Chronic liver disease (CLD) is associated with tissue-destructive fibrosis. Considering that common mechanisms drive fibrosis across etiologies, and that steatosis is an important cofactor for pathology, we performed RNA sequencing on liver biopsies of patients with different fibrosis stages, resulting from infection with hepatitis C virus (HCV) (with or without steatosis) or fatty liver disease. In combination with enhanced liver fibrosis score correlation analysis, we reveal a common set of genes associated with advanced fibrosis, as exemplified by those encoding the transcription factor ETS-homologous factor (EHF) and the extracellular matrix protein versican (VCAN). We identified 17 fibrosis-associated genes as candidate EHF targets and demonstrated that EHF regulates multiple fibrosis-associated genes, including VCAN, in hepatic stellate cells. Serum VCAN levels were also elevated in advanced fibrosis patients. Comparing biopsies from patients with HCV with or without steatosis, we identified a steatosis-enriched gene set associated with advanced fibrosis, validating follistatin-like protein 1 (FSTL1) as an exemplar of this profile. In patients with advanced fibrosis, serum FSTL1 levels were elevated in those with steatosis (versus those without). Liver Fstl1 mRNA levels were also elevated in murine CLD models. We thus reveal a common gene signature for CLD-associated liver fibrosis and potential biomarkers and/or targets for steatosis-associated liver fibrosis.
Hepatic expression profiling identifies steatosis-independent and steatosis-driven advanced fibrosis genes

Divya Ramnath,1,2 Katharine M. Irvine,3,4 Samuel W. Lukowski,1,2 Leigh U. Horsfall,3,5 Zhixuan Loh,1,2 Andrew D. Clouston,4 Preya J. Patel,1,5 Kevin J. Fagan,1 Abishek Iyer,1,2 Guy Lampe,6 Jennifer L. Stow,1,2 Kate Schroder,1,2 David P. Fairlie,3,2 Joseph E. Powell,1,2,7 Elizabeth E. Powell,3,5 and Matthew J. Sweet1,2

1Institute for Molecular Bioscience (IMB) and 2IMB Centre for Inflammation and Disease Research, The University of Queensland, Brisbane, Queensland, Australia. 3Centre for Liver Disease Research and 4Faculty of Medicine, Mater Research Institute, Translational Research Institute, The University of Queensland, Brisbane, Queensland, Australia. 5Department of Gastroenterology and Hepatology, Princess Alexandra Hospital, Brisbane, Queensland, Australia. 6Pathology Queensland, Princess Alexandra Hospital, Brisbane, Queensland, Australia. 7Garvan Institute of Medical Research, Sydney, New South Wales, Australia.

Introduction

Hepatitis C virus (HCV) infection affects between 63 and 79 million people worldwide (1) and can lead to liver fibrosis, cirrhosis, hepatic decompensation, and hepatocellular carcinoma (HCC). Multiple direct-acting antiviral medications are now available to effectively treat HCV in over 95% of subjects, and, without advanced fibrosis (METAVIR stage 3–4), eradication of virus is regarded as a “cure.” Following a sustained virological response, there is gradual regression of hepatic fibrosis, even in patients with cirrhosis, and this is associated with a beneficial clinical outcome (2, 3). However, despite improvement in fibrosis, the cirrhotic liver does not show resolution of other key features, such as intrahepatic vascular shunts (4), and the risk of HCC persists (5). Therefore, in these subjects, ongoing surveillance for liver-related complications, particularly HCC, is recommended (6). Although the prevalence of cirrhosis in patients with chronic liver disease (CLD) remains unclear, a recent multicenter analysis of an HCV-infected cohort (n = 9,783 patients) in the US estimated that 28.5% of patients were cirrhotic and that liver disease severity was significantly underdocumented and underdiagnosed (7).

Chronic liver disease (CLD) is associated with tissue-destructive fibrosis. Considering that common mechanisms drive fibrosis across etiologies, and that steatosis is an important cofactor for pathology, we performed RNA sequencing on liver biopsies of patients with different fibrosis stages, resulting from infection with hepatitis C virus (HCV) (with or without steatosis) or fatty liver disease. In combination with enhanced liver fibrosis score correlation analysis, we reveal a common set of genes associated with advanced fibrosis, as exemplified by those encoding the transcription factor ETS-homologous factor (EHF) and the extracellular matrix protein versican (VCAN). We identified 17 fibrosis-associated genes as candidate EHF targets and demonstrated that EHF regulates multiple fibrosis-associated genes, including VCAN, in hepatic stellate cells. Serum VCAN levels were also elevated in advanced fibrosis patients. Comparing biopsies from patients with HCV with or without steatosis, we identified a steatosis-enriched gene set associated with advanced fibrosis, validating follistatin-like protein 1 (FSTL1) as an exemplar of this profile. In patients with advanced fibrosis, serum FSTL1 levels were elevated in those with steatosis (versus those without). Liver Fstl1 mRNA levels were also elevated in murine CLD models. We thus reveal a common gene signature for CLD-associated liver fibrosis and potential biomarkers and/or targets for steatosis-associated liver fibrosis.
Results

Patient characteristics at liver biopsy. Of the 69 patients included in the RNA-sequencing analysis, 55 had chronic HCV and 14 had FLD (12 NAFLD and 2 alcohol-related FLD). The HCV cohort included 26 subjects without steatosis and 29 subjects with steatosis (19, grade 1; 6, grade 2; and 4, grade 3 steatosis). The patient cohort was further subdivided into early and advanced fibrosis. Selected demographic and clinical characteristics of the patient cohort are summarized in Table 1.

Identification of genes associated with advanced fibrosis in CLD patients. As the first step in determining gene signatures associated with advanced liver fibrosis, we undertook RNA sequencing of liver biopsies from patients presenting with different etiologies (HCV or FLD) and fibrosis stages (n = 69 patients). These patients were categorized into early (stages 0–2) or advanced fibrosis (stages 3–4). To filter the data, a threshold of ≥10 counts per million (CPM) in at least 20 samples (29%) was set, with a total of 9,624 genes passing this criterion. Principal component analysis of the corrected counts for the 9,624 genes was performed to identify the components that contribute to maximum variability. As anticipated, the samples clustered based on early or advanced fibrosis (Figure 1A). In Figure 1B, a volcano plot highlights the 168 differentially expressed genes (DEGs) with study-wide significance between early and advanced fibrosis. Of these, 61 showed ≥2-fold difference (Figure 1B, blue), while 107 significant genes were differentially expressed but showed ≤2-fold difference (Figure 1B, orange). The less stringent FDR-corrected analysis revealed 1,595 DEGs between early and advanced fibrosis (Supplemental Data Set 1; supplemental material available online with this article; https://doi.org/10.1172/jci.insight.120274DS1). We next examined the expression patterns of the 168 DEGs across the patient samples. Figure 1C clearly shows a distinct separation in gene expression patterns in liver biopsies of patients with early versus advanced fibrosis. For a more comprehensive overview of the gene expression signature, we subjected the 1,595 (FDR-corrected) DEGs to KEGG pathway analysis and found that many pathways linked with cancer and/or metabolism, for example, the PI3K/Akt signaling pathway, were enriched in this gene set (Figure 1D). Further pathway analysis by MetaCore also identified predicted cellular processes (e.g., extracellular matrix [ECM] remod-
eling) and signaling pathways (e.g., TGF-β, WNTs, and Hedgehog) (Supplemental Data Set 2), which have previously been implicated in CLD progression (21, 22). In addition to the above analyses, we also performed more stringent (≥10 CPM in all samples) and relaxed (≥10 CPM in at least 1 sample) filtering of the data. Using these criteria, 5,404 and 12,596 genes were captured by the stringent and relaxed filtering approaches, respectively. For the stringently filtered set of 5,404 genes, 71 (study-wide significance) and 791 (FDR-corrected) genes were differentially expressed between early- and advanced-stage fibrosis, while, for the gene list generated through more relaxed filtering (12,596 genes), 314 (study-wide significance), and 2,883 (FDR-corrected) genes were identified. These data are also provided in Supplemental Data Set 1. All subsequent analyses focused on the medium filtered set of 9,624 genes.

Identification of a gene signature that correlates with ELF score. The ELF score, calculated from serum levels of TIMP-1, PIIINP, and HA, is widely accepted as a noninvasive tool to assess liver fibrosis severity (19, 23). To determine the genes correlating with ELF score, we took the corrected counts for all 9,624 transcripts and performed a linear regression analysis with the corresponding ELF scores. We identified 37 genes that significantly correlated with the ELF score, with an $r^2$ value of 0.3 or higher (Supplemental Data Set 1). As expected, most of the ELF score–correlated genes (86%) were also identified as DEGs in the early versus advanced fibrosis comparisons above. Genes that showed the strongest correlation ($r^2 \geq 0.39$) are plotted in Figure 2. The genes that most closely correlated with ELF score were cystic fibrosis transmembrane conductance regulator (CFTR) and osteopontin (SPP1), both of which have previously been implicated in regulating liver fibrosis (24–27), thus validating our approach. Interestingly, the transcriptional regulator ETS-homologous factor (EHF), also known as ESE-3, which itself is a DEG (Figure 1 and Supplemental Data Set 1) that correlates with ELF score (Figure 2), has been functionally linked to other genes in this list.

Table 1. Selected demographic and clinical characteristics of the patient cohort for the initial RNA-sequencing analysis

<table>
<thead>
<tr>
<th>Variable</th>
<th>HCV without steatosis</th>
<th>HCV + steatosis</th>
<th>FLD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Early</td>
<td>Advanced</td>
<td>$P$ value</td>
</tr>
<tr>
<td>No. of patients&lt;sup&gt;a&lt;/sup&gt;</td>
<td>16</td>
<td>10</td>
<td>0.92</td>
</tr>
<tr>
<td>Age (yr)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>38 (21–55)</td>
<td>49 (39–64)</td>
<td><strong>0.0031</strong></td>
</tr>
<tr>
<td>Sex (male)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>9 (56)</td>
<td>7 (70)</td>
<td>0.48</td>
</tr>
<tr>
<td>BMI (kg/m&lt;sup&gt;2&lt;/sup&gt;)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>25 (18–29)</td>
<td>27 (17–33)</td>
<td>0.25</td>
</tr>
<tr>
<td>Alcohol (around biopsy) (g/w)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>87 (0–323)</td>
<td>2 (0–10)</td>
<td><strong>0.006</strong></td>
</tr>
<tr>
<td>ELF score&lt;sup&gt;a&lt;/sup&gt;</td>
<td>8.8 (8.06–9.69)</td>
<td>10.3 (8.66–12.64)</td>
<td><strong>0.014</strong></td>
</tr>
<tr>
<td>Viral genotype&lt;sup&gt;a&lt;/sup&gt;</td>
<td>1</td>
<td>10 (62)</td>
<td>6 (60)</td>
</tr>
<tr>
<td>3</td>
<td>5 (32)</td>
<td>3 (30)</td>
<td>0.94</td>
</tr>
<tr>
<td>Other</td>
<td>1</td>
<td>1 (6)</td>
<td>1 (10)</td>
</tr>
<tr>
<td>Steatosis grade&lt;sup&gt;a&lt;/sup&gt;</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1 (6)</td>
<td>5 (38)</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>19</td>
<td>1 (8)</td>
</tr>
<tr>
<td>Inflammation grade&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0</td>
<td>1</td>
<td>6 (23)</td>
</tr>
<tr>
<td>1</td>
<td>12</td>
<td>75</td>
<td>3 (30)</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>19</td>
<td>4 (40)</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>18</td>
<td>3 (30)</td>
</tr>
<tr>
<td>Fibrosis stage&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0</td>
<td>6</td>
<td>23</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>19</td>
<td>NA</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>15</td>
<td>NA</td>
</tr>
<tr>
<td>3</td>
<td>NA</td>
<td>7</td>
<td>27</td>
</tr>
<tr>
<td>4</td>
<td>NA</td>
<td>3</td>
<td>12</td>
</tr>
</tbody>
</table>

<sup>a</sup>Data presented categorically as n (%) and analyzed using $\chi^2$ test. <sup>b</sup>Data presented continuously as median ± range and analyzed using Welch’s $t$ test. $P$ values in bold are statistically significant.
Figure 1. An overview of DEGs in early- versus late-stage liver fibrosis. (A) Principal component analysis of the corrected counts for the filtered genes (9,624 genes), highlighting the components that contribute to maximum variability (n = 69 patients; circles, early stages; triangles, advanced stages). (B) Volcano plot illustrating genes, from the filtered data set, that are significantly differentially expressed between early and advanced stages of liver fibrosis, as identified by RNA sequencing (n = 69 patients; EdgeR-generated FWER value and log2 fold change). (C) A heatmap showing the expression patterns of the 168 DEGs after hierarchical clustering using the complete linkage method (n = 69 patients). (D) KEGG pathway analysis of the DEGs between early- versus late-stage fibrosis (n = 69 patients; 1,595 FDR-corrected list). Circle size correlates with number of genes, and circle color indicates statistical significance.
for example, *CFTR, SPP1,* and *CD24* (28–30). To determine if EHF may be more widely linked to core liver fibrosis-associated genes, we compared the list of 37 ELF score–correlating genes to EHF target genes that were previously identified via a ChIP-seq approach in human airway epithelial cells (31). Indeed, 10 of the 37 ELF score–correlated genes that are expressed in fibrotic liver were also characterized as EHF targets in that study (Figure 3, A and B), representing a significant enrichment (hypergeometric test; *P* = 0.01). Additionally, we used an informatics approach (RcisTarget) (32) to identify the overrepresented transcription factor–binding motifs in the ELF score–correlated genes and found that the ETS motif, to which EHF binds, was significantly enriched (enriched in 9 genes, enrichment score 3.16) (Figure 3C and Supplemental Data Set 1). Combined, these two approaches implicated EHF as a core liver fibrosis regulator (17 of the 37 ELF score–correlated genes. We next used siRNA-mediated gene silencing of *EHF* in LX-2 hepatic stellate cells (HSCs) (Figure 3D) to determine whether this transcription factor is a likely regulator of TGF-β–inducible expression of ELF score–correlated genes. We examined 10 ELF score–correlated genes and found that 4 of these genes, *VCAN* (encoding versican), *DHR52* (encoding dehydrogenase/reductase family member 2), *COL1A1* (encoding collagen type I, α 1 chain) and *DTNA* (encoding dystrobrevin α), were TGF-β regulated (Figure 3, E–H). EHF silencing significantly reduced TGF-β–inducible mRNA expression of *VCAN, DHR52,* and *COL1A1* (Figure 3, E–G), while TGF-β–mediated suppression of *DTNA* mRNA expression was unaffected by *EHF* knockdown (Figure 3H). Silencing of *IRF6,* another epithelial-expressed transcription factor, had no significant effect on TGF-β–inducible *VCAN, DHR52,* or *COL1A1* mRNA levels. Collectively, these data suggest that EHF regulates the expression of multiple liver fibrosis-associated genes.

*Genes encoding extracellular proteins that are associated with progressive liver fibrosis.* In order to identify gene targets that may represent novel diagnostics and/or be amenable to targeting with biologicals, we subjected the list of 168 DEGs to pathway analysis tools and found that 77 of these genes encode secreted and/or cell-surface proteins (Supplemental Data Set 1). Genes encoding extracellular proteins showing greater than 2.5-fold differences between early and advanced stages are highlighted in Table 2. We also performed this analysis on the FDR-corrected gene list (1,595 genes), identifying 550 genes encoding extracellular proteins (Supplemental Data Set 1). To provide further stringency to our analysis, we compared the 168 DEGs to other published data sets, including an RNA sequencing–based study in a NAFLD cohort (33) and 3 microarray-based studies in NAFLD cohorts (34–36). We observed an overlap of 48 genes from our list with at least 3 other published data sets, with 16 of these genes being present in all 5 data sets (Figure 4). Of these 48 genes, 35 encode extracellular proteins (Figure 4B, highlighted in gray). Our attention was drawn to one of these, *VCAN,* which encodes an ECM protein recently implicated in the activation of HSCs (37) and that has also been linked to extracellular proteins (Figure 4B, highlighted in gray). We performed pathway analysis of the FDR-corrected gene list (1,595 genes), identifying 550 genes encoding extracellular proteins (Supplemental Data Set 1). Genes encoding extracellular proteins showing greater than 2.5-fold differences between early and advanced stages are highlighted in Table 2. We also performed this analysis on the FDR-corrected gene list (1,595 genes), identifying 550 genes encoding extracellular proteins (Supplemental Data Set 1). To provide further stringency to our analysis, we compared the 168 DEGs to other published data sets, including an RNA sequencing–based study in a NAFLD cohort (33) and 3 microarray-based studies in NAFLD cohorts (34–36). We observed an overlap of 48 genes from our list with at least 3 other published data sets, with 16 of these genes being present in all 5 data sets (Figure 4). Of these 48 genes, 35 encode extracellular proteins (Figure 4B, highlighted in gray). Our attention was drawn to one of these, *VCAN,* which encodes an ECM protein recently implicated in the activation of HSCs (37) and that has also been linked to inflammation and cancer metastasis (38). *VCAN* mRNA expression strongly correlated with ELF score (*r* value of 0.41) (Figure 2) and was inducible by the profibrotic cytokine TGF-β in an EHF-dependent manner in the LX-2 HSC line (Figure 3E). Figure 5A shows that *VCAN* mRNA was significantly elevated in biopsies of patients with advanced fibrosis. Four different isoforms of *VCAN* (V0–V3) have been described (39) (Figure 5B). To further validate our findings, and determine whether specific *VCAN* isoforms were differentially regulated, we performed qPCR for the different transcripts. The V0 and V1 isoforms were readily detectable and significantly elevated in liver biopsies from patients with advanced disease, whereas the V2 and V3 isoforms were expressed at very low to undetectable levels and showed no significant differences (Figure 5, C–F). We also examined the serum levels of VCAN in an independent cohort of patients with chronic HCV or NAFLD (Supplemental Table 1). Circulating VCAN was detected in only a subset of these patients (26%), and in this subset, levels were significantly elevated in those with advanced fibrosis (Figure 5G).

**Mapping the steatosis-associated liver fibrosis gene signature.** The above approach identified gene expression changes that occurred in early versus advanced stages of fibrosis, irrespective of the underlying causes of CLD. However, our profiling of HCV patients with or without steatosis provided an opportunity to identify steatosis-associated genes linked to advanced fibrosis. To address this, we reanalyzed the data by comparing patients with and without steatosis across all fibrosis stages. Somewhat surprisingly, this approach did not identify statistically significant DEGs (Supplemental Data Set 3). We presume that the strong fibrosis-associated gene signature overrides any steatosis-specific one. Consequently, we next identified DEGs between early and advanced fibrosis in HCV-infected patients with or without steatosis (*n* = 26 and *n* = 29, respectively). Whereas no genes showed study-wide significant differences in early versus advanced fibrosis in HCV patients without steatosis (Figure 6A), 58 genes had study-wide significance for early versus advanced fibrosis in patients with HCV and steatosis (Figure 6B). 22 genes were excluded from further analyses due to high variability between patient samples.
(coefficient of variance > 1), as this confounded data analysis. We further examined the expression patterns of the remaining 36 genes across the HCV patient subgroups (with or without steatosis) and found that most genes show similar trends in patients with early versus advanced fibrosis; however, the effect was more pronounced in those with steatosis (Figure 6C). To identify the subset of these fibrosis-associated genes that was most highly elevated in patients with steatosis, we calculated the fold change in early versus advanced fibrosis in the HCV versus HCV with steatosis patient subgroups. We identified 12 fibrosis-associated genes that showed greater than 3-fold differences between HCV patients with and without steatosis (Figure 6D), of which 10 encoded extracellular proteins (Table 3). Of the genes falling into this category, follistatin-like protein 1 (FSTL1), a member of the SPARC family (40), showed the highest fold difference (5.46-fold).

**FSTL1 as a candidate target for steatosis-associated liver fibrosis.** FSTL1 is known to drive pathology in inflammatory disease settings in mouse models, particularly arthritis (41–47). Among its various activities, it has been reported to promote inflammatory cytokine production, including IL-1β, from macrophages and fibroblasts (41, 47, 48). While FSTL1 mRNA was significantly elevated in CLD...
patients with advanced fibrosis (Figure 7A), it showed only a modest correlation with ELF score (r^2 value of 0.12) (Figure 7B). This lack of correlation is not surprising, given that elevated expression was only apparent in late-stage disease for patients with steatosis. Indeed, FSTL1 was significantly elevated in advanced versus early fibrosis stages in patients with HCV and steatosis but showed no significant difference in patients with HCV infection alone (Figure 7C). To further validate our findings, we examined serum levels of FSTL1 in an independent cohort of patients with chronic HCV or NAFLD (Supplemental Table 1). Circulating FSTL1 was detected in patients with steatosis, with significantly higher levels being apparent in advanced-stage NAFLD patients (Figure 7D). Given its potential as a steatosis-specific marker of advanced liver fibrosis, we also examined serum levels of FSTL1 in an independent NAFLD patient cohort (Supplemental Table 2) and again found that this protein was significantly elevated in patients with advanced fibrosis (Figure 7E). To determine whether similar effects were apparent during murine CLD, we examined Fstl1 mRNA levels in two mouse models of progressive liver fibrosis associated with different types of chronic liver injury, the hepatotoxin thioacetamide (TAA) model and the choline-deficient, ethionine-supplemented (CDE) model that develops steatosis. Fstl1 mRNA levels in the livers of mice subjected to both of these regimes were elevated by comparison with those of control mice (Figure 7, F and G).
Discussion

Our combinatorial approach of expression profiling CLD patient liver biopsies in parallel with ELF score analysis has enabled us to robustly identify genes associated with advanced fibrosis. Importantly, our analyses also enabled us to identify genes that were particularly elevated in steatosis-associated advanced fibrosis. Our clinically well-characterized patient cohort allowed us to perform stringent statistical corrections for confounding factors to identify genes that are altered at the population level. Pathway analysis of advanced fibrosis-associated genes revealed those involved in ECM remodeling, cell adhesion, cancer, TGF-β signaling, epithelial-to-mesenchyme transition, and immune response pathways (Supplemental Data Set 2). Many of the liver fibrosis-associated pathway signatures that were identified, including the dysregulation of several metabolic pathways (Figure 1D), are broadly consistent with those reported in other published studies (33, 35, 49–52). Furthermore, some of the individual DEGs have strong associations with liver fibrosis (27, 53–56).

We added another layer of cross-validation to our analysis of fibrosis-associated gene expression by correlating patient ELF scores with gene expression. The ELF score correlates strongly with liver-related clinical outcomes (20), and the 37 genes identified through this approach (Supplemental Data Set 1) will be of particular interest for future functional studies. For example, CFTR and SPP1 showed the strongest correlation with ELF score (Figure 2). Approximately 30% of patients with cystic fibrosis develop liver fibrosis (53), and Cftr deficiency in mice predisposes to the development of liver fibrosis upon exposure to a profibrogenic xenobiotic (24). Similarly, SPP1 was recently identified as a mediator of liver fibrosis via activation of HSC (27). Several other ELF score–correlated genes, including DCDC2 and ITGBL1 (57), and CD24 and BICC1 (58), have emerging genetic or functional links with liver fibrosis. Furthermore, the identification of the Ets-family transcription factor EHF, which has been linked to the regulation of the ELF score–correlated genes CFTR, SPP1, and CD24 (28–30), implies that this transcriptional regulator may be an important driver of fibrogenesis. Indeed, ITGBL1, DCDC2, CD24, and EHF (all ELF score–correlated genes) were identified in independent studies as genes closely associated with HBV-associated liver fibrosis (57) and/or cirrhosis (59). Interestingly, 17 of the 37 ELF score–correlated genes, which were all expressed in fibrotic liver, were identified as EHF targets in airway epithelial cells (31) and/or were enriched for the EHF-binding motif, as identified by RcisTarget. Although EHF has been reported to act as a transcriptional repressor (60), individual members of the ETS family can act as both transcriptional activators and repressors in different contexts (61, 62). For example, EHF promoted expression of the death receptor DR-5 in HCT116 colorectal cancer cells (63) but inhibited Ras-inducible expression of target genes in differentiated epithelial cells (60). In our studies, silencing of EHF in the LX-2 HSC line impaired TGF-β–inducible mRNA expression of VCAN, DHRS2, and COL1A1 (Figure 3, E–G), thereby supporting the view that this transcription factor promotes expression of multiple genes that are associated with advanced fibrosis. Of note, however, EHF silencing in LX-2 cells did not affect TGF-β–mediated suppression of DTNA (Figure 3H), and the 6 other ELF score–correlated genes that were examined were...
not regulated by TGF-β in these cells (data not shown). It is possible that these genes are controlled by other profibrotic and/or proinflammatory stimuli and/or are expressed in other cell types, given that only TGF-β responses in HSCs were examined here. Whether EHF regulates the expression of these genes in such settings remains to be determined, and it is also conceivable that only a subset of candidate EHF targets identified in our analyses are bona fide targets of this transcription factor. Regardless of this, future functional studies on the role of EHF in the progression and severity of liver fibrosis are clearly warranted. Recent studies have revealed that genetic variation in EHF may be associated with severity of airways disease in cystic fibrosis patients (64–66), a subset of which also develop liver disease (53). By refining the list of DEGs to those encoding extracellular proteins, we describe gene candidates that may be amenable to biological-based therapeutic targeting. A similar approach was recently pursued in a NAFLD patient cohort, where 110 genes were found to be cell-surface proteins (33). For example, were also identified as differentially expressed extracellular proteins in our study. By comparing the advanced fibrosis-associated gene set from this study (168 genes) with those derived from other CLD expression profiling studies (33–36), we were able to refine the list to 35 genes encoding extracellular targets that were common to at least 4 of 5 data sets, including our own (Figure 4B). Several of these (e.g., SPP1 and CFTR) have strong genetic or functional links to liver fibrosis, but we focused on validating expression of VCAN (Figure 5), given the limited information that exists on its role in this disease process. Interestingly, we found that the V0 and V1 isoforms were elevated in advanced liver fibrosis (Figure 5, C and D), whereas other alternatively spliced transcripts were expressed at low to undetectable levels (Figure 5, E and F).
is in keeping with other studies demonstrating tissue-specific expression of VCAN isoforms (67–70). The $V_0$ and $V_1$ isoforms are also reportedly expressed in some tumors (71, 72), whereas the $V_2$ isoform has been demonstrated to inhibit cell proliferation (73). Thus, it is tempting to speculate that selective expression of the $V_0$ and $V_1$ isoforms in advanced fibrosis may have some role in HCC progression. Recently, VCAN was suggested to have a role in hepatic fibrosis, with in vitro studies showing that it promotes HSC activation (37). However, its role in driving inflammation and liver fibrosis during CLD is unknown. Based on existing literature, it might be predicted to have dual roles in liver disease processes, promoting both HSC activation and inflammation. VCAN appears to exert several distinct proinflammatory functions. First, its presence in the ECM promotes leukocyte adhesion and inflammatory responses (74). Second, it can directly trigger Toll-like receptor 2–mediated activation of macrophages, a response that has been linked to metastasis in a breast cancer model (38). VCAN is processed by members of the a disintegrin and metalloproteinase with thrombospondin type-1 motifs (ADAM-TS) family of proteases, including ADAM-TS1, the expression of which was also elevated in patients with advanced fibrosis (Supplemental Data Set 1). Members of this family have previously been functionally linked to liver fibrosis pathology (75, 76). Therefore, it is possible that processing of VCAN generates a profibrotic activity that contributes to disease progression. However, VCAN has a well-recognized role in maintaining the ECM structure by interacting with other ECM components, such as hyaluronan, fibronectin, and type-I collagen (77). A contribution of full-length VCAN to the progression of liver fibrosis is therefore also possible.

Our comparison of patients with HCV with patients with HCV and those with HCV and steatosis enabled us to identify genes that were most highly expressed in patients with both steatosis and advanced fibrosis. This approach revealed 12 such genes (Figure 6D), 10 of which encoded extracellular proteins (Table 3). Since inflammation, associated cell death, and aberrant wound healing are central to the progression of liver fibrosis in NAFLD (78), we pursued FSTL1 as a steatosis-linked gene, because of its strong links to both inflammation and fibrosis (40, 41, 43–45, 79). FSTL1 is a secreted glycopro-
tein belonging to the SPARC family (40). Although it is known to drive pathology in mouse models of inflammatory diseases, particularly arthritis (43, 45, 46), the mechanisms involved remain unclear. However, a consistent theme is that it is linked to proinflammatory cytokine production from multiple cell types. For example, FSTL1 promoted inflammatory cytokine production, including IL-1β, from macrophages and fibroblasts (44, 48). Overexpression of Fstl1, by in vivo adenoviral transduction in DBA/1 mice, promoted expression of the proinflammatory cytokines TNF-α, IL-1β, and IL-6 in the liver (47). Although it is currently unclear which receptors are required for FSTL1-mediated inflammatory responses, recent studies have implicated both the NLRP3 inflammasome and IL-17 signaling in downstream responses, both of which are strongly implicated in liver fibrosis and adipogenesis (48, 80–85).

Importantly, FSTL1 has also recently been linked to pathology in pulmonary (79) and hepatic fibrosis (86) in mice. Silencing Fstl1 using shRNA reduced Col1a1 mRNA expression and macrophage accumulation, while increasing myofibroblast density and Timp-1 mRNA expression in carbon tetrachloride injury-mediated liver fibrosis in mice (86). Furthermore, an anti-FSTL1 blocking antibody protected against bleomycin-induced lung fibrosis, with the mechanism linked to a reduction in TGF-β-induced ECM production and myofibroblast differentiation (79). Haplodeletion of Fstl1 in mice protected against bleomycin-induced epithelial injury and impaired profibrotic signaling responses, restoring the balance between TGF-β and BMP signaling. Thus, the emerging links among FSTL1, inflammation, adipogenesis, and fibrosis, together with our evidence of elevated FSTL1 expression in steatosis-associated liver fibrosis, suggest that this secreted factor may have a role in the progression of NAFLD.

Since advanced fibrosis and cirrhosis are linked to age, this is a confounding factor that complicates interpretation of most expression profiling studies of CLD patient samples. Thus, our RNA-sequencing findings, and other similar studies (33, 35, 36, 49–51), require independent validation studies for individual genes and/or proteins of interest. In this study, we have verified that levels of circulating VCAN (Figure 5G) and FSTL1 (Figure 7, D and E) were significantly higher in patients with advanced fibrosis compared...
with those with early-stage disease in independent, age-matched cohorts. These findings are consistent with the elevated hepatic mRNA levels of VCAN and FSTL1 in patients with advanced fibrosis, as identified in our initial expression profiling analysis (Figure 5A and Figure 7A). Our data are therefore consistent with these two genes being predictive of etiology-independent and steatosis-enriched advanced fibrosis, respectively, rather than merely being associated with aging. The genetic and functional links that exist between many other genes that we identified and liver fibrosis further support the notion that the gene sets identified in this study are bona fide indicators of CLD. In summary, in this study, we have used a stringent approach to identify hepatic gene expression profiles associated with advanced fibrosis in general as well as genes enriched in the setting of comorbid steatosis. These data sets should provide important resources for the pursuit of novel liver fibrosis targets and/or biomarkers, as exemplified by the validation of VCAN and FSTL1 in this study.

Methods

Clinical information for patient cohorts. This is a retrospective, cross-sectional study in 69 patients with chronic HCV or FLD who had undergone a liver biopsy between 2000 and 2015 at the Princes Alexandra Hospital, Brisbane, Australia. Liver tissue and sera were collected at the time of liver biopsy following an overnight fast for 8–10 hours and stored at –80˚C. Diagnosis of liver disease was based on standard biochemical and serological assays and consensus histological assessment of the liver biopsy by 2 blinded, experienced pathologists (ADC and GL) as previously reported (23). All patients with chronic HCV were HCV RNA positive at the time of liver biopsy and/or serum collection. A diagnosis of NAFLD required demonstration of hepatic steatosis in the presence of metabolic risk factors and the exclusion of substantial alcohol consumption (≥20 g/d) within the preceding 5 years, steatosis-inducing drugs, or CLD (including a prior history of alcohol-related liver disease). 78% of liver biopsies were ≥15 mm in length or showed definite cirrhosis. Fibrosis was assessed using a modified METAVIR score in these patients (87). Advanced fibrosis was defined as fibrosis stage 3 and 4. The METAVIR scoring system was used to assess hepatic inflammatory activity (87). Steatosis was diagnosed when it affected >5% of the parenchyma, and it was graded as mild (5%–33%, grade 1), moderate (34%–66%, grade 2), or severe (>66%, grade 3). Clinical/demographic data (BMI, age, alcohol consumption, metabolic comorbidities) were obtained from the medical record. An ADVIA Centaur XP system (Siemens Healthcare Diagnostics) was used to quantify serum HA, PIIINP, and TIMP-1 according to manufacturer’s instructions. The ELF score was autocalculated by the instrument (23). Subsequently, two independent patient cohorts were used for serum analyses: a cohort of 35 patients with chronic HCV or NAFLD and a cohort of 25 NAFLD/NASH patients. In these cohorts, fibrosis was assessed using a modified METAVIR score (87) in the cohort with chronic HCV or NAFLD and the Brunt score (88) for the NAFLD/NASH cohort.

Animal models. 4- to 5-week-old C57BL/6 female mice were treated with or without 300 mg/l TAA (MilliporeSigma) supplemented in drinking water for up to 12 weeks. For the CDE model, a diet deficient in choline was fed to 4- to 5-week-old C57BL/6 female mice together with drinking water supplemented with 0.15% ethionine (MilliporeSigma) for up to 3 weeks. Control mice were fed normal chow. Mouse livers were harvested at the endpoint of experiments for gene expression analyses.

Cell culture and gene silencing. The human HSC line LX-2 (provided Grant Ramm, Queensland Institute for Medical Research Berghofer, Brisbane, Queensland, Australia) was cultured in DMEM (Life Technologies) with 2% fetal bovine serum, 20 U/ml penicillin, and 20 mg/ml streptomycin. For gene-silencing experiments, EHF siRNAs (premade Stealth siRNA IDs: HSS120049, HSS178139, Life Technologies) and IRF6 siRNA (Stealth sense: CAGGCACCCUAUCAGCCCCUCUAUA; antisense: UAUAGAAGGGCGUGUAAG-GUGCCUG; Life Technologies; used as a relevant control gene) were diluted in Opti-MEM reduced serum media (Life Technologies) to a final concentration of 200 pM, after which they were transfected into LX-2
cells using Lipofectamine 2000 reagent (2.5 μl/transfection, Life Technologies) by the reverse transfection method for 24 hours, as described by the manufacturer. Cells were then stimulated with 10 ng/ml TGF-β (R&D Systems) for 24 hours or were left untreated, after which RNA was extracted and qPCR was performed.

**RNA purification and qPCR.** Total RNA was isolated from human liver biopsy tissue, mouse liver, or LX-2 cells using TRI Reagent (MilliporeSigma), as per the manufacturer's instructions. Up to 1 μg of the extracted RNA was used to generate cDNA using random hexamer (for human samples) or oligodT (for mouse samples) priming and SuperScript III reverse transcriptase (Invitrogen). The synthesized cDNA was diluted 10-fold in RNase-free water and used for qPCR analysis. Gene expression, relative to the control gene 18S ribosomal RNA (18S) (for human biopsy samples) or hypoxanthine-guanine phosphoribosyltransferase (HPRT) (for human and mouse samples), was determined by qPCR using the ΔCt method, SYBR Green PCR Master Mix (Applied BioSystems), and ViiA7 RT-PCR systems (Applied BioSystems). Primers used to amplify specific genes are listed in Supplemental Table 3.

**RNA sequencing and analyses.** RNA sequencing was performed at the IMB Sequencing Facility at The University of Queensland. RNA quality was assessed using the Agilent 2100 BioAnalyzer, and samples with an RNA integrity number ≥ 6.0 were included for RNA sequencing. Up to 1 μg RNA was used to synthesize mRNA libraries using the TruSeq Stranded Total RNA (Ribo-Zero GOLD) library preparation kit (Illumina), as per the manufacturer's instructions. Sequencing was performed using NextSeq 75 cycle (1 × 75 bp) high-output run (Illumina) on a NextSeq500 sequencing system (Illumina). Sequence reads were mapped to

**Figure 7. FSTL1 is elevated in patients with steatosis-associated late-stage fibrosis.** (A) A scatter plot showing the log₂-corrected counts of FSTL1 for all patients from the RNA-sequencing analysis (n = 69 patients; horizontal bars indicate mean ± SEM; circles, early stages; triangles, advanced stages). (B) Correlation analysis showing FSTL1 expression as the independent variable and ELF score as the dependent variable in HCV (blue), HCV with steatosis (red), NAFLD (orange), and alcoholic FLD (brown) patients (n = 69 patients). (C) FSTL1 corrected counts were plotted in early (E) versus advanced (A) stages of fibrosis from patients with HCV alone or HCV and steatosis (n = 55 patients; horizontal bars indicate mean ± SEM; circles, early stages; triangles, advanced stages). (D and E) Circulating levels of FSTL1 in sera from (D) patients with chronic HCV or NAFLD (see Supplemental Table 1 for details of the cohort; n = 35 patients; detectable levels in 29 patients; circles, early stages; triangles, advanced stages) and (E) patients with NAFLD (see Supplemental Table 2 for details of the cohort; n = 25 patients; detectable levels in 18 patients; circles, early stages; triangles, advanced stages) were determined by ELISA (data represented as mean ± SEM). (F and G) mRNA levels of Fstl1 were measured, relative to the housekeeping gene Hprt, in (F) 6- and 12-week-old TAA- or sham-treated mice (n = 5 mice per group) and (G) 1- and 3-week-old CDE- or sham-treated mice (n = 4 mice per group) (data represented as mean ± SEM). For A and C, FWER values were calculated using Bonferroni multiple-testing corrections; for D and G, FDR values were calculated using nonparametric ANOVA (Kruskal-Wallits test) followed by Benjamini-Hochberg multiple-testing corrections; and for E and F, P values were calculated using the Mann-Whitney test. P or FDR ≤ 0.05 was considered statistically significant.
the human hg19 reference genome using STAR aligner v2.3.0e (89), with 12.79–19.5 million reads per sample, and the Sequence Alignment/Map output files were converted into Binary Alignment/Map format using Samtools v0.1.19 (90). Transcript read counts were obtained using the HTSeq python package (91), following which the CPM value was obtained using the edgeR package (92). Genes that passed a threshold of 10 CPM or more in all samples (high stringency), 10 CPM or more in at least 20 patient samples (medium level stringency), or 10 CPM or more in at least 1 patient sample (relaxed stringency) were retained. Most subsequent analyses focused on the medium level stringency gene set. Filtered genes were subjected to normalization by the trimmed means of M value method (Supplemental Figure 1) (93). Normalization was performed after filtering the data, as has been previously recommended for analysis of RNA-sequencing data (94). Correlation analysis was performed on the clinical parameters (sex, ethnicity, etiology, smoking habit, technical factors, alcohol consumption, BMI, age, fibrosis stage, and ELF score) for all patients to determine confounding factors. Factors that correlated more than 25%, such as age, were considered strong confounders and were not corrected for in subsequent analyses. Genes were then subjected to a batch-effect correction based on a generalized linear model and analyzed for differential gene expression by likelihood-ratio test using the edgeR package in R (version 3.3.2)/Bioconductor (95). Differential gene expression analyses were performed by comparing early versus advanced stages of fibrosis across different etiologies and in HCV patients with or without steatosis. Multiple testing was performed using Bonferroni corrections (for family-wise error rate-adjusted [FWER-adjusted] or study-wide significant gene list) or Benjamini-Hochberg corrections (for a less stringent, FDR-adjusted gene list). FWER or FDR values of 0.05 or less were considered significantly different. For ELF score correlation analysis, a linear model was generated using the corrected counts for all genes (y variable) and ELF score (x variable) for each patient. The genes strongly fitting the model were then subjected to multiple testing corrections using the Bonferroni approach, as described above.

Pathway analysis, enrichment analysis, and meta-analysis. Pathway analysis was performed using the GeneGo MetaCore pathway analysis tool (Thomas Reuters) and KEGG pathway analysis via ClusterProfiler package in R (96). For extracellular protein analysis, GO analysis via ClusterProfiler package in R (96) and MetaCore (GO localization) tools were used. For transcription factor-target enrichment analysis, a hypergeometric test was performed on the ELF score–correlated gene set and 1,000 randomized gene sets using the phyper function in R (97). Transcriptional factor motif enrichment analysis was performed on the ELF score–correlated gene set using the RcisTarget package in R (32). For meta-analysis, lists of DEGs were taken from published data sets (33–35, 98). The Vennerable package (https://github.com/js229/Vennerable, commit 46057c9) was used to create the Venn diagram comparing all the data sets.

ELISA. Serum was isolated from patient blood, and VCAN and FSTL1 levels were quantified using a human VCAN ELISA kit (CUSABIO, CSB-E11884h) and human FSTL1 ELISA kit (R&D Systems, RDSDY1694), respectively, according to the manufacturer’s protocol.

Statistics. Statistical analyses of RNA-sequencing data are described above. For other experiments, data were combined from 3 or more independent experiments and statistical analysis was performed using RStudio (version 0.99.442) or GraphPad Prism (version 7) software. Statistical tests used for individual experiments include Bonferroni multiple-testing corrections, nonparametric ANOVA (Kruskal-Wallis test) followed by Benjamini-Hochberg multiple-testing corrections, Mann-Whitney test (2-tailed), 2-way ANOVA followed by Benjamini-Hochberg multiple-testing corrections, χ² test, and Welch’s t test (2-tailed) and are indicated in individual figure legends. P values less than or equal to 0.05 were considered as statistically significant.

Study approval. Informed written consent was obtained from each patient, and the protocol was approved by the Metro South Health and the University of Queensland Human Research Ethics Committees (ethics reference no. HREC/99/QPAH76 and 2003000092). All animal work was carried out with the approval of the University of Queensland Animal Ethics Committee (animal ethics clearance no. IMB/509/15/UQ and IMB/364/15/CIDRR).

Data availability. The raw and processed data from this study have been submitted to ArrayExpress (https://www.ebi.ac.uk/arrayexpress/) under accession E-MTAB-6863.

Author contributions

DR designed and performed experiments; acquired, analyzed, and interpreted the data; and drafted the manuscript. KMI contributed to study design and data interpretation and critically revised the manuscript. SWL contributed to data analyses and interpretation and critically revised the manuscript. LUH collected clinical samples and information from patients. ZL performed the animal model studies. ADC provided clini-
cal expertise and histological assessment of patient samples and critically revised the manuscript. PJP and KJF collected clinical samples and provided assistance with clinical information. AI assisted with the animal model studies. GL provided histological assessment of patient samples. JLS and KS contributed to study design, supervision, and data interpretation. DPF contributed to study design, supervision of the animal model experiments, and data interpretation. JEP contributed to study design and data interpretation, provided expertise in statistical analyses, and critically revised the manuscript. EEP designed and supervised the study, collected clinical samples and data, and drafted and critically revised the manuscript. MJS designed and supervised the study, interpreted the data, and drafted and critically revised the manuscript.

**Acknowledgments**

This work was supported by the IMB Centre for Inflammation and Disease Research at The University of Queensland as well as CSL Limited. The latter had no role in study design or data analyses. MJS is supported by a National Health and Medical Research Council (NHMRC) Senior Research Fellowship (APP1107914), DPF is supported by an NHMRC Senior Principal Research Fellowship (APP1117017), and KS is supported by an NHMRC Career Development Fellowship (APP1141131). We thank Siemens Healthineers for providing the ELF test.

Address correspondence to: Matthew Sweet, Institute for Molecular Bioscience, The University of Queensland, Brisbane, Qld 4072, Australia. Phone: 61.7.3346.2082; Email: m.sweet@imb.uq.edu.au. Or to: Elizabeth Powell, Centre for Liver Disease Research, The University of Queensland, Translational Research Institute, Brisbane, Qld 4102, Australia. Phone: 61.7.3443.8015; Email: e.powell@uq.edu.au.

---


83. Giles DA, Moreno-Fernandez ME, Dravanovic S. IL-17 axis driven inflammation in non-alcoholic fatty liver disease progression. Curr Drug Targets. 2015;16(12):1315–1323.


