

AperTO - Archivio Istituzionale Open Access dell'Università di Torino

qFIBS: A Novel Automated Technique for Quantitative Evaluation of Fibrosis, Inflammation, Ballooning, and Steatosis in Patients With Nonalcoholic Steatohepatitis

This is a pre print version of the following article:

Original Citation:

Availability:

This version is available <http://hdl.handle.net/2318/1725563> since 2020-01-28T12:06:00Z

Published version:

DOI:10.1002/hep.30986

Terms of use:

Open Access

Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.

(Article begins on next page)

qFIBS: A NOVEL AUTOMATED TECHNIQUE FOR QUANTITATIVE EVALUATION OF FIBROSIS, INFLAMMATION, BALLOONING, AND STEATOSIS IN PATIENTS WITH NON-ALCOHOLIC STEATOHEPATITIS

Feng Liu^{1*}, George Boon-Bee Goh^{2*}, Dina Tiniakos³, Aileen Wee⁴, Wei-Qiang Leow⁵, Jing-Min Zhao⁶, Hui-Ying Rao¹, Xiao-Xiao Wang¹, Qin Wang¹, Wei-Keat Wan⁵, Kiat-Hon Lim⁵, Manuel Romero-Gomez⁷, Salvatore Petta⁸, Elisabetta Bugianesi⁹, Chee-Kiat Tan², Stephen A. Harrison¹⁰, **Quentin M. Anstee^{11†}, Pik-Eu Jason Chang^{2†}, Lai Wei^{1†}**

*These authors share co-first authorship

† These authors share co-senior and corresponding authorship

Institutional affiliations:

1. *Peking University People's Hospital, Peking University Hepatology Institute, Beijing Key Laboratory of Hepatitis C and Immunotherapy for Liver Diseases, Beijing 100044, China.*
2. *Department of Gastroenterology and Hepatology, Singapore General Hospital, Singapore 169856, Singapore.*
3. *Institute of Cellular Medicine, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, UK & Department of Pathology, Aretaieion Hospital, Medical School, National & Kapodistrian University of Athens, Athens, Greece.*
4. *Department of Pathology, Yong Loo Lin School of Medicine, National University of Singapore, National University Hospital, 5 Lower Kent Ridge Road, Singapore 119074, Singapore.*
5. *Department of Anatomical Pathology, Singapore General Hospital, Singapore 169856, Singapore.*
6. *Department of Pathology and Hepatology, Beijing 302 Hospital, Beijing 100039, China. E-mail: jmzhao302@163.com*
7. *Unit for the Clinical Management of Digestive Diseases. Centro para la Investigacion Biomedica en Red de Enfermedades Hepaticas y Digestivas*

(CIBEREHD), Institute of Biomedicine Seville (IBIS), Virgen del Rocio University Hospital, University of Seville, Seville, Spain.

- 8. Sezione di Gastroenterologia ed Epatologia, Dipartimento di Medicina Interna e Specialistica, DIBIMIS, Università di Palermo, Palermo, Italy.*
- 9. Division of Gastroenterology, Department of Medical Sciences, University of Turin, Italy.*
- 10. University of Oxford, Radcliffe Department of Medicine, Oxford, United Kingdom.*
- 11. Institute of Cellular Medicine, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, United Kingdom; Liver Unit, Newcastle Upon Tyne Hospitals NHS Trust, Freeman Hospital, Newcastle upon Tyne, United Kingdom.*
E-mail: quentin.anstee@newcastle.ac.uk

Abbreviations:

ALT, alanine transaminase; AST, aspartate transaminase; AUC, area under curve; AUROC, area under the receiver operating characteristic; BMI, body mass index; CI, confidence interval; CRN, Clinical Research Network; CV, central vein; FFPE, formalin fixed paraffin embedded; FV, fat vacuoles; GGT, gamma-glutamyl transferase; MRE, MR elastography; MRI-PDFF, magnetic resonance imaging derived proton density fat fraction; NAFLD, nonalcoholic fatty liver disease; NAS, NAFLD activity score; NASH, nonalcoholic steatohepatitis; NPV, negative predictive value; PS, perisinusoidal; PPV, Positive predictive value; PT, portal tract; ROC, receiver operating characteristic; SAF, Steatosis, Activity, Fibrosis; SHG, second harmonic generation; TPEF, two photon excitation fluorescence.

Key words: Quantitative evaluation; qFIBS; NASH; qFibrosis; automated

Running heads: Quantitative assessment for fibrosis, inflammation ballooning, and steatosis

Grant Support:

This work was supported by China-Singapore International S&T Cooperation Program (2016YFE0116800) and grants from the China National Science and Technology Major Project for Infectious Diseases Control during the 13th Five-Year Plan Period (2018ZX09201002-001-005) and National Natural Science Foundation of China (NSFC) (81870407). QMA, DT, EB, MRG and SP are collaborators in the EPoS (Elucidating Pathways of Steatohepatitis) consortium “European NAFLD Registry” funded by the Horizon 2020 Framework Program of the European Union under Grant Agreement 634413. This research work was supported by the Newcastle NIHR Biomedical Research Centre and also supported by a project collaboration agreement between Singapore General Hospital, Diagnostics Development (DxD) Hub of Exploit Technologies Pte Ltd, Singapore and HistoIndex Pte Ltd, Singapore. Collaborating parties invested contributions to enable the research.

† Co-corresponding Author:

Lai Wei, MD

Director and Professor

Peking University Hepatology Institute, Peking University People's Hospital Beijing

Key Laboratory of Hepatitis C and Immunotherapy for Liver Diseases,

No.11, Xizhimen South Street, Beijing 100044, China.

Tel: +8610-88325730.

E-mail: weilai@pkuph.edu.cn

Associate Professor Jason Chang Pik Eu, MBBS, MRCP

Department of Gastroenterology and Hepatology

Singapore General Hospital

Outram Road, Singapore 169608.

Tel:+65 63214684

E-mail: jason.chang@singhealth.com.sg

Prof Quentin M. Anstee PhD, FRCP

Institute of Cellular Medicine, Faculty of Medical Sciences, Newcastle University,
4th Floor, William Leech Building, Framlington Place, Newcastle upon Tyne NE2
4HH, UK; and at the Liver Unit, Newcastle Upon Tyne Hospitals NHS Trust,
Freeman Hospital, Freeman Road, High Heaton, Newcastle upon Tyne NE7 7DN,
UK.

Tel: +44 (0) 191 208 7012

E-mail: quentin.anstee@newcastle.ac.uk

Authors' contribution:

Study design: LW, PEJC, GBBG, QMA and DT

Liver biopsy pathology assessment: DT, AW, JMZ, WQL, WWK and LKH

Data collection: FL, LW, PEJC, GBBG, TCK, JMZ, HYR, XXW, QW, WQL, WWK,
LKH, QMA, DT, RGM, PS, BE

Data interpretation: LW, FL, AW

Statistical analysis: FL, GBBG, PEJC

Manuscript writing: FL, LW, GBBG, PEJC, QMA, DT, SAH

Critical revision of the manuscript: LW, PEJC, GBBG, TCK, QMA, AW, DT, RGM, PS,
BE, WQL, SAH, WWK and LKH

Disclosures:

LW is consulting for Abbvie, Allergan/Tobira, Novartis Pharma AG, Pfizer Ltd., Galmed, Gilead, HistoIndex,, Janssen. His institution has received research grant funding from Abbvie, BMS, and Gilead.

QMA is coordinator of the H2020 EPoS and IMI2 LITMUS consortia. His institution has received research grant funding from Abbvie, Allergan/Tobira, AstraZeneca, GlaxoSmithKline, Novartis Pharma AG, Pfizer Ltd., Vertex; has served as a consultant to Abbott Laboratories, Allergan/Tobira, E3Bio, Eli Lilly & Company Ltd., Galmed, Genfit SA, Gilead, Grunthal, HistoIndex, Imperial Innovations, Intercept., Inventiva, Janssen, Kenes, Madrigal, MedImmune, NewGene, NGMBio, Novartis,

Novo Nordisk, Pfizer, Raptor Pharma, Servier; and he has received lecture fees from Abbott Laboratories, Allergan/Tobira, BMS, Clinical Care Options, Falk, Genfit SA, Gilead.

SAH is has served as a consultant to Echosens, Allergan, Metacrine, Perspectum, Prometheus, Galmed, CiVi Biopharma, Corcept, Madrigal, Pfizer, NGM Bio, BMS, Gilead, Intercept, HistoIndex, Cirius, Axcella, Genfit, Novo Nordisk, Novartis, PPD, Medpace, IQVIA, Cymabay, Chronic Liver Disease Foundation, Innovate, Albireo, Hightide, Terns, Consynance, Galectin, Second Genome, Akero. He served as a speaker or a member of a speaker's bureau: Abbvie, Alexion. He has received grants for clinical research from: Gilead, Intercept, Genfit, Cirius, NGM Bio, Novo Nordisk, Novartis, Galmed, Immuron, Galectin, Madrigal, Conatus, Pfizer, Tobira/Allergan, Cymabay

DT has served as consultant to Intercept and has received grants from Histoindex.

The other authors have nothing to disclose.

ABSTRACT

Background: Nonalcoholic steatohepatitis (NASH) is a common cause of chronic liver disease. Clinical trials use the NASH CRN system for semiquantitative histological assessment of disease severity. Interobserver variability may hamper histological assessment and diagnostic consensus is not always achieved. We evaluate a novel second harmonic generation/two photon excitation fluorescence (SHG/TPEF) imaging-based tool to provide an automated quantitative assessment of histological features pertinent to NASH.

Method: Images were acquired by SHG/TPEF from 219 NAFLD/NASH liver biopsy samples from 7 centers in Asia and Europe. These were used to develop and validate *qFIBS*: a computational algorithm that quantifies key histological features of NASH.

Results: *qFIBS* was developed based on *in silico* analysis of selected signature parameters for four cardinal histopathological features, i.e. fibrosis (*qFibrosis*), inflammation (*qInflammation*), hepatocyte ballooning (*qBallooning*) and steatosis (*qSteatosis*), treating each as a continuous rather than categorical variable. Automated *qFIBS* analysis outputs showed strong correlation with each respective component of the NASH CRN scoring ($P<0.001$) [*qFibrosis* ($r=0.776$), *qInflammation* ($r=0.557$), *qBallooning* ($r=0.533$) and *qSteatosis* ($r=0.802$)] and high AUROC values [*qFibrosis* (0.870-0.951) (95%CI, 0.787–1.000; $P<0.001$), *qInflammation* (0.820-0.838) (95%CI, 0.726-0.933; $P<0.001$), *qBallooning* (0.813-0.844) (95%CI, 0.708-0.957; $P<0.001$) and *qSteatosis* (0.939-0.986) (95%CI, 0.867-1.000; $P<0.001$)] and was able to distinguish differing grades/stages of histological disease. Performance of *qFIBS* was best when assessing degree of steatosis and fibrosis, but performed less well when distinguishing severe inflammation and higher ballooning grades.

Conclusions: *qFIBS* is an automated tool that accurately quantifies the critical components of NASH histological assessment. It offers a relatively simple tool that could potentially aid reproducibility and standardization of liver biopsy assessments required for NASH therapeutic clinical trials.

INTRODUCTION

Nonalcoholic fatty liver disease (NAFLD) is an important cause of chronic liver

disease worldwide, with an estimated prevalence reported to be between 20 to 30% in North America, northern Europe, Australia, Japan, India and China ^[1-2]. NAFLD encompasses a wide spectrum of clinical histological phenotypes, ranging from pure steatosis without significant liver injury to nonalcoholic steatohepatitis (NASH); a constellation of features exemplifying hepatocellular injury and inflammation, with or without fibrosis or cirrhosis, occurring in the absence of significant alcohol intake. NASH is a major cause of liver-related morbidity and mortality ^[3,4], and has become one of the leading indications for liver transplantation ^[5-8]. However, at present, there is no licensed or universally accepted liver-specific therapy for NASH. Accordingly, there remains a pressing need to develop new effective therapeutics.

Despite recent advances in the field, the development pathway for effective therapies in NASH is complex ^[9,10]. Well-defined endpoints are essential if therapeutic NASH trials are to robustly assess efficacy of treatment. These endpoints need to be accurate, reproducible and clinically meaningful ^[11-13]. Although non-invasive biomarkers, including imaging techniques, are promising tools to assess disease severity in NAFLD, none have yet been validated or qualified as trial endpoints and none can accurately assess grade of steatosis, ballooning of hepatocytes or liver inflammation. Therefore, histological assessment remains the, albeit imperfect, reference standard for diagnosing NASH and continues to play an integral role in clinical practice ^[14].

Several semiquantitative histological scoring systems to diagnose and grade NASH have been proposed ^[15-18], the most commonly used are the NASH Clinical Research Network (CRN) “NAFLD activity score” (NAS) and the FLIP “Steatosis, inflammatory Activity and Fibrosis score” (SAF). Whilst these scoring systems do differ in important ways, they all measure four key histological characteristics: steatosis, hepatocellular injury (hepatocyte ballooning), lobular inflammation and fibrosis. Although these scoring systems have been used in clinical trials ^[19-21], there remain concerns regarding potential interobserver variability and absence of clear diagnostic

consensus that are accentuated when these tools are applied by non-specialist hepatopathologists. In addition, current scoring systems only provide a non-linear, semiquantitative or categorical assessment of disease ^[18]. This may limit precision and granularity of data, particularly in the context of subtle changes with therapy. As such, there is a need for standardized, continuous quantitative scales that intuitively may perform better, especially when measuring interval changes and therapeutic responses. Therefore, there is an imperative to establish new, accurate and automated quantitative scoring systems that encompasses all key histopathological characteristics of NASH.

With recent advancement in ultra-fast lasers, intrinsic optical signals like auto-fluorescence from biological samples could be imaged by using second harmonic generation (SHG) and two-photon excited fluorescence (TPEF) with unstained slide. Wang et al ^[22] first established a SHG-based quantification of fibrosis-related parameters (Q-FP) model in NAFLD that offered a more refined assessment of collagen architectural changes along a continuous, quantitative and scale. Liu et al ^[23] and Chang et al ^[24] subsequently described the use of SHG to auto-quantify collagen deposition and assess fibrosis in pediatric and adult patients with NAFLD. To further explore the utility and potential of SHG in the context of NAFLD, we carried out an international, multicenter study to develop and validate a novel SHG-based automated quantitative evaluation tool “qFIBS”, for the evaluation of fibrosis as well as lobular inflammation, ballooning and steatosis, with the goal of increasing the precision of disease severity assessment following the confirmation of a NAFLD/NASH diagnosis by an expert hepatopathologist^[22-25]. This present study comprises a large multicenter cohort that includes patients from different ethnicities, age ranges (pediatric to adult), and varying severity of NAFLD, from simple steatosis through to cirrhosis.

MATERIALS AND METHODS

Patients and Tissue Preparation

Liver tissue from 219 adult patients (≥ 18 years old) with biopsy-proven NAFLD/NASH were included in the current study. Cases were recruited at seven international specialist centers (Peking University People's Hospital, China; Beijing 302 Hospital, China; Singapore General Hospital, Singapore; Freeman Hospital, Newcastle upon Tyne, UK; University of Turin, Italy; University of Palermo, Italy; and Virgen del Rocio University Hospital, Spain). Patients with liver disease of other etiologies, such as alcoholic or drug-induced liver disease, autoimmune liver disease, viral hepatitis, cholestatic or genetic liver diseases were excluded. A separate cohort of pediatric NAFLD cases were also assessed as part of the validation study (Supplementary data method). This study was approved by the Ethics Committee of all the participating hospitals.

The 219 samples were sectioned into 4-5 μm thick for SHG-imaging and stained with H&E and Masson trichrome for histological assessment. Clinical and pathological characteristics were obtained from the patients' medical records. Liver histology from all cases was jointly assessed by three expert hepatopathologists (DT, AW, WQL) and consensus scores determined according to the NASH CRN System (NAS = sum of steatosis scored 0–3, ballooning 0–2, and lobular inflammation 0–3) ^[15]. Liver fibrosis was scored as 0 - 4 using the NASH CRN scoring system, with 4 indicating cirrhosis ^[15,16]. NAFLD was defined by the presence of $\geq 5\%$ macrovesicular or mixed steatosis. NASH was defined as $\text{NAS} \geq 4$ with a score of at least 1 in each individual component of the score ^[25]. The minimum length of liver biopsy specimens was 10 mm.

Image acquisition

All imaging was conducted by trained technicians on identical equipment (Genesis® system, HistoIndex Pte. Ltd., Singapore) according to a standardized operating procedure. Of the 219 samples, 69 samples were imaged in China, 108 samples in

Singapore, and 42 samples from European centers were imaged in the UK. This process utilized SHG microscopy to visualize collagen and TPEF microscopy to visualize the other histological structures (i.e. inflammation, hepatocyte ballooning and steatosis). The same image acquisition parameters were used for the samples at all the centers. The samples were laser-excited at 780 nm, SHG signals were recorded at 390 nm, and TPEF signals were recorded at 550 nm. Image tiles were acquired at 20X magnification with 512 X 512 pixel resolution with a dimension of 200 X 200 μm^2 . Multiple adjacent image tiles were captured to encompass the whole tissue area in each slide.

Establishment & Validation of qFIBS (qFibrosis, qInflammation, qBallooning and qSteatosis)

The 219 samples were assigned *a priori* using stratified randomization by an experienced statistician to one of two groups: two-thirds (146 samples) were assigned to a Training group and the remaining third (73 samples) to a Validation group (Fig. 1).

Using the NASH CRN scoring system ^[15] as the reference standard, automated measures of fibrosis, inflammation, hepatocyte ballooning and steatosis were developed in the Training group and validated in the Validation group. The sequential procedure for establishing the four indices comprised: (1) **detection** of collagen, inflammatory cells, ballooned cells and fat vacuoles in different regions of the lobules (Fig. 2); (2) **quantification** of defined architectural parameters that were characteristic of each histopathological feature (Fig.1); (3) selection of the most significant parameters (Fig.1); and (4) **model construction**, combination of parameters into a single “signature” index for each of the four histological components. These were individually termed qFibrosis, qInflammation, qBallooning and qSteatosis (Fig. 1). The severity of all NASH pathological characteristics was calculated with the four established indices, together comprising the qFIBS index. Detailed descriptions of the protocols are provided in Supplementary Materials and

Methods.

Statistical analysis

The Spearman nonparametric method was used to estimate the correlation between the quantified indices (qFibrosis, qInflammation, qBallooning and qSteatosis) and NASH CRN defined semiquantitative pathological categories. The area under the receiver operating characteristic (AUROC) analysis was performed to evaluate the accuracy of the quantified indices for prediction of the different scores of histological characteristics. The cutoff values were determined by Youden's index. The sensitivity and specificity of the four indices for prediction of the different scores were calculated. Statistical significance level was set at $P < 0.05$. Statistical analyses were done with MATLAB R2015a (MathWorks, USA).

RESULTS

Summary of clinical data

The demographic, biochemical, and histological characteristics of patients with biopsy-proven NAFLD are summarized in Table 1. The Training and Validation groups were comparable with respect to demographics, laboratory and histological data. There were no significant differences between the two groups in terms of age, gender, ethnicity, body mass index (BMI), and prevalence of diabetes mellitus and hypertension. There were also no statistically significant differences with regards to serum alanine transaminase (ALT), aspartate transaminase (AST), gamma-glutamyl transferase (GGT), alkaline phosphatase (ALP), total bilirubin, albumin, fasting glucose, triglycerides and platelet count. Similarly, the fibrosis stage, inflammation, ballooning and steatosis scores were evenly distributed between the two groups.

Fibrosis, inflammation, ballooning and steatosis assessments

Liver fibrosis, inflammation, hepatocyte ballooning and steatosis severity were quantified using the SHG/TPEF system (Fig. 3). All parameters were quantified across the whole slides. 128 collagen and 45 steatosis parameters were quantified at the central vein (CV), portal tract (PT) and perisinusoidal (PS) regions, 63 inflammation and 39 ballooning parameters were quantified at the PT and lobular (CV and PS) regions. The details of all parameters measured are included in the supplementary results.

Establishment of the models for fully quantitative assessments of qFIBS component models and dynamic ranges

The number of parameters required to optimally assess each histological feature was refined during the algorithm training process in the discovery cohort in order to develop the four separate component models comprising qFIBS. Ultimately, qFibrosis included 17 parameters, qInflammation required 25 parameters, qBallooning needed 13 parameters, and qSteatosis just 5 parameters in the respective models. Each model outputs a numerical index that indicates the severity of the corresponding histological characteristic. The value ranges of the four indices

are positive integers ranging between 0 - 6.55 for qFibrosis, 0.68 - 3.38 for qInflammation, 0.22 - 1.96 for qBallooning, and 0.83 - 2.18 for qSteatosis (Supplementary Table 9).

The comparison of H&E and Masson trichrome staining with SHG/TPEF images of liver biopsy tissue at different scales of scoring and the quantitative values of qFibrosis, qInflammation, qBallooning and qSteatosis, are illustrated (Supplementary Fig. 1 - 4). Each qFIBS component was strongly correlated with the relevant NASH CRN component ($P<0.001$ for fibrosis, lobular inflammation, ballooning, and steatosis). (Supplementary Tables 5 - 8).

Validation of qFIBS models demonstrates correlation with histologically determined NASH CRN scores

The performance of the qFIBS models were next assessed in a separate validation cohort (n=73 cases), results of these analyses are summarized in Figure 4 with a detailed breakdown of performance characteristics (sensitivity, specificity, Positive predictive value (PPV) and negative predictive value (NPV)) provided in Table 2.

qFibrosis, which is principally determined by degree of zone 3 perisinusoidal fibrosis, correlated strongly with histological fibrosis stage ($r=0.776$) and could accurately differentiate fibrosis stages, area under curve (AUC) >0.87 (0.870 - 0.951) (95% confidence interval (CI), 0.754 - 1.000) (Supplementary Table 5, Fig.4 and Table 2). Similarly, qInflammation increased with histologically determined inflammation grade ($r=0.557$) and could differentiate grades of inflammation, AUC >0.82 (0.820 - 0.838) (95% CI, 0.726 – 0.933) (Supplementary Table 6, Fig.4 and Table 2). qBallooning, which was principally determined by features localized to the portal tract and lobular areas, also increased with histological ballooning score ($r=0.533$) and could distinguish different grades of ballooning, AUC >0.813 (0.813 - 0.844) (95% CI, 0.708 - 0.957) (Supplementary Table 7, Fig.4 and Table 2). qSteatosis exhibited the highest level of correlation with histological steatosis score ($r=0.802$) and performed especially well, differentiating grades of steatosis with very high levels of accuracy,

AUC > 0.939 (0.939 - 0.986) (95% CI, 0.867 – 1.000) (Supplementary Table 8, Fig.4 and Table 2).

To demonstrate the generalizable nature of the qFIBS analysis, further validation in a pediatric NAFLD/NASH cohort was also undertaken. The AUC values of performance results in adult-pediatric groups were Fibrosis, 0.807 - 0.961; qInflammation, 0.803 - 0.855; qBallooning, 0.810 - 0.834; qSteatosis, 0.961 - 0.976. These results were comparable with the adult-only analysis (see Supplementary Table 10).

DISCUSSION

We report the development and validation of a novel *in silico* process for automated NAFLD/NASH biopsy evaluation (qFIBS), that standardizes the histological assessment of NASH severity following confirmation of diagnosis by an expert heptaopathologist. At present, histological assessment of severity at baseline, and subsequent change in response to intervention, are assessed by expert histopathologists and remains an essential endpoint in Phase II/III clinical trials for NASH [27, 28]. The majority of studies use the NASH CRN system for semiquantitative histological assessment to grade disease activity and stage fibrosis [10,11]. However, inter- and intra-observer variability may impair the accuracy and reproducibility of these histological assessments. Indeed, clear diagnostic consensus regarding the presence or absence of steatohepatitis among pathologists is not always feasible [15,17, 29]. These issues serve to highlight the need for a more standardized, accurate and precise approach to grading and staging NASH in liver biopsies once the diagnosis of NAFLD/NASH has been established. They also imply a paradox that makes addressing this need all the more challenging: the histological reference standard for NASH, against which any novel tool is assessed, is itself inherently imperfect and unable to produce a completely error-free classification with respect to the presence or absence, or severity, of the target condition or feature. Although not unique to liver histopathology, such situations are methodologically challenging to address [30].

Notable issues with the traditional histological scores relate to the nature of semiquantitative grading systems employed by both NAS and SAF, which conflate anatomical distribution of specific lesions with features of severity in order to facilitate data capture by the human eye. These scores necessitate that continuous measures be assigned into discrete categorical grading bins [26-32], a constraint that inevitably leads to discrepancies related to inter- and intra-observer judgement, especially with cases at the margins between two categories. It also produces a blunting of sensitivity to detect change as semiquantitative grades may fail to report clinically

relevant but modest alterations in severity that do not transition across a predefined but arbitrary categorical boundary. This phenomenon is well illustrated by the breadth of disease that is encompassed by stage F3 fibrosis in the Brunt/Kleiner NASH CRN classification. In F3, anatomical portal-portal and/or portal-central bridging is the defining feature, yet no weight is given to density of collagen deposition or consequent effect on lobular ultrastructure. Intuitively, quantitative assessment of changes on a continuous scale in liver lesions should provide a more tractable and sensitive reference when dissecting disease progression and regression in NASH.

In longitudinal studies, the strongest histological long-term prognostic factor has been repeatedly found to be stage of liver fibrosis, with patients with advanced fibrosis (stages F3 and F4 [cirrhosis]) having the highest rates of liver-related and all-cause morbidity and mortality ^[3,33,34]. Therefore, presence of advanced fibrosis and especially cirrhosis is considered a generally-accepted surrogate for liver-related outcomes in NASH trials. Staging of fibrosis based on the NASH CRN underestimates perisinusoidal fibrosis and, as alluded to earlier, does not provide a distinction between biopsies with rare or short septa from those with numerous septa (unlike the distinction between F2 and F3 with the METAVIR score for chronic hepatitis). This is a limitation of the current reference standard and means there is no clear-cut threshold to differentiate significant fibrosis from those with advanced fibrosis. Analysis of biopsies during the development of the qFibrosis has established that zone 3 perisinusoidal fibrosis is highly correlated with classical fibrosis stage and that measurement of zone 3 fibrosis facilitates accurate discrimination of fibrosis stages, whilst also providing a continuous assessment of fibrosis severity. Wang et al has shown SHG can differentiate subtle differences between fibrosis stages 1a, 1b, 1c (NASH-CRN system), and differences in zonal distribution of fibrosis in cirrhotic patients ^[22]. If validated in long term studies against hard clinical outcomes, these data suggest that qFibrosis may be used to assess fibrosis change as a primary outcome and could potentially provide the clearest answer to the question of clinically relevant therapeutic response.

When assessing grade of disease, the NAS comprises the sum of 3 discrete semiquantitative parameters: steatosis, lobular inflammation and hepatocyte ballooning ^[20,35]. The validity of this combined score has previously been questioned by the histopathology specialist community as it is heavily weighted towards degree of steatosis and so fails to adequately distinguish differences in disease activity from amount of steatosis ^[27,30]. We employed TPEF to accurately assess the tissue microstructure and autofluorescent properties when the biopsy samples are irradiated with an exogenous or endogenous laser ^[36,37]. Based on TPEF signals and quantification, three separate parameters are described: qInflammation, qBallooning and qSteatosis. Whilst not perfect, each is focused upon a given histological feature and showed a good correlation with the relevant NAS component with high AUROC values. That some PPV/NPV values when discriminating between intermediate grades of ballooning or inflammation (Table 2) were modest may, at least in part, be due to classification errors in the reference standard as, within the classic accuracy concept, any discrepancy is assumed to be a failure of novel test and therefore a “false” positive/negative result ^[30].

Similar efforts on automated biopsy analysis has been made in preclinical models of NASH using the open-source pathology software QuPath ^[38]. Several techniques have been employed to automate steatosis measurement, however many require the use of specialized histological stains that are not compatible with routine formalin fixed paraffin embedded (FFPE) processed samples combined with digital image analysis ^[39]. Automatic quantification of lobular inflammation and ballooning on digital images of H&E stained slides has also been reported, but AUC values were poor (0.45 for lobular inflammation and 0.46 for ballooning) ^[40]. Others have demonstrated that liver fibrosis could be quantified using a range of techniques such as collagen proportionate area ^[41] and dual-photon microscopy ^[22-24,42] but these studies have tended to be in mixed etiology cohorts, or derived from single centers, and histological characteristics other than fibrosis were not assessed. In contrast, the

current study derives NAFLD/NASH cases associated samples from a large international multicenter collaboration and so provides an opportunity to not only comprehensively assess key histological features of NASH, but also to demonstrate utility across differing ethnicities, environmental/nutritional backgrounds and age groups (adult and pediatric disease).

Although the automated evaluation system - qFIBS is very precise and highly reproducible, it deserves further commentary. By using unstained slides to further eliminate potential variation caused by staining, qFIBS could most accurately distinguish different grades of steatosis and different stages of fibrosis, but was less accurate in distinguishing severe inflammation and higher ballooning grades. While the initial data is indeed promising, the algorithm will require further development and refinement. For steatosis, the system only measures empty spaces since fat has been lost during the processing technique. The size of the fat vacuole may have relevance in the clinical evolution or improvement of NASH. Histologically, the pathologist can distinguish between macrovesicular steatosis comprising large and small droplet fat from microvesicular steatosis. However, the system currently recognizes predominantly large droplet fat that pushes the hepatocyte nucleus to one side. Furthermore, the difference in the definitions of mild, moderate and severe steatosis is a limitation of the automated technique ^[43-45]. Correlation with more testing features such as hepatocyte ballooning remains moderate. Hepatocytes exhibiting ballooning change may not necessarily be enlarged and, separately, ballooning cannot always be clearly differentiated from feathery degeneration. Despite extensive cross-validation, double-reporting and consensus building amongst the expert pathologists, there remained inter-observer variability that limited the fidelity of the reference standard.

In the context of inflammation, NASH-type lobular inflammation is difficult to distinguish from inflammation from other etiologies as the system currently cannot distinguish the various inflammatory cell types. Therefore, the technology is not well

suited to establishing a *de novo* diagnosis but rather is a tool with which to robustly quantify disease severity once the diagnosis of NAFLD has been established. While qFIBS offers several novel applications for assessment of NASH histology, this technology does not replace the need for invasive liver biopsy and cannot replace the key role of the histopathologist in clinical practice. The information provided by qFIBS is limited to the specific assessment of fibrosis, lobular inflammation, ballooning and steatosis components and does not provide a diagnostic assessment of the liver biopsy specimen. In other words, qFIBS like NASH CRN scoring system is to be applied only after a histological diagnosis of NASH has been made based on established criteria including the pattern of tissue injury. qFIBS is thus most suitable as an adjunct tool that provides additional value to the efficiency and reliability of the histopathologist in clinical practice. The primary utility of qFIBS may be in standardizing histological assessments in therapeutic NASH trials, many of which require serial liver biopsies for assessment of improvement in the various phenotypes of NASH. In order to demonstrate sensitivity for detecting a therapeutic response, the qFIBS tool will require further validation in paired biopsy samples before and after treatment for NASH. The automated fibrosis scoring provided by SHG microscopy helps to minimize interobserver variability and, as has been demonstrated in the current study, permits image acquisition to occur at widely distributed clinical sites rather than necessitating shipping of samples for central histology review. Most importantly, by expression in an automated continuous scale, qFIBS may provide more precise quantification that allows for detection of subtle interval nuances that may not be so apparent using traditional approach.

In conclusion, our study demonstrates that it is feasible to develop a system using SHG/TPEF to automatically quantify the four cardinal features needed to characterize NASH, namely, fibrosis, lobular inflammation, hepatocyte ballooning and steatosis. These findings are significant as accurate continuous measurements are potentially more useful than semiquantitative scores to measure NASH activity and quantify patients' response to therapeutics used in clinical trials or patient care.

This automated quantification system has the potential to support translational and clinical research as it addresses several unmet needs for precision medicine.

Acknowledgments

The authors thank Dean C. S. Tai for his constructive comments on technical aspects of this study and Ya-Yun Ren for performing data analysis.

References

1. Younossi ZM, Koenig AB, Abdelatif D, Fazel Y, Henry L, Wymer M. Global epidemiology of nonalcoholic fatty liver disease-Meta-analytic assessment of prevalence, incidence, and outcomes. *Hepatology*. 2016; 64:73-84.
2. Younossi Z, Anstee QM, Marietti M, Hardy T, Henry L, Eslam M, et al. Global burden of NAFLD and NASH: trends, predictions, risk factors and prevention. *Nat Rev Gastroenterol Hepatol* 2018; 15:11-20.
3. Dulai PS, Singh S, Patel J, Soni M, Prokop LJ, Younossi Z, et al. Increased risk of mortality by fibrosis stage in nonalcoholic fatty liver disease: Systematic review and meta-analysis. *Hepatology*. 2017; 65:1557-1565.
4. Charlton MR, Burns JM, Pedersen RA, Watt KD, Heimbach JK, Dierkhising RA. Frequency and outcomes of liver transplantation for nonalcoholic steatohepatitis in the United States. *Gastroenterology*. 2011; 141: 1249- 1253.
5. Wong RJ, Cheung R, Ahmed A. Nonalcoholic steatohepatitis is the most rapidly growing indication for liver transplantation in patients with hepatocellular carcinoma in the U.S. *Hepatology*. 2014; 59:2188-95.
6. Kemmer N, Neff GW, Franco E, Osman-Mohammed H, Leone J, Parkinson E, et al. Nonalcoholic fatty liver disease epidemic and its implications for liver transplantation. *Transplantation*. 2013; 96:860–862.
7. Wong RJ, Aguilar M, Cheung R, Perumpail RB, Harrison SA, Younossi ZM, et al. Nonalcoholic steatohepatitis is the second leading etiology of liver disease among adults awaiting liver transplantation in the United States. *Gastroenterology* 2015; 148: 547-555.
8. Younossi ZM, Loomba R, Anstee QM, Rinella ME, Bugianesi E, Marchesini G, et al. Diagnostic modalities for nonalcoholic fatty liver disease, nonalcoholic steatohepatitis, and associated fibrosis. *Hepatology*. 2018; 68:349-360.
9. Rotman Y, Sanyal AJ. Current and upcoming pharmacotherapy for non-alcoholic fatty liver disease. *Gut*. 2017; 66:180-190.

10. Sanyal AJ, Brunt EM, Kleiner DE, Kowdley KV, Chalasani N, Lavine JE, et al. Endpoints and clinical trial design for nonalcoholic steatohepatitis. *Hepatology*. 2011; 54:344-353.
11. Sanyal AJ, Friedman SL, McCullough AJ, Dimick-Santos L; American Association for the Study of Liver Diseases; United States Food and Drug Administration. Challenges and opportunities in drug and biomarker development for nonalcoholic steatohepatitis: findings and recommendations from an American Association for the Study of Liver Diseases-U.S. Food and Drug Administration Joint Workshop. *Hepatology*. 2015; 61:1392-1405.
12. Patel YA, Imperial JC, Muir AJ, Anstee QM, DeBrotta D, Dimick-Santos L, et al; Liver Forum's Data Standardization Working Group. Baseline Parameters in Clinical Trials for Nonalcoholic Steatohepatitis: Recommendations from the Liver Forum. *Gastroenterology*. 2017; 153:621-625.
13. Caussy C, Reeder SB, Sirlin CB, Loomba R. Non-invasive, quantitative assessment of liver fat by MRI-PDFF as an endpoint in NASH trials. *Hepatology*. 2018 68(2):763-772.
14. Burt AD, Lackner C, Tiniakos DG. Diagnosis and assessment of NAFLD: Definitions and Histopathological Classification. *Semin Liver Dis* 2015; 35(3):207-20.
15. Kleiner DE, Brunt EM, Van Natta M, Behling C, Contos MJ, Cummings OW, et al. Design and validation of a histological scoring system for nonalcoholic fatty liver disease. *Hepatology*. 2005; 41: 1313-1321.
16. Brunt EM, Janney CG, Di Bisceglie AM, Neuschwander-Tetri BA, Bacon BR. Nonalcoholic steatohepatitis: a proposal for grading and staging the histological lesions. *Am J Gastroenterol*. 1999; 94:2467-2474.
17. Bedossa P, FLIP Consortium. Utility and appropriateness of the fatty liver inhibition of progression (FLIP) algorithm and steatosis, activity, and fibrosis (SAF) score in the evaluation of biopsies of nonalcoholic fatty liver disease. *Hepatology* 2014; 60: 565-575.
18. Brunt EM. Nonalcoholic fatty liver disease and the ongoing role of liver biopsy evaluation. *Hepatol Commun*. 2017;1: 370-378.

19. Bedossa P. Diagnosis of non-alcoholic fatty liver disease/ non- alcoholic steatohepatitis: Why liver biopsy is essential. *Liver Int.* 2018 Feb; 38 Suppl1:64-66.
20. Brunt EM, Kleiner DE. Challenges in the hepatic histopathology in non-alcoholic fatty liver disease. *Gut.* 2017; 66:1539-1540.
21. Bedossa P. Pathology of non-alcoholic fatty liver disease. *Liver Int.* 2017;37 Suppl 1:85 - 89.
22. Wang Y, Vincent R, Yang J, Asgharpour A, Liang X, Idowu MO, et al. Dual-photon microscopy-based quantitation of fibrosis-related parameters (q-FP) to model disease progression in steatohepatitis. *Hepatology.* 2017; 65:1891-1903.
23. Liu F, Zhao JM, Rao HY, Yu WM, Zhang W, Theise ND, et al. Second Harmonic Generation Reveals Subtle Fibrosis Differences in Adult and Pediatric Nonalcoholic Fatty Liver Disease. *Am J Clin Pathol.* 2017; 148:502-512.
24. Chang PE, Goh GBB, Leow WQ, Shen L, Lim KH, Tan CK. Second harmonic generation microscopy provides accurate automated staging of liver fibrosis in patients with non-alcoholic fatty liver disease. *PLoS One.* 2018; 13: e0199166.
25. Brunt EM. Nonalcoholic Fatty Liver Disease: Pros and Cons of Histologic Systems of Evaluation. *Int J Mol Sci.* 2016;17(1). pii: E97.
26. European Association for the Study of the Liver (EASL); European Association for the Study of Diabetes (EASD); European Association for the Study of Obesity (EASO). EASL-EASD-EASO Clinical Practice Guidelines for the management of non-alcoholic fatty liver disease. *J Hepatol.* 2016;64:1388-1402.
27. Siddiqui MS, Harrison SA, Abdelmalek MF, Anstee QM, Bedossa P, Castera L, et al. Case definitions for inclusion and analysis of endpoints in clinical trials for nonalcoholic steatohepatitis through the lens of regulatory science. *Hepatology.* 2018; 67:2001-2012
28. Sanyal AJ, Friedman SL, McCullough AJ, Dimick-Santos L, American Association for the Study of Liver D, United States F, Drug A. Challenges and opportunities in drug and biomarker development for nonalcoholic steatohepatitis: findings and recommendations from an American Association for the Study of Liver Diseases-

- U.S. Food and Drug Administration Joint Workshop. *Hepatology* 2015;61:1392-1405.
29. Bedossa P, Poitou C, Veyrie N, Bouillot JL, Basdevant A, Paradis V, et al. Histopathological algorithm and scoring system for evaluation of liver lesions in morbidly obese patients. *Hepatology* 2012;56:1751-1759.
 30. Rutjes AW, Reitsma JB, Coomarasamy A, Khan KS, Bossuyt PM. Evaluation of diagnostic tests when there is no gold standard. A review of methods. *Health Technol Assess* 2007;11:iii, ix-51.
 31. Kleiner DE, Bedossa P. Liver histology and clinical trials for nonalcoholic steatohepatitis-perspectives from 2 pathologists. *Gastroenterology*. 2015; 149: 1305-1308.
 32. Filozof C, Chow SC, Dimick-Santos L, Chen YF, Williams RN, Goldstein BJ, et al. Clinical endpoints and adaptive clinical trials in precirrhotic nonalcoholic steatohepatitis: Facilitating development approaches for an emerging epidemic. *Hepatol Commun*. 2017;1:577-585.
 33. **Ekstedt M, Hagstrom H**, Nasr P, Fredrikson M, Stal P, Kechagias S, et al. Fibrosis stage is the strongest predictor for disease-specific mortality in NAFLD after up to 33 years of follow-up. *Hepatology* 2015;61:1547-1554.
 34. Angulo P, Kleiner DE, Dam-Larsen S, Adams LA, Bjornsson ES, Charatcharoenwitthaya P, et al. Liver Fibrosis, but No Other Histologic Features, Is Associated With Long-term Outcomes of Patients With Nonalcoholic Fatty Liver Disease. *Gastroenterology* 2015;149:389-397.
 35. Kleiner DE. Histopathology, grading and staging of nonalcoholic fatty liver disease. *Minerva Gastroenterol Dietol*. 2018; 64:28-38.
 36. Wang S, Du H, Lin B, Liao C, Zhu X, Wang X, et al. Spatial and temporal identification of cerebral infarctions based on multiphoton microscopic imaging. *Biomed Opt Express*. 2018; 9:2312-2325.
 37. Thomas G, van Voskuilen J, Gerritsen HC, Sterenborg HJ. Advances and challenges in label-free nonlinear optical imaging using two-photon excitation

- fluorescence and second harmonic generation for cancer research. *J Photochem Photobiol B*. 2014; 141: 128-138.
38. Levene AP, Kudo H, Armstrong MJ, Thursz MR, Gedroyc WM, Anstee QM, et al. Quantifying hepatic steatosis - more than meets the eye. *Histopathology* 2012; 60: 971-981.
 39. E. Rexhepaj, D. Nathalie, N. Benoit, H. Alison, B. Carole, R. Walczak, S. Megnien, D. Hum, B. Staels, J. Brozek. A deep-learning approach for pattern recognition allows rapid and reproducible quantification of histological NASH parameters: Integration into the QuPath platform. *J Hepatol*. 2018; 68(S1):S123.
 40. Vanderbeck S, Bockhorst J, Kleiner D, Komorowski R, Chalasani N, Gawrieh S. Automatic quantification of lobular inflammation and hepatocyte ballooning in nonalcoholic fatty liver disease liver biopsies. *Hum Pathol*. 2015; 46:767-775.
 41. Tsochatzis E, Bruno S, Isgro G, Hall A, Theocharidou E, Manousou P, et al. Collagen proportionate area is superior to other histological methods for subclassifying cirrhosis and determining prognosis. *J Hepatol* 2014; 60: 948-954.
 42. Pirhonen J, Arola J, Sädevirta S, Luukkonen P, Karppinen SM, Pihlajaniemi T, et al. Continuous Grading of Early Fibrosis in NAFLD Using Label-Free Imaging: A Proof-of-Concept Study. *PLoS One*. 2016; 25;11: e0147804.
 43. **Nativ NI, Chen AI**, Yarmush G, Henry SD, Lefkowitz JH, Klein KM, et al. Automated image analysis method for detecting and quantifying macrovesicular steatosis in hematoxylin and eosin-stained histology images of human livers. *Liver Transplant*. 2014; 20: 228-236
 44. Vanderbeck S, J. Bockhorst, R. Komorowski, D.E. Kleiner, S. Gawrieh. Automatic classification of white regions in liver biopsies by supervised machine learning. *Hum. Pathol*. 2014; 45:785-792
 45. Goceri E, Shah ZK, Layman R, Jiang X, Gurcan MN. Quantification of liver fat: A comprehensive review. *Comput Biol Med*. 2016;71: 174- 189.

Figure Legend

Figure 1. Flowchart for the development of the four components of qFIBS (qFibrosis, qInflammation, qBallooning, qSteatosis). Abbreviations: BJ, BeiJing Center; SGH, Singapore General Hospital; EU, European centers.

Figure 2. Schematic illustration of the detection algorithms of fibrosis, inflammation, ballooning and steatosis in livers with nonalcoholic steatohepatitis.

Figure 3. Liver biopsy with NASH: (A) and (B) Steatosis with ballooning of hepatocytes, mild portal inflammatory cell infiltrates, and perisinusoidal fibrosis (Hematoxylin & Eosin and Masson). (C) Raw image of same area highlighting collagen fibers green (SHG/TPEF). (D) Segmented image with differential labeling of collagen in portal tract (PT) (yellow), central vein (CV) (purple) and perisinusoidal (PS) (blue) regions (SHG/TPEF). (E) Inflammation in the same respective areas was correspondingly labeled (SHG/TPEF). (F) Ballooning of hepatocytes was scattered (SHG/TPEF). (G) Fat vacuoles (FV) were distributed throughout the regions (SHG/TPEF).

Figure 4. Box-Whisker plots of values of each qFIBS component distribution relative to the NASH CRN Score component category in the adult Validation Group. Results for the Validation cohort are shown. For each stage/score/grade, maximum and minimum values are indicated by horizontal lines at the bottom and top of each stage/score/grade entry, the white box in the middle represents data points in the 25% to 75% interquartile range, and the line through the middle of the central white box represents the median value. Note: r value was calculated according to Spearman method.