qFibrosis: A fully-quantitative innovative method incorporating histological features to facilitate accurate fibrosis scoring in animal model and chronic hepatitis B

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.
Fibrosis: A fully-quantitative innovative method incorporating histological features to facilitate accurate fibrosis scoring in animal model and chronic hepatitis B patients

Shuoyu Xu1,2,3,*, Yan Wang4,5,6,*, Dean C.S. Tai1,*, Shi Wang7, Chee Leong Cheng7, Qiwen Peng1,2, Jie Yan1,3, Yongpeng Chen4, Jian Sun4, Xieer Liang4, Youfu Zhu4, Jagath C. Rajapakse2,3,8,11, Roy E. Welsch2,9, Peter T.C. So2,3,10,11, Aileen Wee7,12,*, Jinlin Hou4,*, Hanry Yu1,2,3,6,11,13

1Institute of Bioengineering and Nanotechnology, 31 Biopolis Way, #04-01, Singapore 138669, Singapore; 2Computation and System Biology Program, Singapore-MIT Alliance, 4 Engineering Drive 3, E4-04-10, Singapore 117576, Singapore; 3Biosystems and Micromechanics IRG, Singapore-MIT Alliance for Research and Technology, 1 CREATE Way, #04-13/14 Enterprise Wing, Singapore 138602, Singapore; 4Institute of Hepatology, Department of Infectious Diseases, Nanfang Hospital, Southern Medical University, 1838 North Guangzhou Avenue, Guangzhou 510515, China; 5Departments of Hepatobiliary Surgery and Pathology, Zhejiang Hospital, Southern Medical University, 253 Gongye Dadao Avenue, Guangzhou 510280, China; 6Department of Physiology, Yong Loo Lin School of Medicine, National University of Singapore, 2 Medical Drive, MD9 #04-01, Singapore 117597, Singapore; 7Department of Pathology, National University Hospital, 5 Lower Kent Ridge Road, Singapore 119074, Singapore; 8Bioinformatics Research Center, School of Computer Engineering, Nanyang Technological University, 50 Nanyang Avenue, Singapore 639798, Singapore; 9Sloan School of Management, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA; 10Department of Mechanical Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA; 11Department of Biological Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA; 12Department of Pathology, Yong Loo Lin School of Medicine, National University of Singapore, 5 Lower Kent Ridge Road, Singapore 119074, Singapore; 13The Mechanobiology Institute, 5A Engineering Drive 1, T-Lab #05-01, Singapore 117411, Singapore

Background & Aims: There is increasing need for accurate assessment of liver fibrosis/cirrhosis. We aimed to develop qFibrosis, a fully-automated assessment method combining quantification of histopathological architectural features, to address unmet needs in core biopsy evaluation of fibrosis in chronic hepatitis B (CHB) patients.

Methods: qFibrosis was established as a combined index based on 87 parameters of architectural features. Images acquired from 25 Thioacetamide-treated rat samples and 162 CHB core biopsies were used to train and test qFibrosis and to demonstrate its reproducibility. qFibrosis scoring was analyzed employing Metavir and Ishak fibrosis staging as standard references, and collagen proportionate area (CPA) measurement for comparison.

Results: qFibrosis faithfully and reliably recapitulates Metavir fibrosis scores, as it can identify differences between all stages in both animal samples (p<0.001) and human biopsies (p<0.005). It is robust to sampling size, allowing for discrimination of different stages in samples of different sizes (area under the curve (AUC): 0.93–0.99 for animal samples; 1–16 mm²; AUC: 0.84–0.97 for biopsies: 10–44 mm in length). qFibrosis can significantly predict staging underestimation in suboptimal biopsies (<15 mm) and under- and over-scoring by different pathologists (p<0.001). qFibrosis can also differentiate between Ishak stages 5 and 6 (AUC: 0.73, p = 0.008), suggesting the possibility of monitoring intra-stage cirrhosis changes. Best of all, qFibrosis demonstrates superior performance to CPA on all counts.

Conclusions: qFibrosis can improve fibrosis scoring accuracy and throughput, thus allowing for reproducible and reliable analysis of efficacies of anti-fibrotic therapies in clinical research and practice.

© 2014 European Association for the Study of the Liver. Published by Elsevier B.V. All rights reserved.
Excessive accumulation of extracellular matrix (ECM) results in fibrosis, which is the hallmark of chronic liver diseases (CLD) [1]. Progression of liver fibrosis is closely related to the development of major complications of CLD [2]. Chronic hepatitis B (CHB), a leading global health burden, is the major cause of cirrhosis and liver cancer [3]. With recent advances in efficacious antiviral therapies, the endpoint of fibrosis/cirrhosis regression can be achieved in long-term treatment of CHB [4]. Herein lies an increasing need for accurate and precise assessment of fibrosis, a prognostic indicator of chronicity and CLD sequelae, in order to facilitate and monitor the effective utilization of therapeutic advances [5].

Liver biopsy has long been the gold standard for fibrosis assessment in CLD [6]. It has the capability of providing histopathological information on various morphological parameters that have been clinically validated for their pathophysiological relevance, but are not obtainable with non-invasive techniques [7,8] such as liver stiffness measurements [9] and biochemical markers [10]. Currently, liver biopsy-based assessment remains the standard reference for monitoring therapeutic responses in both clinical research trials and actual practice [4,11].

However, conventional histological staging of fibrosis in liver biopsy is semiquantitative and highly subjective to sampling error and observer variations, as it basically relies on a global assessment of architectural distortion and associated fibrosis. It cannot sufficiently and reliably reflect the complicated pathophysiological functional status of the liver, which is incumbent for diagnostic decision-making in current CLD management [5,12]. Furthermore, cirrhosis has recently been redefined to be a dynamic process with intra-stage progressive/regressive changes [13]; in this regard, the International Liver Pathology Study Group has called for biopsy-based histological markers that can quantify and predict intra-stage cirrhosis changes [14]. Thus, technologies that can provide feasible solutions to these issues may potentially improve fibrosis assessment in CLDs such as CHB.

Image-based morphometric analysis of biopsy samples has been explored as an alternative to histological staging systems [15]. The current method of choice is collagen proportionate area (CPA) measurement, which quantifies the extent of collagenous ECM deposition without incorporating architectural information about the damaged tissue landscape [15–17]. CPA correlates well with late stages of fibrosis but is highly sensitive to sample size [18]. Clinical applicability of CPA is still being critically evaluated.

The strengths and limitations of current assessment systems motivated us to develop an innovative method – qFibrosis for liver biopsy assessment, based on the strategy of combining pathology-relevant collagen architectural features with automated computer-aided image analysis tools. With input of imaging data from the liver sample, qFibrosis can automatically compute the fully quantitative fibrosis scores based on the respective collagen architectural features. Such a strategy potentially overcomes some limitations of the biopsy-based histological fibrosis assessment with a more accurate and quantitative staging methodology. Here we report the development of qFibrosis and verify its potential as a fibrosis assessment tool in both animal model and CHB patients.

Materials and methods

Thioacetamide-induced liver fibrosis in rats

The Thioacetamide (TAA)-induced animal model is used for studying liver fibrosis in rats. All the protocols for studying TAA-induced liver fibrosis rat models were reviewed and approved by the Biological Resource Centre (BRC) Institutional Animal Care and Use Committee (IACUC). Twenty-five rats were randomly separated into 5 groups, representing 5 time points – without drug treatment, and treated with TAA for 4, 8, 10, and 12 weeks. Liver specimens from the left lateral lobe of each animal were formalin-fixed, paraffin-embedded, and sectioned into consecutive slices of 50 μm for direct SHG-imaging and Masson Trichrome staining for histological examination [19]. Scoring was performed by an experienced pathologist using the Metavir fibrosis staging system [18].

Clinical biopsy samples

Clinical biopsy samples from two independent cohorts were included: 107 non-fragmented liver core biopsies for algorithm training and testing, and another well-balanced 55 long core biopsy samples for demonstrating the technology reproducibility and robustness. Both cohort samples were from CHB patients in Nanfang Hospital (Guangzhou, China). The clinical study was conducted according to the Declaration of Helsinki guidelines and approved by the Ethical Committee of Nanfang Hospital. All patients have given written informed consent for liver biopsy as well as permission for use of their medical records. The average length of the 107 biopsies was 16.7 ± 5.4 mm (minimum length: 10 mm, maximum length: 30 mm). The average length of the 55 biopsies was 30.4 ± 4.4 mm (minimum length: 25 mm, maximum length: 44 mm).

All the liver biopsy specimens were routinely processed by formalin fixation and paraffin-embedding, sectioned at 5 μm thickness for SHG-imaging, and then stained with Masson Trichrome for histological assessment. Biopsy samples were read independently by one hepatopathologist (A.W.) and one junior pathologist (W.S.), and staged using Metavir and Ishak fibrosis scoring systems. The detailed distribution of all biopsies, together with their Metavir fibrosis stages is summarized in Supplementary Table 1.

Image acquisition

The 107 samples for training and testing qFibrosis were imaged by the system of second harmonic generation/two photon excitation fluorescence (SHG/TPEF) microscopy established and adjusted as previously reported [19] at the Institute of Bioengineering and Nanotechnology, Singapore. Image acquisition was performed with a 20× objective on unstained sections of the tissue samples. To cover most of the sample areas, 3 nine-by-nine multi-tile images were acquired for the animal samples with a final image size of 16 mm² (4 × 4 mm); and up to 10 three-by-three multi-tile images for each human biopsy sample with final image size of 1.8 mm² (1.35 × 1.35 mm). The additional 55 samples to demonstrate reproducibility and robustness (or the degree of insensitivity to different image acquisition methods) were imaged by Genesis system (Histolindex, Singapore), an SHG/TPEF technology-based commercial device, at Southern Medical University (Guangzhou, China). Image acquisition parameters for these samples were set the same as the ones for the former cohort samples.

Establishing and measuring qFibrosis

The procedure for establishing qFibrosis includes (i) identification of different collagen patterns, (ii) extraction of collagen architectural features, and (iii) statistical analysis of features of the respective collagen patterns, which were then combined into a single index. Detailed descriptions of the protocols are provided in Supplementary Materials and methods.

The acquired images of samples were processed and calculated with the established qFibrosis. A numerical value between 0 and 1 was assigned to each sample while the higher value indicates more severe fibrosis.

Statistical analysis

The two-tailed Wilcoxon rank-sum test was performed to estimate the statistical differences of CPA and qFibrosis index between different Metavir and Ishak fibrosis stages, and differences of clinical measurements between Ishak stages 5 and 6. The DeLong test was used to compare the receiver-operating-characteristics curves (ROC) and area under ROCs (AUCs) of qFibrosis and CPA. The stepwise logistic regression was performed to find the best combination of markers to differentiate Ishak stages 5 and 6. Statistical significance level was set as p < 0.05.
Results

qFibrosis, an automated assessment of changes in collagen patterns and quantification of liver fibrosis

We employed the Metavir fibrosis staging system to illustrate the histopathological architectural features of the various collagen patterns acquired in CHB [20] (Fig. 1A). The main collagen patterns, namely, portal collagen (portal expansion), septal collagen (bridging fibrosis), and fibrillar collagen (fine collagen distributed in the pericellular/perisinusoidal space of Disse) were identified through image acquisition and processing, and translated into quantitative parameters to build up qFibrosis indices (Fig. 1B and C). In the statistical analysis framework of qFibrosis (Fig. 1C), a list of 87 collagen architectural features (Supplementary Tables 2-4) was categorized into 3 groups, namely, portal, septal, and fibrillar collagen; feature selection was performed to identify the most important architectural features [21]; principal component analysis was used to reduce the dimension of the selected features [22]; and multinomial logistic regression was performed to combine the principal components of the 3 subgroups (subindices) into a single index, qFibrosis. The potential use of qFibrosis in routine clinical practice is illustrated in Fig. 2.

qFibrosis scoring can faithfully replicate Metavir fibrosis staging

We first investigated the performance of qFibrosis to replicate the fibrosis scores obtained with conventional histological assessment such as Metavir staging system. qFibrosis reflected a continuum of fibrosis progression that was consistent with Metavir fibrosis stages in both animal model and CHB patients, of which the values are summarized in Tables 1 and 2, respectively.

In the rat model, 75 liver tissue images (16 mm²) were quantified with 15 images from each stage. qFibrosis values increased with fibrosis progression and showed significant differences between all the stages (p <0.001) (Fig. 3A). CPA showed drastic changes only in late stages and could not differentiate between early stages (stages 1 and 2) (Fig. 3B). In the CHB biopsies, qFibrosis values, obtained from 69 biopsies longer than 15 mm, success-fully differentiated between all stages (p <0.05) (Fig. 3C). In comparison, CPA could only differentiate between stages 3 and 4 (stages 1 vs. 2, p = 0.124; stages 2 vs. 3, p = 0.194) (Fig. 3D).

qFibrosis is less sensitive to sampling error

Sampling error is a major limitation when applying quantification methods such as CPA [18]. To assess the sensitivity of qFibrosis to sampling error, we first performed a proof-of-concept demonstration with animal samples. Different sizes were divided from a large-size section of liver containing a sufficient number of portal tracts for accurate scoring by an experienced pathologist. Images of the large sections were cropped to simulate samples of varying sizes (Supplementary Fig. 1). The coefficient of variation (CV) of qFibrosis was calculated for each sample at different sizes; the CV values gradually increased from 18% to 28% whilst the sample sizes decreased from 8 mm² to 1 mm² (Fig. 4A). In contrast, the CV of CPA increased more drastically from 20% to 46% for the same sample size (Fig. 4A). The CV of qFibrosis was significantly smaller than that of CPA for samples sizes at 4 mm² (p = 0.02), 2 mm² (p <0.001), and 1 mm² (p <0.001).

The performances of qFibrosis vs. CPA for fibrosis scoring with different sample sizes were evaluated with ROC analysis (Supplementary Fig. 2, Supplementary Table 5). The AUC values of qFibrosis decreased slightly along with the reduction in sample sizes (Fig. 4B). CPA achieved similar AUC values as qFibrosis using large samples at 16 mm²; however, the AUC values of CPA decreased drastically when the sample sizes were reduced (Fig. 4B). The differences of AUC between qFibrosis and CPA values became significant at half (8 mm²) stages 0 vs. 1, 2, 3, 4 and 0, 1 vs. 2, 3, 4; p <0.05, respectively) to one fourth (4 mm²) (all stages, p <0.001) of the original size for differentiating liver fibrosis stages. The performance achieved by qFibrosis at 1 mm² sample size (AUC: 0.95–0.93) was similar to that obtained by CPA at sample size of 8 mm² (AUC: 0.96–0.90).

In the clinical scenario provided by the CHB biopsies, the AUC values of qFibrosis for the detection of different stages of fibrosis on 69 samples longer than 15 mm were from 0.92 to 0.84, while the AUC values of CPA were smaller (0.76–0.71) (Fig. 4C, Supplementary Table 6). We further evaluated qFibrosis on all 107 non-fragmented human core biopsy samples that included both long (>15 mm) and short samples (<15 mm), as short samples are unavailable in routine clinical practice. AUC values of qFibrosis were maintained at higher than 0.8 for detection of significant fibrosis and cirrhosis, whereas the AUC of CPA dropped to 0.71 (Fig. 4D, Supplementary Table 6). Thus, we demonstrated that qFibrosis can potentially differentiate fibrosis at different stages in core biopsy samples of different sizes.

qFibrosis can aid in correction of sampling error-mediated intra-observer variation

Short core biopsy samples are known to yield underestimated scores in fibrosis staging [23]. We simulated the scenario of a pathologist scoring short core biopsy samples to investigate whether qFibrosis can identify the potential underestimation. We used all 69 good quality (>15 mm) biopsy samples to train a multinomial logistic regression model and applied it to the remaining 38 suboptimal (<15 mm) biopsy samples to obtain qFibrosis values. The underestimation of fibrosis stages by pathologists on suboptimal biopsy samples was accurately predicted by qFibrosis (Fig. 4E). CPA cannot predict the underestimation in all stages except for stage 4 (Fig. 4E). It is generally accepted that there is rare underestimation of Metavir F4 samples. Therefore, qFibrosis can potentially aid pathologists to adjust for the degree of aggressive vs. conservative scoring decisions to compensate for sampling error-mediated intra-observer variation.

qFibrosis can aid in correction of inter-observer variation

We investigated whether qFibrosis can identify the trend of deviation of a pathologist’s scoring with reference to an experienced pathologist’s scores. All 107 human samples were independently scored by two pathologists, A and B. Cohen's and Fleiss's kappa statistics were used to assess the inter-observer agreement between two (Cohen's) or any number (Fleiss's) of observers. The Cohen's kappa of the scores from the two pathologists was 0.40 (p <0.001), suggesting a fair but not strong agreement (Supplementary Fig. 3). Forty-nine out of 107 human samples with available FibroScan® measurements were chosen to compare the consistency of fibrosis scores by the two pathologists. The
Fig. 1. Schematic illustration of qFibrosis establishment. (A) Representation of changes in collagen patterns in chronic liver disease based on Metavir staging system. Portal, septal and fibrillar collagen are denoted in blue, green and red, respectively. (B) The 3 types of collagen patterns are shown in Thioacetamide (TAA)-induced rat liver samples with normal and advanced fibrosis, as visualised by Masson Trichrome-stained, TPEF/SHG and processed images. (C) Computation framework to establish qFibrosis.

Please cite this article in press as: Xu S et al. qFibrosis: A fully-quantitative innovative method incorporating histological features to facilitate accurate fibrosis scoring in animal model and chronic hepatitis B patients. J Hepatol (2014), http://dx.doi.org/10.1016/j.jhep.2014.02.015
cut-off values of non-invasive fibrosis markers, such as FibroScan®, APRI, and FIB-4, to predict cirrhosis (F4) or significant fibrosis (F2-4) were established in large cohort studies of CHB patients [24–26]. The scores from pathologist A were more consistent with all the clinical markers (with higher Fleiss's kappa indicating stronger overall agreement) than pathologist B (Supplementary Table 7). Thus, scores from pathologist A were used to train the multinomial logistic regression model to yield $q$Fibrosis values for all 107 samples. Compared to the scores from pathologist A, the scores from pathologist B were overestimated and underestimated by 3.7% and 42%, respectively. Such over- and underestimation can be accurately predicted by $q$Fibrosis but not by CPA (Fig. 4F). Thus, $q$Fibrosis can aid in the correction of inter-observer variation in fibrosis assessment by serving as a reliable proxy for experienced pathologists.

$q$Fibrosis can aid in detection and monitoring of intra-stage cirrhosis changes

To differentiate intra-stage cirrhosis changes, we calculated $q$Fibrosis values from 43 human samples that were categorized as cirrhosis (F4) on Metavir and under two substages 5 and 6 according to Ishak staging. $q$Fibrosis accurately differentiated these two substages ($p = 0.008$) with AUC of 0.73 whereas CPA failed to do so ($p = 0.302$) (Fig. 4G, Supplementary Fig. 4A). We also investigated whether the combination of $q$Fibrosis with non-invasive clinical markers would improve the detection of intra-stage cirrhosis. Nine routine clinical biomarkers and stiffness measurement by FibroScan® were first assessed in 17 of the 43 Metavir F4 samples, which had complete clinical data; only FibroScan® could differentiate intra-stage cirrhosis changes.

### Table 1. $q$Fibrosis values of Thioacetamide-treated animal samples.

<table>
<thead>
<tr>
<th>Fibrosis-metavir score</th>
<th>Percentage of total</th>
<th>25th percentile</th>
<th>75th percentile</th>
<th>Median</th>
<th>Mean</th>
<th>SEM</th>
</tr>
</thead>
<tbody>
<tr>
<td>F0 (n = 15)</td>
<td>20.0</td>
<td>0.016</td>
<td>0.144</td>
<td>0.049</td>
<td>0.074</td>
<td>0.017</td>
</tr>
<tr>
<td>F1 (n = 15)</td>
<td>20.0</td>
<td>0.157</td>
<td>0.410</td>
<td>0.258</td>
<td>0.266</td>
<td>0.040</td>
</tr>
<tr>
<td>F2 (n = 15)</td>
<td>20.0</td>
<td>0.350</td>
<td>0.504</td>
<td>0.429</td>
<td>0.434</td>
<td>0.031</td>
</tr>
<tr>
<td>F3 (n = 15)</td>
<td>20.0</td>
<td>0.743</td>
<td>0.854</td>
<td>0.751</td>
<td>0.779</td>
<td>0.025</td>
</tr>
<tr>
<td>F4 (n = 15)</td>
<td>20.0</td>
<td>0.968</td>
<td>1.000</td>
<td>0.998</td>
<td>0.956</td>
<td>0.021</td>
</tr>
</tbody>
</table>

SEM, standard error of mean.

### Table 2. $q$Fibrosis values in 69 chronic hepatitis B core liver biopsies longer than 15 mm.

<table>
<thead>
<tr>
<th>Fibrosis-metavir score</th>
<th>Percentage of total</th>
<th>25th percentile</th>
<th>75th percentile</th>
<th>Median</th>
<th>Mean</th>
<th>SEM</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1 (n = 12)</td>
<td>17.4</td>
<td>0.257</td>
<td>0.493</td>
<td>0.374</td>
<td>0.411</td>
<td>0.051</td>
</tr>
<tr>
<td>F2 (n = 9)</td>
<td>13.0</td>
<td>0.445</td>
<td>0.783</td>
<td>0.590</td>
<td>0.607</td>
<td>0.073</td>
</tr>
<tr>
<td>F3 (n = 18)</td>
<td>26.1</td>
<td>0.640</td>
<td>0.919</td>
<td>0.809</td>
<td>0.761</td>
<td>0.047</td>
</tr>
<tr>
<td>F4 (n = 30)</td>
<td>43.5</td>
<td>0.776</td>
<td>0.995</td>
<td>0.933</td>
<td>0.892</td>
<td>0.023</td>
</tr>
</tbody>
</table>

SEM, standard error of mean.
amongst these markers (Table 3). By stepwise logistic regression analysis, including all 10 markers together with qFibrosis and CPA, the combination of qFibrosis, FibroScan\(^n\), and international normalized ratio (INR) was the most predictive for differentiating intra-stage cirrhosis (Supplementary Table 8); the AUC improved from 0.81 (qFibrosis only) to 0.93 (combination of qFibrosis, FibroScan\(^n\), and INR) (Supplementary Fig. 4B). Thus, qFibrosis can differentiate intra-stage cirrhosis changes alone or in combination with FibroScan\(^n\) and INR.

Validation of qFibrosis on an independent cohort of CHB biopsy samples

We further tested the reliability of qFibrosis on images acquired by a commercial SHG/TPEF imaging device on another 55 core samples of mixed CLDs for scoring the amount of fibrosis via blinded evaluation with CHB samples; thus, promoting the readiness of its inherent full-quantification algorithm; it could potentially have discriminative power for precisely reflecting the dynamics of fibrosis/cirrhosis progression or regression.

Employing the similar imaging technique, Gailhouste et al. first comprehensively validated SHG on 119 clinical liver tissue samples of mixed CLDs for scoring the amount of fibrosis via detecting fibrillar collagen density, which is similar to CPA measurement [16]. Our present study is innovative in its strategy for establishing the qFibrosis index with histopathological architectural features by quantitatively defining the spatial parameters of fibrillar collagen. Another distinct contribution of our study is that qFibrosis was specially trained and validated with CHB samples; thus, promoting the ready applicability of our method to align closely with clinical practice of this particular disease.

We further analysed the performance of qFibrosis against CPA. While CPA showed limitations in discrimination accuracy and higher sensitivity to sampling error, as reported previously [11,18], qFibrosis exhibited significantly improved capacity to...
overcome the above limitations (Figs. 3 and 4). Considering the strategy taken for the qFibrosis design, it is rational that qFibrosis would behave more similarly to a conventional histological assessment system than CPA. This partly accounts for the robustness of qFibrosis to sample size-dependent sampling error (i.e., sample adequacy). On the other hand, CPA has significant

Please cite this article in press as: Xu S et al. qFibrosis: A fully-quantitative innovative method incorporating histological features to facilitate accurate fibrosis scoring in animal model and chronic hepatitis B patients. J Hepatol (2014), http://dx.doi.org/10.1016/j.jhep.2014.02.015.

Fig. 4. qFibrosis is superior to collagen proportionate area (CPA) in resolving biopsy-related issues of sampling error, inter-/intraobserver variation, and intra-stage discrimination in cirrhosis. (A) Coefficient of variance of qFibrosis and CPA in animal samples (p < 0.05). (B) Area under ROC curve (AUC) values of qFibrosis and CPA at all sample sizes in animal study. (C) Performances of qFibrosis and CPA in differentiating all fibrosis stages for human core biopsies including short ones (≥10 mm in length). (D) Performances of qFibrosis and CPA at all sample sizes in animal study. (E) Comparison of the capability to highlight potential underscoring of suboptimal biopsy samples to address size-dependent sampling error-mediated intraobserver variation by qFibrosis and CPA, respectively. The qFibrosis values of suboptimal biopsy samples scored as stages 1 to 3 are significantly higher than the qFibrosis values of good quality biopsy samples scored as the same stages, indicating that these suboptimal samples are underscored and may belong to later fibrosis stages (p < 0.05). (F) Comparison of the capability to predict interobserver over-/underestimation of biopsy samples by qFibrosis and CPA, respectively. The values of qFibrosis can significantly reflect the scoring-deviation under the same stages except for F4 (p < 0.05). (G) Differentiation of Ishak stage 5 from 6 by qFibrosis (p = 0.008) and CPA (p = 0.302).
deviation between different samples of the same stage scores, since histological staging and amounting fibrosis (CPA) are entirely different assessments [11,29]. Collectively, the results not only support the technical strength of qFibrosis for future applications, but also support our hypothesis that the improved discriminative power of qFibrosis is due to the additional input of histopathological architectural features.

We also showed that the performance of qFibrosis is reproducible between the original cohort of 107 samples and the independent cohort of 55 samples imaged by different SHG imaging devices. qFibrosis is highly reproducible when the image quality is reasonably consistent. The variations of SHG images due to different sample processing procedures or different imaging systems can be corrected using different optical settings and be calibrated according to Guilbert et al. [30]. Since qFibrosis depends on the spatial collagen information relative to tissue architectural features rather than on the absolute quantity of collagen intensity signal, this method is necessarily less sensitive to imaging or staining quality variation than other intensity-dependent methods such as CPA. Thus, qFibrosis is a robust quantitative staging method that would be suitable for the potential multicenter clinical research studies where sample processing and image acquisition variation would be unavoidable.

There are several exciting areas in which qFibrosis may have a role in the near future. First is in the realm of antiviral clinical research and management of CHB. It has been verified recently that long-term effective antiviral therapy can lead to regression of liver fibrosis and cirrhosis in CHB patients [4]. Histological analysis is currently the standard reference for performing the evaluation. Primarily established and validated with CHB samples, qFibrosis may soon be sufficiently improved to serve as an automatic and reliable adjunctive tool for liver biopsy evaluation. Second would be potentially in the area of cirrhosis assessment. Regression of cirrhosis has now been observed with the availability of HBV potent therapy [31], bringing with it increasing requirement for the assessment of regression and substaging [5,14]. qFibrosis could differentiate between Ishak stages 5 and 6; with the potential to detect the changes of histological patterns in cirrhosis progression or regression through the quantitative classification of different collagen patterns. In future studies correlating with or complementing other clinical markers such as FibroScan® [32] and hepatic vein pressure gradient (HVPG) [33] or complications of cirrhosis, we could further develop qFibrosis into a tool to aid monitoring the cirrhosis dynamics. Last but not least, since experienced hepatopathologists are a rare breed in most setups, qFibrosis might act as a valuable aid to pathologists to produce consistent staging of liver fibrosis; as well as to provide on-site expert consultation to the non-expert pathologists. In laboratories without SHG microscopy, qFibrosis values can be obtained from images of stained biopsy samples using routine light microscopy, as long as accurate identification of collagen can be ensured. The examples of both Masson Trichrome-stained and Sirius Red-stained images with qFibrosis evaluation are shown in Supplementary Fig. 6.

Performance of qFibrosis can be affected by the quality of samples in the training set; as evident from the higher AUC values obtained in staging larger animal samples in our results. This is because establishment of qFibrosis is basically the generation of an algorithm by training-and-learning with the sample's imaging data. Thus, in future work, more qualified biopsy samples for training purpose can further improve the performance of qFibrosis; and recruiting a larger set of samples for a multicentre clinical study would be necessary to generalize the capability of qFibrosis and validate its clinical applications. Fragmented biopsies are unavoidable in clinical practice due to the selection of sampling techniques [34]. In this study, we only selected non-fragmented biopsies for training and testing qFibrosis for accuracy in histological scoring reference [12,23]. Our observation that qFibrosis is less sensitive to sampling error suggests that fragmented biopsies contain information potentially extractable by qFibrosis. Further studies with large cohorts would be needed to evaluate the degree of tolerable fragmentation, to repair and partially reconnect the fragments by extrapolation, and to complement the non-fragmented biopsies in applications.

Since its target sample is liver biopsy, however superior the diagnostic utility of qFibrosis, invasiveness of the assessment is still an inherent limitation against its widespread application in preventive medicine. However, since liver biopsy provides comprehensive information not only on fibrosis but also on necroinflammation, steatosis, and other specific histopathological features, the role of qFibrosis can be expanded with the aid of TPEF or other staining or non-staining imaging modalities to quantify these other features. It is noteworthy to emphasize that caution needs to be exercised when applying qFibrosis to samples with significant amounts of these other histopathological features.

In summary, qFibrosis has been established and validated to provide quantitative scores incorporating histopathological features for liver fibrosis evaluation. It faithfully recovers the staging results of Metavir histological assessment system; while in the meantime, effectively ameliorating the inherent issues of the current systems regarding sampling error and observer variation. It can also differentiate stages 5 and 6 in the Ishak system suggesting the possibility of monitoring intra-stage cirrhosis changes. qFibrosis can potentially be a valuable tool to enhance the utility of liver biopsy for accurate and objective assessment of fibrosis in clinical research and management of CLD.

Conflict of interest

Dean C.S Tai co-founded and currently works in Histolindex Pte Ltd (Histolindex), a medical imaging device company. His contri-

<table>
<thead>
<tr>
<th>Table 3. Comparison of qFibrosis, collagen proportionate area (CPA) and clinical parameters for Ishak stages 5 and 6.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ishak stage 5</td>
</tr>
<tr>
<td>CPA (%)</td>
</tr>
<tr>
<td>qFibrosis (a.u.)</td>
</tr>
<tr>
<td>ALT (IU/L)</td>
</tr>
<tr>
<td>AST (IU/L)</td>
</tr>
<tr>
<td>ALB (g/L)</td>
</tr>
<tr>
<td>TBIL (mol/L)</td>
</tr>
<tr>
<td>INR (a.u.)</td>
</tr>
<tr>
<td>PT (s)</td>
</tr>
<tr>
<td>PLT (10^5/L)</td>
</tr>
<tr>
<td>FibroScan® (kPa)</td>
</tr>
<tr>
<td>APRI (a.u.)</td>
</tr>
<tr>
<td>FIB-4 (a.u.)</td>
</tr>
</tbody>
</table>

Data are Mean ± SD. ALT, alanine transaminase; AST, aspartate transaminase; ALB, albumin; TBIL, total bilirubin; INR, international normalized ratio; PT, prothrombin time; PLT, platelets; APRI, AST-to-platelet ratio index; and FIB-4, FIB-4 index.
Author contributions
S.X. developed image analysis tools, performed data analysis, designed the experiments, and wrote the manuscript. Y.W., J.H., and H.Y. designed the experiments, performed data analysis, and wrote the manuscript. D.T., J.R., R.W., and P.S. designed the experiments. S.W. and C.C. performed pathology scoring. Q.P. and wrote the manuscript. X.L., J.S., Y.C., and Y.Z. collected clinical samples and data. A.W. performed pathology scoring, designed the experiments, and wrote the manuscript.

Acknowledgements
This work was supported in part by the Institute of Bioengineering and Nanotechnology, Biomedical Research Council, Agency for Science, Technology and Research, Singapore; grants from Janssen, Computation and system biology program Singapore-MIT Alliance, Biosystems and Micromechanics IRG, Singapore-MIT Alliance for Research and Technology and Mechanobiology Institute of Singapore (R-714-001-003-271) to Jinlin Hou; and the National Natural Science Foundation of China, China (81371603) to Hanry Yu; the National Science and Technology Major Project of China (2012ZX10002003) to Jinlin Hou; and the National Natural Science Foundation of China, China (81371603) and Southern Medical University Zhujiang Hospital Young Talent Project funding (2012003) to Yan Wang. We thank Drs Zhenhai Zhang, Miguel Esteban, and Yanan Du for their constructive comments on the manuscript preparation; Mr. Haolin Liu, Misses Xiaoli Huang and Chuwen Lin for performing image acquisition with Genesis; and Dr. Karoline Lackner (Medical University of Graz, Austria) for providing the image of Sirius red stained liver tissue.

Supplementary data
Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jhep.2014.02.015.

References
Research Article


