFs marker panel, TCGA database analysis, transcriptome profiling

Introduction

Hepatocellular carcinoma (HCC) is one of the leading causes of cancer death in both men and women worldwide.1 There is an increased number of activated fibroblasts, also known as the cancer-associated fibroblasts (CAFs) within the HCC microenvironment, which derives mainly from the activation of hepatic stellate cells (HSCs). Multiple studies have shown that CAFs can be a poor prognostic factor in HCC.2–4 There are ample evidences that suggest the active involvement of CAFs in the progression of different cancers.5,6 Although CAFs bear the features of activated fibroblasts, they are identified merely by one or two fibroblastic markers in most studies, eg, the α-smooth muscle actin (α-SMA).4,7 Therefore, the role of CAFs needs to be further characterized by more fibroblastic markers before their involvement in HCC progression can be better studied.

Not until recently, there has been little progress in the characterization of these heterogeneous stromal cells. Increased numbers of CAFs subpopulations have been...
found owing to the differential expression of fibroblastic panels in these cells, which is explored, instead of a single marker.\textsuperscript{8–10} It is believed that these stromal cells may share dissimilar fibroblastic patterns in different cancers and cancer stages,\textsuperscript{4,11–13} while some markers of CAFs are known to affect many aspects of HCC progression.\textsuperscript{7,14} In this sense, defining a certain panel of the fibroblastic markers can help to identify the CAFs coevolving with HCC.

The screening for fibroblastic markers can be performed for differential expression profiles in cancer tissues to obtain a more complete panel in HCC. As an integrated source for the sequencing data of most known human cancers, the Cancer Genome Atlas (TCGA) offers a good platform for the investigation. In addition, the known markers of activated fibroblasts are categorized as mesenchymal cell markers, cytokines, chemokines, growth factors, extracellular matrix components, and related enzymes.\textsuperscript{4,7}

The aim of this study was to explore and identify the marker panel of CAFs associated with tumor progression in HCC. The differential expression profiles between 1) normal liver tissues and HCC tissues and 2) early stage and advanced-stage tumors were retrieved from TCGA database. Based on these, the identified fibroblastic markers were associated with the pathological progression of HCC. The correlation of the marker panel with clinical outcomes was then investigated. The panel was validated in the in vitro model of induced CAFs, ie, the HSC cell line activated by HCC cancer cell, as described in previous study,\textsuperscript{15} and in the postsurgical specimens from HCC patients.

Materials and methods

HCC mRNA sequencing data from the public TCGA database

The HCC mRNA sequencing data were acquired in the open-source TCGA database, from the National Cancer Institute GDC data portal, USA (https://portal.gdc.cancer.gov/). We searched for the HCC mRNA sequencing data by the following preconditions: 1) transcriptome profiling; 2) gene expression quantification; 3) RNA sequencing; 4) primary cancer site defined as the liver; 5) TCGA program; 6) pathological diagnosis as the liver HCC; and 7) both genders, all races, all ages, and all follow-up status. Altogether, 371 cases were identified with 424 sequencing files (RNA sequencing data form of the Illumina HiSeq 2000 RNA Sequencing platform). After excluding cases with incomplete clinical information, we included 366 cases for further analysis. There were 50 cases with sequencing data of the normal/nontumor liver tissues.

Data integration and differential gene screening in software R

Screening the differentially expressed genes for further CAF expression profile analysis was performed in the software R. The 366 included samples were classified according to pathological stages (I, II, III, and IV). Subsequently, two different patterns of early stage and advanced-stage groups were defined: 1) pathological stage I and stage II–IV, and 2) pathological stage I–II and stage III–IV respectively. Then, the differential gene expressions were screened between 1) the normal and tumor tissues, and 2) the early stage and advanced-stage tumor tissues using the edgeR package in R 3.4.3. The cutoff set as a false discovery rate was <0.05, and the genes with |logFC| of >1 were considered as significantly changed. The differential genes obtained were then used for further analysis using the pheatmap package in the software R.

CAFs expression profiles in different stages of HCC

The differential genes obtained were matched with the fibroblast profile.\textsuperscript{4,7} Subsequently, the relative expression profiles of CAFs in different HCC stages were plotted in the GraphPad Prism version 6.00 for Windows (GraphPad Software, La Jolla, CA, USA).

Integration of all markers and determination of low/high expression in each panel

To evaluate the correlation between different panels and clinical outcomes, we integrated the expression of all markers in each panel for every patient. In the integration process, we first assumed the same weighted score to each marker. For example, 1/12 was applied to the expression of each marker in each panel for every patient. In the integration process, we first assumed the same weighted score to each marker. For example, 1/12 was applied to the expression of each marker in the 12-CAF marker panel. Thereafter, the weighed expressions of these markers were added up to become one integrated expression. Accordingly, the integrated expressions of these panels were made in each patient. The cutoff values of high and low expression were then set at 50% of the integrated expressions among all patients. Subsequently, these patients were divided into two groups in each CAF panel, ie, the low and high expression groups.

Patient specimens

Thirty fresh tumor tissue samples and 10 normal liver tissues were collected right after surgical resections (total number = 40), between the year 2010 and 2017, in The Fifth Affiliated Hospital, Sun Yat-sen University, Guangdong.
Province, China. These specimens were then immediately stored at –80°C. All patients received no treatments before their first surgical resections in our center. All tumor cases were pathologically confirmed HCC. The clinicopathological characteristics are shown in Table S1. The normal liver tissues were collected from trauma patients who underwent partial hepatectomy with no other known diseases. The TNM staging were determined according to the 2010 American Joint Committee on Cancer staging system. All samples were collected after obtaining written informed consent from the patients in our center, and the study was approved by the Institutional Ethical Review Board of The Fifth Affiliated Hospital, Sun Yat-sen University.

**Cell lines and cell culture**

The HCC cancer cell line Huh7 was obtained from Cancer Center, Sun Yat-sen University, Guangzhou, Guangdong, China. The HSC line LX2 (Merck Millipore, Billerica, MA, USA) was obtained from Fudan University, Shanghai, China. The use of the cell lines was approved by the Institutional Ethical Review Board of The Fifth Affiliated Hospital, Sun Yat-sen University. LX2 was cultured in RPMI-1640 medium (Gibco, Thermo Fisher Scientific, Waltham, MA, USA) and Huh7 was cultured in DMEM medium (Gibco). All culture media were supplemented with 10% FBS, 50 U/mol penicillin, and 50 mg/mL streptomycin (Sigma-Aldrich Co., St Louis, MO, USA). The cells were maintained in a humidified incubator at 37°C with 5% CO2. Conditioned medium (CM) from Huh7 was collected, centrifuged at 3,000 g for 20 minutes, and then filtered with 0.45 µM filters. The CM was supplied with RPMI-1640 containing 10% FBS (1:1) for the induction of LX2 cells. LX2 was induced in the CM for at least 48 hours.

**Immunofluorescence staining**

LX2 was plated, induced, and grown on glass coverslips up to 60% confluence. After being fixed in 4% paraformaldehyde, the cells were incubated in primary rabbit anti-α-SMA (Abcam, Ab32575, 1:500 dilutions) overnight at 4°C. Afterward, the cells were washed and incubated for 1-hour with fluorescent-conjugated secondary antibody (BA1105, 1:500 dilutions; BOSTER Biological Technology, Wu Han, China). Finally, cells were washed and stained by DAPI (BOSTER Biological Technology). Then, the cells were observed and imaged by the fluorescence microscope (ZEISS, Oberkochen, Germany).

**Quantitative real-time PCR**

Total RNA from cultured cells were extracted by using RNeasy Plus (Takara Bio, Da Lian, China). NanoDrop 2000 (Thermo Fisher Scientific, Waltham, MA, USA) was used to determine the concentration of extracted RNA. TaKaRa One Step RNA PCR Kit (AMV) (Takara Bio) was used for cDNA synthesis. The PowerUp SYBR Master Mix Applied Biosystems (Invitrogen, Thermo Fisher Scientific) was used for the detection of relative gene expression. Each PCR was run in at least three independent experiments (Eppendorf Realplex System, Eppendorf AG, Hamburg, Germany). The comparative Ct method (ΔCt) was used for the relative expression calculation, which was normalized using β-actin as the internal control. The primer sequences were either designed on the NCBI website (http://www.ncbi.nlm.nih.gov/tools/primer-blast) or obtained online from the PrimerBank website (http://pga.mgh.harvard.edu/primerbank/). The primer sequences are shown in Table S2.

**Statistical analysis**

The SPSS statistical software (version 16.0; IBM SPSS, Armonk, NY, USA) and the GraphPad Prism software (version 6.0) were used for statistical analyses. All survival statistics (including the survival curve plotting) were done by the Kaplan–Meier survival method with log-rank test in SPSS. Cox proportional hazards regression, univariate survival analysis, and associations among different factors (chi-squared test) were performed using SPSS16.0. P-value <0.05 was considered significant.

**Results**

**Differential expression profile in HCC tumor tissues**

To set the background genes for further analysis, we compared the sequencing data between HCC and the adjacent nontumor samples (Figure 1A and Table S3). Meanwhile, we found the genes involved in tumor development by comparing the data in the early stage and the advanced-stage tumors. The definitions of different stages have been described in Materials and methods. Subsequent results were shown between stage I and stage II–IV tumors (Figure 1B and Table S4); stage I–II and stage III–IV tumors (Figure 1C and Table S5).

**Identify the CAF marker panel associated with pathological progression of HCC**

Next, among the identified gene profiles, we searched for fibroblastic markers according to published articles.7 As a result, the CAF profile differentially expressed in cancer was identified between normal liver and tumor (Figure S1A and Table S6). And the stage-associated CAF profile was
found between early stage and advanced-stage HCC tumors (Figure S1B and Table S7; Figure S1C and Table S8). Then we crossed the three CAF profiles to reach a specific panel associated with the progression of HCC (Tables S6 and S8). Subsequently, 12 markers were identified for the panel (Figure S1D and Table 1).

The 12 markers consisted of fibroblast growth factors (FGF5, FGF8, FGF17, FGF19, FGF4, and FGF23), matrix metalloproteinase-1 (MMP1), members from the family of ADAMs (a disintegrin and metalloproteinase, ADAM32 and ADAM18), C-X-C motif chemokine ligand-5 (CXCL5), and members of the insulin-like growth factor family of signaling molecules (IGFL2 and IGFL1).

To verify the expression patterns of the 12 markers in different HCC stages, we analyzed the expression of each gene according to the TNM staging. The relative expression profile was plotted for each marker (Figure S2). The panel was grouped into four different subpanels based on the stage expression patterns of these markers (Table 1). The association of these panels and the clinical features in HCC was subsequently explored.

**Association of the CAF panels and clinical outcomes in HCC patients**

Integration of all markers and determination of low/high expression of the panels have been described in the Materials and methods.

The association of different CAFs gene panels and patient clinical features is shown in Table S9. To evaluate the association between the expression of different panels and HCC patient survival, we built the overall survival curves using Kaplan–Meier method (Figure 2).
As a result, the high expression of the panel with advanced markers predicted worse patient survival (12-CAF marker panel, \( P = 0.001 \); 11-CAF marker panel, \( P < 0.001 \); 8-advanced CAF marker panel, \( P < 0.0001 \); Kaplan–Meier, log-rank test). Meanwhile, the expression of the panels with only early stage markers showed no influence on HCC patient survival (4-early CAF marker panel, \( P = 0.608 \); 3-early CAF marker panel, \( P = 0.671 \); log-rank test).

Furthermore, the multivariate analysis showed that the high expression of the panel with only advanced-stage markers was an independent risk factor for HCC patient survival (8-advanced CAF marker, Cox regression, \( P < 0.01 \)) (Table 2).

To reach a more precise conclusion, we performed stratified analysis of the 8-advanced marker panel. The stratifications performed were according to the two early marker panels (4- and 3-early CAF marker panel, Figure S3). The result showed that patients with high expression of the advanced marker panel significantly correlated with patient’s poor survival, regardless of the expression levels of the early markers (Figure S3, Table S10).

Taken together, the 12-CAF marker panel was significantly associated with the clinical progression of HCC. The subpanel analysis further confirmed that the panel consisted of markers from different tumor stages and was parallel to clinical features and patient prognosis.

### Table 1: The markers included in different subpanels.

<table>
<thead>
<tr>
<th>Gene set</th>
<th>Markers</th>
</tr>
</thead>
<tbody>
<tr>
<td>12-CAF marker panel</td>
<td>FGF5, CXCL5, IGFL2, MMP1, ADAM32, ADAM18, IGFL1, FGF8, FGF17, FGF19, FGF4, and FGF23</td>
</tr>
<tr>
<td>11-CAF marker panel</td>
<td>FGF5, CXCL5, IGFL2, MMP1, ADAM32, ADAM18, IGFL1, FGF8, FGF17, FGF19, and FGF4</td>
</tr>
<tr>
<td>8-advanced CAF marker panel</td>
<td>FGF5, CXCL5, IGFL2, MMP1, ADAM32, ADAM18, IGFL1, and FGF8</td>
</tr>
<tr>
<td>4-Early CAF marker panel</td>
<td>FGF17, FGF19, FGF4, and FGF23</td>
</tr>
<tr>
<td>3-Early CAF marker panel</td>
<td>FGF17, FGF19, and FGF4</td>
</tr>
</tbody>
</table>

**Abbreviations:** ADAM, a disintegrin and metalloproteinase; CAFs, cancer-associated fibroblasts; CXCL, C-X-C motif chemokine ligand; FGF, fibroblast growth factor; IGFL, IGF-like family member; MMP, matrix metalloproteinase.
In vitro confirmation of the 12-marker panel in HCC-induced CAFs model

We confirmed the panel in HCC-induced CAFs model, ie, the HSCs cell line LX2 that was activated by the HCC cancer cell line Huh7. As in most HCC studies, the activation of LX2 (induced CAFs) was indicated by the increased expression of α-SMA (Figure 3A, Figure S4) and fibroblast activation protein (Figure S4). In the induced CAFs model, we validated that all 12 markers were increased in the CAFs model upon Huh7 induction, using the quantitative real-time PCR assay (Figure 3B).

In vivo validation of the 12-marker panel in patient specimens from our center

The 12 markers were further validated in the postsurgical tissues from patients in our center. In consistent with the results found in TCGA data, the eight advanced markers (FGF5, CXCL5, IGFLL2, MMP1, ADAM32, ADAM18, IGFLL1, and FGF8) were upregulated in stage III–IV tumors. FGF17 were significantly increased in stage II and stage III–IV HCC tumors. FGF19 and FGF4 were significantly higher in stage I tumors. The change of FGF23 was insignificant, although there was discernable upregulation in some tumor tissues (Figure 4).

Discussion

In this study, we have successfully identified 12 markers for the CAFs panel in HCC, which was associated with the pathological progression of tumor. Furthermore, subpanel analysis confirmed that the panel correlated with both the TNM stages and the clinical progression. In vitro study also verified that all 12 markers were upregulated in the induced model of CAFs. The markers were validated in the postsurgical tissues from HCC patients in our center.

The 12-marker panel offers a solution concerning the molecular phenotyping of CAFs in HCC. Although there are emerging studies for various subpopulations of CAFs, their presence in HCC still depends on the application of individual markers.7–10,13 According to the previous reports, the 12-marker panel built in this study is also based on the expression patterns of fibroblast markers.8,10 Because it is our major concern to identify the CAFs that can coevolve with HCC, the panel includes general markers not only in the cancer tissues but also associated with both the early- and advanced-stage HCC, pathologically.

The panel recognized in our study should be considered as the marker pool for the identification of several subpopulations related to the process of HCC progression. In general, patients with high expression of the 12-marker CAFs panel suffered from increased clinical events of tumor progression. And it is in consistence with the results of previous studies that increased CAF features in patients indicate poor prognosis.18–20

HSCs have been reported to be the important source for CAFs in HCC.4 Meanwhile, they show enhanced features

<table>
<thead>
<tr>
<th>Factors</th>
<th>OS</th>
<th>Univariate</th>
<th>Multivariate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age: ≤68 vs &gt;68</td>
<td>0.890</td>
<td>0.044</td>
<td></td>
</tr>
<tr>
<td>Gender: male vs female</td>
<td>0.428</td>
<td>0.073</td>
<td></td>
</tr>
<tr>
<td>Race: White vs Asian vs Black vs others</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hepatitis/alcohol/other risk factors: yes vs no</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TNM staging: I vs II vs III vs IV</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vascular invasion: yes vs no</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tumor grade (1 vs 2 vs 3 vs 4)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AFP (μg/L): ≤20 vs &gt;20</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Platelet count (&lt;10^12/L): &lt;100 vs 100–300 vs &gt;300</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Albumin (g/L): &lt;35 vs 35–50 vs &gt;50</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total bilirubin, mg/dL (μmol/L): ≤1.0 (17.1) vs &gt;1.0 (17.1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Child–Pugh classification: A vs B vs C</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12-CAF marker combination: low vs high expression</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11-CAF marker panel: low vs high expression</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8-Advanced CAF marker panel: low vs high expression</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4-Early CAF marker panel: low vs high expression</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3-Early CAF marker panel: low vs high expression</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: p<0.05. Bold values indicate statistical significance.

Abbreviations: AFP, α-fetoprotein; CAFs, cancer-associated fibroblasts; OS, overall survival; HCC, hepatocellular carcinoma.
The total number of patients is 40, as for each group, n=10. *P<0.05; **P<0.01; ***P<0.001; ****P<0.0001.

Abbreviation: HCC, hepatocellular carcinoma.
of CAFs after receiving stimulations from cancer cells. It is also suggested that CAFs and activated HSCs may share a lot of common gene signatures.\textsuperscript{15,21–23} Therefore, the in vitro model used in our study may resemble the activated CAFs. As a validation of the specificity of the panel, all 12 markers are found to be upregulated in the in vitro CAFs model induced by HCC cancer cell line.

But it is more likely that CAFs bear changing expression of these markers across different tumor stages. As depicted in TCGA data and validated in the postsurgical tissues in our center, the 12-marker panel consists of at least two distinct subpanels, the early and advanced stage. The two subpanels are in parallel with the pathological and clinical prognostic features that correspond to different tumor stages. To recognize even more precise subpopulations by using the panel, a more stage-specific cell model is in need. One of the solutions is the primary culture of CAFs from human HCC tissues. Moreover, the relative expression of these markers should also be validated by more methods, eg, the immunohistochemistry staining in vivo and the flow cytometry in vitro. In fact, the crosstalk between cancer cells and these stromal fibroblasts remains largely unknown.\textsuperscript{4} Hopefully, exploring the 12-marker panel in future studies could benefit and boost the understanding of the cancer–stroma relationship.

There are some frequently used markers for the CAF studies in HCC, eg, the α-SMA and fibroblast-specific protein.\textsuperscript{3,11,18,24} But these markers failed to be included in the panel. It is thus highly suggested that they are not good indicators for the progression-related CAFs that we try to define here. It brings up the concern again that the study of CAFs using one or two of the fibroblastic markers is indeed not specific enough. However, these markers are sensitive to indicate the presence of activated fibroblasts in tumor tissues.\textsuperscript{7}

In summary, we report a 12-marker panel of CAFs in HCC, which is associated with both pathological and clinical progressions. The panel is validated in the in vitro CAFs model and in the HCC tumor tissues in vivo. By using such a panel, it will facilitate us to identify the specific CAFs in the process of HCC development. It can also help us understand how the microenvironment can adapt to different HCC developmental stages in coming studies.

Acknowledgment
This work was supported by the Fundamental Research Funds for the Central Universities (Sun Yat-sen University, Grant 17ykpy60).

Disclosure
The authors report no conflicts of interest in this work.

References

