Histopathology

Histopathology 2021, **79**, 791–800. DOI: 10.1111/his.14424

Diagnostic challenges of focal nodular hyperplasia: interobserver variability, accuracy, and the utility of glutamine synthetase immunohistochemistry

Daniel J Rowan,[1] Daniela S Allende,[2] Andrew M Bellizzi,[3] Ryan M Gill,[4] Xiuli Liu,[5] Catriona A McKenzie,[6,7] Roger K Moreira,[1] Taofic Mounajjed,[1] Samar Said,[1] Maria Westerhoff,[8] Sarah M Jenkins,[9] Kenneth P Batts,[10] Lawrence J Burgart,[10] Laura W Lamps[8] & Rondell P Graham[1]

[1]*Department of Laboratory Medicine and Pathology, Division of Anatomic Pathology, Mayo Clinic, Rochester, MN,* [2]*Department of Pathology, Cleveland Clinic, Cleveland, OH,* [3]*Department of Pathology, University of Iowa, Iowa City, IA,* [4]*Department of Pathology, University of California-San Francisco, San Francisco, CA,* [5]*Department of Pathology, Immunology, and Laboratory Medicine, University of Florida, Gainesville, FL, USA,* [6]*Department of Tissue Pathology and Diagnostic Oncology, Royal Prince Alfred Hospital, Camperdown,* [7]*New South Wales Health Pathology and Faculty of Medicine, University of Sydney, Sydney, Australia,* [8]*Department of Pathology, University of Michigan, Ann Arbor, MI,* [9]*Division of Biomedical Statistics and Informatics, Mayo Clinic, Rochester, MN, and* [10]*Hospital Pathology Associates and MNGI,Minneapolis, MN, USA*

Date of submission 22 February 2021
Accepted for publication 30 May 2021
Published online *Article Accepted* 3 June 2021

Rowan D J, Allende D S, Bellizzi A M, Gill R M, Liu X, McKenzie C A, Moreira R K, Mounajjed T, Said S, Westerhoff M, Jenkins S M, Batts K P, Burgart L J, Lamps L W & Graham R P (2021) *Histopathology* **79**, 791–800. https://doi.org/10.1111/his.14424

Diagnostic challenges of focal nodular hyperplasia: interobserver variability, accuracy, and the utility of glutamine synthetase immunohistochemistry

Aims: The diagnosis of focal nodular hyperplasia (FNH) and the interpretation of glutamine synthetase (GS) staining can be challenging on biopsies. We aimed to evaluate the reproducibility of needle biopsy diagnosis of FNH, the effect of GS immunohistochemistry on FNH diagnosis, and which histological features are most useful for the diagnosis of FNH.

Methods and results: The study included virtual needle biopsies generated from 75 resection specimens (30 FNHs, 15 hepatocellular adenomas, 15 hepatocellular carcinomas, and 15 non-lesional liver specimens). Pathologists were reasonably accurate (83.1%) in the diagnosis of FNH with haematoxylin and eosin alone. Ductular reaction and nodularity had the highest sensitivity for a diagnosis of FNH (88.1% and 82.2%, respectively), whereas central scar was the most specific feature (90.6%). The presence of two or more of the classic histological features had 89.6% sensitivity and 86.2% specificity for a diagnosis of FNH. Diagnostic accuracy was significantly higher with the addition of a GS stain. A map-like GS staining pattern was highly specific (99.3%) for FNH. However, GS staining was interpreted as non-map-like in 14.4% of reviews of true FNH cases, and overall interobserver agreement for interpretation of the GS staining pattern was only moderate (kappa = 0.42).

Conclusions: Pathologists are reasonably accurate in the diagnosis of FNH on virtual biopsies, and GS staining improves accuracy. However, a subset of FNH cases remain challenging. Steatosis and a pseudo-map-like GS staining pattern were associated with increased difficulty. Therefore, although a map-like GS staining pattern is useful for confirmation of a diagnosis, the lack of a map-like GS staining pattern on needle biopsy does not necessarily exclude a diagnosis of FNH.

Address for correspondence: Rondell P. Graham, MBBS, Department of Laboratory Medicine and Pathology, Mayo Clinic, 200 First Street SW, Rochester, MN 55905, USA. e-mail: graham.rondell@mayo.edu

Keywords: accuracy, focal nodular hyperplasia, glutamine synthetase, interobserver variability, liver, needle biopsy, virtual biopsy

Introduction

Focal nodular hyperplasia (FNH) is a benign, non-neoplastic liver lesion that is thought to be a hyperplastic response to abnormalities in hepatic blood flow.[1–3] It is the second most common benign liver lesion (after haemangioma),[3] and is most commonly diagnosed in female patients aged <40 years.[4] The diagnosis can often be made by imaging when the characteristic features of central scar surrounded by a homogeneous lesion without a capsule are identified,[5] and, because the lesion is benign and complications are rare, surgical resection is typically not required for asymptomatic cases.[6] In FNH cases that are not radiographically typical, or when there is a high degree of clinical suspicion for malignancy,[7] however, needle biopsy may be undertaken, as a definite diagnosis on biopsy allows for non-operative management.[8]

The characteristic histological features of FNH are nodular hepatocellular parenchyma with an absence of normal portal tracts, associated fibrous septa/central scar containing abnormal thick-walled blood vessels, and ductular reaction.[9] In some FNH cases, the presence of macrovesicular steatosis,[1] steatohepatitic features[10] or cholestatic features may contribute to diagnostic confusion. The histological differential diagnosis of FNH, particularly on biopsy specimens, includes hepatocellular adenoma (particularly the inflammatory subtype), well-differentiated hepatocellular carcinoma, cirrhosis, nodular regenerative hyperplasia, liver adjacent to other mass lesions (an FNH-like response has been described adjacent to other types of tumour)[11], steatohepatitis with centrizonal arteries,[12] and occasionally normal liver. A map-like glutamine synthetase (GS) immunohistochemical staining pattern is useful in supporting the diagnosis of FNH,[13] and has been shown to improve diagnostic accuracy.[14,15] However, recognition of the histological features and interpretation of the GS staining can be quite challenging on biopsy specimens, and the reproducibility of GS interpretation in this setting has not been specifically examined.[16,17] Additionally, the ideal gold standard for assessing the accuracy of biopsy interpretation would be to also examine paired resection specimens, but most studies that have investigated the utility of GS immunohistochemistry (IHC) have used resection specimens alone, or biopsies without paired resection specimens. Because FNH biopsies with subsequent paired resection specimens are difficult to find in large numbers, we generated virtual needle biopsy specimens from scanned whole-slide images from resection specimens.

The primary aims of this study were to evaluate the reproducibility of the diagnosis of FNH on virtual needle biopsy specimens; to evaluate the effect of GS IHC on FNH diagnosis and the reproducibility of GS interpretation; to examine which histological features are most useful for the diagnosis of FNH on needle biopsies; and to examine the effect of the number of core biopsies on diagnosis.

Materials and methods

CASE SELECTION AND IHC

The study included 75 resection specimens with the following diagnoses, which were retrieved from the case files of a single institution: FNH (n = 30), non-lesional liver without advanced fibrosis (n = 15), hepatocellular adenoma (n = 15) [inflammatory subtype (n = 8), hepatocyte nuclear factor 1A-inactivated subtype (n = 3), and unclassified subtype (n = 4)], and well-differentiated hepatocellular carcinoma (n = 15). GS staining was performed on a representative block from each case. GS (Millipore, Temecula, CA, USA; catalogue no. MAB302, clone GS-6) was used at a 1:2000 titre diluted with Ventana Antibody Diluent with Casein on the Ventana Benchmark Ultra (Roche, Indianapolis, Indiana, USA).

PREPARATION OF 'VIRTUAL NEEDLE BIOPSIES' AND REVIEW

One representative haematoxylin and eosin (H&E) slide and the corresponding GS slide from each resection specimen were scanned at ×40 magnification on the Aperio ScanScope AT2 brightfield instrument (Leica Biosystems, Buffalo Grove, Illinois, USA) at a resolution of 0.25 μm per pixel. The scan output of the ScanScope AT2 comprised 24-bit contiguous pyramid tiled TIFFs, with the digital slide file

(.svs) format being standard pyramid-tiled TIFFs with JPEG compression at a compression quality setting of 70. Virtual 18G needle biopsies were randomly generated from the scanned images with digital imaging software. By use of the grid overlay and EXTRACT REGION tools in IMAGESCOPE (Leica Biosystems), random full-resolution strip images (virtual needle biopsies), 0.1 cm in width and no longer than 3.1 cm, were manually created from the scans obtained with the AT2 Aperio scanner. For cases with paired H&E and GS slides, matching regions from the H&E and GS slides were extracted. The central *x*-axes and *y*-axes in the original scans were used for positional reference of the horizontal and vertical strips, respectively. Three liver pathologists (L.W.L., L.J.B., and K.P.B.) confirmed the diagnosis on the basis of whole slides from the resection specimen for each case, and confirmed that the virtual biopsies sampled the lesion in all cases. All three reviewers agreed with the original diagnosis in all cases, and their diagnoses were used as the gold standard. All cases contained lesional material within at least one virtual biopsy core. Of the 30 total FNH cases, two separate cases contained one core with lesion sampled and one core without lesion sampled. In two separate FNH cases, one core contained lesional material and the other core contained only the edge of the lesion.

STUDY DESIGN AND INTERPRETATION OF VIRTUAL NEEDLE BIOPSIES

Eight additional liver pathologists reviewed virtual biopsies from each case in four separate, independent rounds, as shown in Figure 1. In each round, reviewers were asked to select one diagnosis for each case from the following list: FNH, hepatocellular adenoma, hepatocellular carcinoma, normal liver, bile duct adenoma, cholangiocarcinoma, or descriptive. If the descriptive diagnostic category was selected, reviewers were asked to provide a descriptive diagnosis in free text format. Within the description, reviewers stated whether a specific diagnosis was favoured. Subsequently, the diagnoses and descriptions were reviewed, and the responses were grouped into categories as shown in Table S1 to allow for statistical analysis.

For each case in round 1, reviewers recorded whether each of the following histological features of FNH was present: central scar, bile ductular reaction, nodular hepatocellular parenchyma, and abnormal vessels. They also indicated whether they would request a GS stain for diagnosis, on the basis of review of the H&E slide.

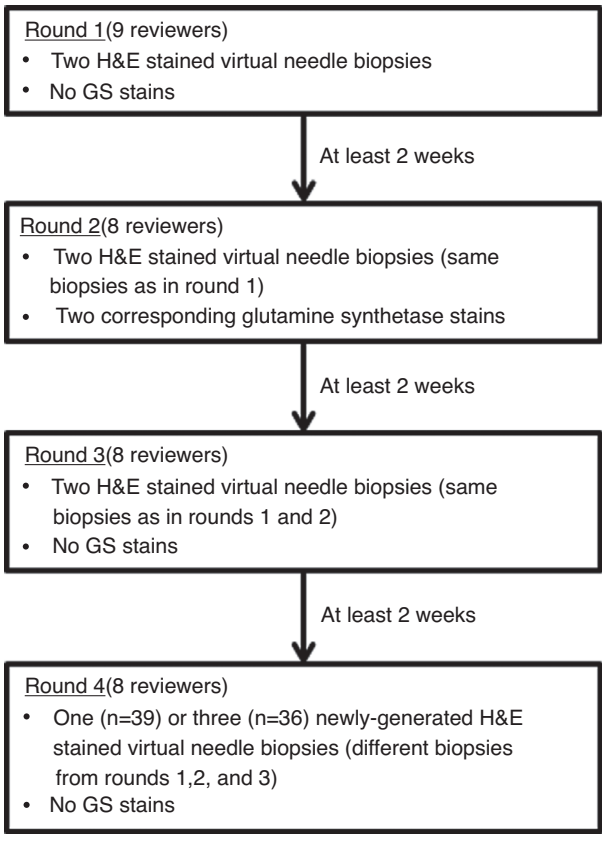

Figure 1. Study design. Pathologists reviewed slides in four separate, independent rounds. In round 2, the utility of the glutamine synthetase immunohistochemical stain was assessed. Round 4 allowed for assessment of the effect of the number of cores on diagnosis. [Colour figure can be viewed at wileyonlinelibrary.com]

STATISTICAL ANALYSIS

For calculation of diagnostic accuracy, definite or probable diagnoses for each diagnostic category were considered to represent agreement with the true diagnosis. For example, a reviewer's diagnosis of definite FNH or probable FNH was considered to represent a correct diagnosis for a true FNH case. The diagnostic accuracy (percentage correctly diagnosed as compared with the true diagnosis) was compared between rounds, as well as between specific characteristics within rounds, by the use of logistic regression models with generalised estimating equations (GEEs) to account for correlated data within reviewer. These models were applied to all cases, as well as to the subset of true FNH cases. In round 1, the percentages of cases classified with histological features (central scar, ductular reaction, nodularity, and abnormal thick-walled vessels) were compared between true FNH cases and non-FNH

cases by the use of logistic regression models with GEEs. Among the true FNH cases, the nodular features were also compared between cases that reviewers classified as definite FNH, probable FNH, or non-FNH, by the use of logistic regression models with GEEs. The sensitivity (among the true FNH cases) and specificity (among the true non-FNH cases) of the histological features, of the number of histological features present in each review and for map-like GS staining patterns from round 2 were calculated. In round 2, the agreement among the reviewers with respect to map-like staining of the cases was quantified with the Fleiss kappa statistic (applicable for multiple raters), along with the 95% confidence interval (CI). Chi-square tests were used to test for differences between proportions. *P*-values of <0.05 were considered to be statistically significant. All analyses were performed with SAS version 9.4 (SAS Institute, Cary, NC, USA).

POST-REVIEW ANALYSIS OF FNH HISTOLOGY AND GS IHC

After the four study rounds, the FNH cases and associated GS stains were separately reviewed by two authors (D.J.R. and R.P.G.) to identify histological or immunohistochemical features that may have contributed to diagnostic difficulty or poor interobserver agreement.

Results

CLINICAL AND IMAGING CHARACTERISTICS OF FNH CASES

The FNH cases included in this study were resected for three main reasons: symptoms ($n = 19$), diagnostic uncertainty of imaging or biopsy ($n = 8$), and incidental resection because the patient was undergoing an unrelated surgical operation ($n = 3$) (one patient had a separate hepatocellular adenoma, one patient had a hepatic haemangioma, and one patient had a low-grade appendiceal mucinous neoplasm). The diagnostic impression on imaging for the 30 FNH cases was: consistent with FNH ($n = 18$), favour hepatocellular adenoma ($n = 4$), and indeterminate ($n = 8$). Six of the resected FNHs had been previously biopsied, with the following biopsy diagnoses: FNH ($n = 1$), hepatocellular adenoma ($n = 3$), and well-differentiated hepatocellular neoplasm ($n = 2$). After resection, no patients with FNH or hepatocellular adenoma had tumour recurrence.

ACCURACY OF H&E IN DIAGNOSING FNH

There were a total of 2475 reviews in the study by nine liver pathologists over four rounds (Table S2). FNH cases were accurately diagnosed as definite FNH in 78.4% of reviews and as probable or definite FNH in 83.1% of reviews with H&E alone (rounds 1, 3 and 4 combined) (Table 1). The 127 (16.9%) reviews of true FNH cases that were not recognised as definite or probable FNH with H&E were diagnosed as follows: descriptive diagnosis (47 reviews), benign lesion (45 reviews), definite or probable non-lesional tissue (21 reviews), definite or probable neoplasm (13 reviews), and possible hepatocellular carcinoma (one review). The diagnostic accuracy for true FNH cases was similar in rounds 1 and 3 (79.3% and 83.3%, respectively). The number of tissue cores present for evaluation (as assessed in round 4) did not significantly correlate with the diagnostic accuracy for true FNH cases. The diagnostic accuracies were 86.7% for FNH cases with one tissue core and 87.5% for those with three tissue cores ($P = 0.76$). Intraobserver agreement on the H&E diagnosis of FNH cases ranged from 40.0% to 96.7% among reviewers.

In FNH cases for which the radiological findings were diagnostic of FNH, the diagnostic accuracy of reviewers using H&E was significantly higher (86.9%) than that for FNH cases for which the radiological findings were indeterminate or favoured a neoplastic lesion (72.5%) ($P = 0.00005$).

HISTOLOGICAL FEATURES OF FNH

Central scar, ductular reaction, nodularity and abnormal vessels were all identified more frequently in true FNH cases than non-FNH cases ($P = 0.008$, $P = 0.003$, $P = 0.003$, and $P = 0.006$, respectively) (Figure S1). In reviews that correctly diagnosed FNH (definite or probable FNH), each of the four histological features were identified more frequently than in reviews that did not recognise FNH (central scar 61.2% versus 26.7%, $P < 0.00001$; ductular reaction 90.2% versus 80.4%, $P = 0.04$; nodularity 87.9% versus 60.7%, $P < 0.00001$; and abnormal vessels 75.7% versus 30.4%, $P < 0.00001$).

Among the four FNH cases that were most commonly missed with H&E alone, central scar and nodularity were identified significantly less frequently than in the remaining FNH cases (central scar 22.2% versus 59.0%, $P = 0.0004$; nodularity 36.1% versus 89.3%, $P \leq 0.00001$), whereas there were no significant differences between the frequency of ductular reaction (80.6% versus 89.3%, $P = 0.13$) and the

Table 1. Reviewer diagnoses for focal nodular hyperplasia (FNH) cases by round

Reviewer diagnoses (%)	Round 1 H&E	Round 2 H&E, GS	Round 3 H&E	Round 4 H&E	Combined rounds 1, 3, and 4 H&E
Definite FNH	74.4	88.3	77.1	84.2	78.4
Probable FNH	4.8	2.9	6.3	2.9	4.7
Benign lesion	9.3	0.4	6.7	1.7	6.0
Definite neoplastic	2.6	1.7	1.7	0.0	1.5
Probable neoplastic	0.0	0.8	0.4	0.4	0.3
Possible HCC	0.4	0.0	0.0	0.0	0.1
Non-lesional/normal liver	3.0	2.5	1.3	1.3	1.9
Probable non-lesional	1.9	0.4	0.4	0.4	0.9
Descriptive	3.7	2.9	6.3	9.2	6.3

GS, glutamine synthetase; H&E, haematoxylin and eosin; HCC, hepatocellular carcinoma.

frequency of abnormal vessels (52.7% versus 68.4%, $P = 0.07$).

When central scar, ductular reaction, nodularity and abnormal vessels were all identified, a definite or probable diagnosis of FNH was correctly made with H&E alone in 96.1% of reviews; in 84.3% with three features, 65.2% with two features, and 44.0% with one feature. Ductular reaction and nodularity were the histological features with the highest sensitivity for a diagnosis of FNH (88.1% and 82.2%, respectively) (Table 2). Central scar was the most specific feature for the diagnosis of FNH (specificity 90.6%). The presence of two or more of the above histological features had 89.6% sensitivity and 86.2% specificity for a diagnosis of FNH on biopsy.

Table 2. Sensitivity and specificity by histological feature and combination of histological features [round 1 data; round 2 for glutamine synthetase (GS)]

Feature(s)	Sensitivity (%)	Specificity (%)
Central scar	54.1	90.6
Ductular reaction	88.1	88.9
Nodularity	82.2	84.2
Abnormal vessels	66.2	76.8
All four histological features	38.2	99.8
Three or more histological features	64.1	96.5
Two or more histological features	89.6	86.2
One or more histological features	98.9	58.0
Map-like GS staining pattern	85.6	99.3

GS- glutamine synthetase

REVIEWER REQUESTS FOR GS STAINS IN TRUE FNH CASES

On the basis of review of H&E-stained needle biopsies in round 1, reviewers requested a GS stain for diagnosis in 84.1% (227/270) (reviewer range: 33.3–100%) of reviews of true FNH cases. A GS stain was requested in 80.6% (reviewer range: 17.8–100%) of true FNH cases that were diagnosed as definite FNH with H&E alone, in 92.3% of cases that were diagnosed as probable FNH, and in 94.6% of cases that were not recognised as FNH.

Among the four reviewers who did not request a GS stain for all FNH cases, identification of a central scar or abnormal thick-walled vessels significantly correlated with a decision not to order a GS stain. Central scar was identified in 60.5% of reviews in which no GS stain was requested and in 40.2% of reviews in which a GS stain was requested ($P = 0.04$). Abnormal vessels were identified in 94.7% of reviews in which no GS stain was requested and in 58.5% of reviews in which a GS stain was requested ($P = 0.00006$). There was no significant correlation between the presence of ductular reaction or nodularity and the decision to order a GS stain.

Table 3. Diagnostic accuracy by round

Accuracy (%)	Round 1 H&E Two cores	Round 2 H&E, GS Two cores	Round 3 H&E Two cores	Round 4 H&E All cases	Round 4 H&E One core	Round 4 H&E Three cores
Overall accuracy	82.2	92.5	86.2	85.2	85.9	84.4
Accuracy for true FNH cases	79.3	91.3	83.3	87.1	86.7	87.5

FNH, focal nodular hyperplasia; GS, glutamine synthetase; H&E, haematoxylin and eosin.

UTILITY AND REPRODUCIBILITY OF GS STAINING INTERPRETATION

The overall diagnostic accuracies in rounds 1 and 3 (H&E slide-only rounds) were 82.2% and 86.2%, respectively. The overall diagnostic accuracy in round 2 (H&E and GS stain round) was 92.5%, which was higher than the accuracy in rounds 1 and 3 ($P = 0.02$ and $P = 0.04$, respectively). With regard to FNH, the diagnostic accuracies for true FNH cases were 79.3% in round 1 and 83.3% in round 3, as compared with 91.3% in round 2 ($P = 0.08$ and $P = 0.14$, respectively) (Table 3).

The GS staining pattern was interpreted as map-like in 85.6% of reviews of true FNH cases. Reviewers unanimously agreed on interpretation of GS staining in 20 of the 30 FNH cases. Twenty-six (14.4%) reviews of GS staining in true FNH cases interpreted the staining pattern as non-map-like. The reviewer interpretations could be classified into three groups: (i) increased/diffuse pattern (10 reviews); (ii) patchy staining pattern (eight reviews); and (iii) normal/non-specific pattern (two reviews). The reviewer diagnoses for these cases are shown in Table S3. There was one FNH case in which the GS stain failed, as shown by no perivenular staining in the adjacent normal liver, but was interpreted as negative (six reviews), rather than failed. There was a moderate level of interobserver agreement on the interpretation of the GS staining pattern as map-like in cases of true FNH (kappa = 0.42, 95% CI 0.14–0.71).

In only one non-FNH case was a map-like GS staining pattern interpretation proffered. This case was an inflammatory-type hepatocellular adenoma, and the GS staining pattern was interpreted as map-like by two separate reviewers (Figure 2). The GS staining pattern in all other non-FNH cases was interpreted as non-map-like.

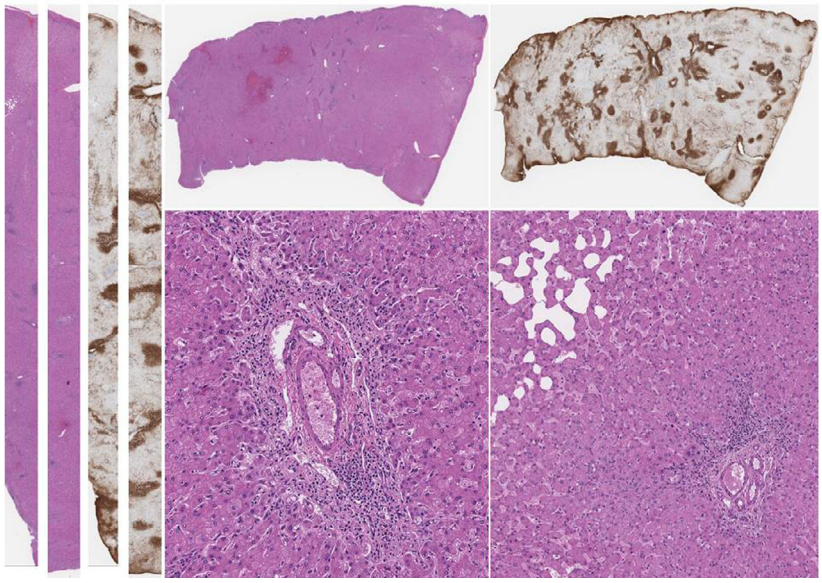

Figure 2. Hepatocellular adenoma (inflammatory subtype) in which two reviewers interpreted the glutamine synthetase (GS) staining pattern as map-like. Other reviewers described the GS staining as increased, negative, focal, and 'weird'.

Discussion

The diagnosis of FNH on needle biopsy specimens can be challenging, even for experienced liver pathologists, because the characteristic histological features are not always sampled, and, as discussed above, the differential diagnosis is fairly broad. Several studies have shown that a map-like GS staining pattern is a sensitive and specific finding supporting a diagnosis of FNH, primarily in resection specimens.[13–15] However, this has not been studied extensively in needle biopsy specimens, or in needle biopsy specimens paired with subsequent resection specimens. The ability to diagnose FNH with confidence on needle biopsy specimens has significant clinical implications, as this is the decision point at which future surgical intervention can be avoided. Additionally, prior studies on FNH and GS staining interpretation have relied on expert consensus opinion as the gold standard diagnosis, rather than a comparison of needle biopsies with resection specimens. Thus, we sought to determine the reproducibility of FNH diagnosis and GS staining interpretation among a multi-institutional group of pathologists with experience in liver pathology.

This study was carried out with virtual needle biopsies that were digitally generated by the use of whole-slide images of slides from resection specimens to mimic 18G needle biopsy specimens. This approach had the major advantage of allowing the paired resection specimen to serve as the gold standard for diagnosis. This methodology also had the advantage of allowing the generation of multiple different virtual needle core biopsy specimens from each scanned resection, producing more virtual biopsies for evaluation to examine the effect of the number of core biopsies on diagnosis. A similar approach has previously been used to evaluate sampling variability in liver fibrosis.[18]

Our findings support the concept that FNH can be confidently diagnosed with H&E alone on many needle biopsy specimens, but also reinforce the fact that the diagnosis is difficult in approximately 15–20% of cases. Of reviews of FNH needle biopsies, 83.1% were correctly diagnosed as definite or probable FNH with H&E alone (78.4% were called definite FNH and 4.7% were called probable FNH). Interestingly, the diagnostic accuracy obtained with H&E was significantly lower among FNH cases that were indeterminate according to radiological features (despite reviewers being blinded to the radiology findings).

When reviewers had access to both H&E and GS slides, the diagnostic accuracy for true FNH cases increased to 91.3%. A map-like GS staining pattern, when present, was highly specific (99.3%) for FNH. However, the overall interobserver agreement for interpretation of the GS staining pattern was only moderate (kappa = 0.42). Although there was unanimous agreement among reviewers on the interpretation of GS staining in 20 FNH cases, there was disagreement on the interpretation of GS staining among reviewers for the remaining 10 FNH cases. Among these cases, the non-map-like GS staining patterns described by reviewers were: patchy (eight reviews), increased but not definitely map-like (seven reviews), negative (six reviews, all the same case), diffuse (three reviews), and non-specific (one review). In the four true FNH cases for which the majority opinion was that the GS staining pattern was non-map-like, there was lower diagnostic accuracy [the round 2 (with GS) accuracy was 59.4%; the round 1 and round 3 (without GS) accuracy was 83.2%], showing that, in some FNH cases, the lack of a clear map-like GS staining pattern may draw pathologists away from a diagnosis of FNH.

A review of GS staining for the FNH cases in this study with the poorest agreement among reviewers highlighted two features that contributed to disagreement. First, the presence of ≥20% macrovesicular steatosis within the lesional cells appeared to contribute to poor agreement. Three true FNH cases had ≥20% macrovesicular steatosis, which we believe contributed to varied interpretations, owing to the effect of the steatosis on the GS staining pattern (Figure 3). The presence of focal or absent GS expression in FNH cases in the presence of prominent steatosis has been previously described.[13] Second, a 'pseudo-map-like' GS staining pattern was the other contributor to poor agreement in four cases. FNH cases with a pseudo-map-like staining pattern on needle biopsy had less intense GS staining, more focal or narrower anastomosing areas, or expression limited to peripheral parts of the tumour nodules (Figure 4). These patterns are similar to those described by Joseph et al., who coined the term 'pseudo-map-like'.[16] In their study, they noted a pseudo-map-like GS staining pattern in four of 24 FNH biopsy cases, and we noted this pattern in a somewhat similar proportion of our cases (four of 30 cases). The cause of this variant pattern of expression is unknown. It is also important to note that the pseudo-map-like GS staining pattern is not specific for FNH, and a similar pattern was seen in one inflammatory hepatocellular adenoma (Figure 2) in our study and in 15% of hepatocellular adenomas in the study by Joseph et al.

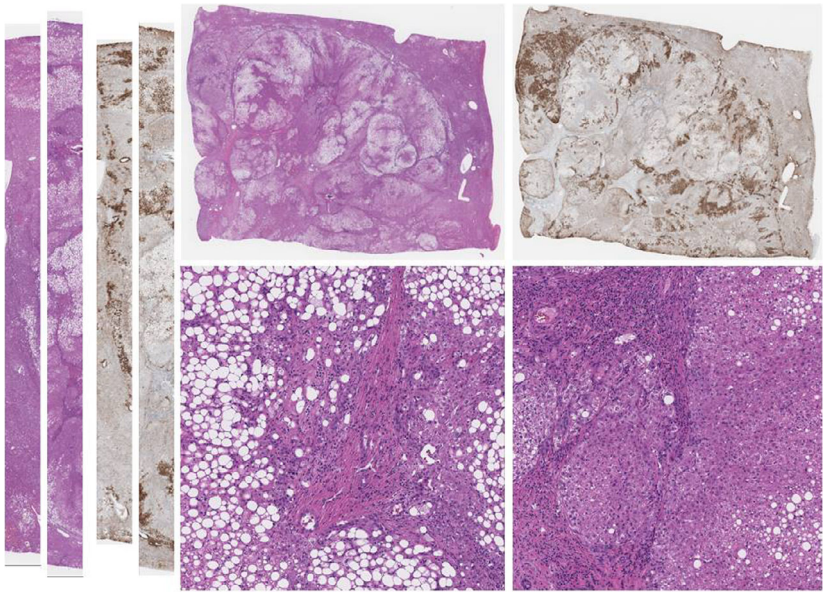

Figure 3. A focal nodular hyperplasia (FNH) case with prominent steatosis and high interobserver variation in the interpretation of glutamine synthetase (GS) staining. Three reviewers described the GS staining pattern as map-like, and three reviewers described GS staining as patchy. Using H&E staining alone, 22 of 25 reviewers diagnosed the case as definite or probable FNH. In round 2, with H&E and GS staining, all reviewers diagnosed the case as probable or definite FNH.

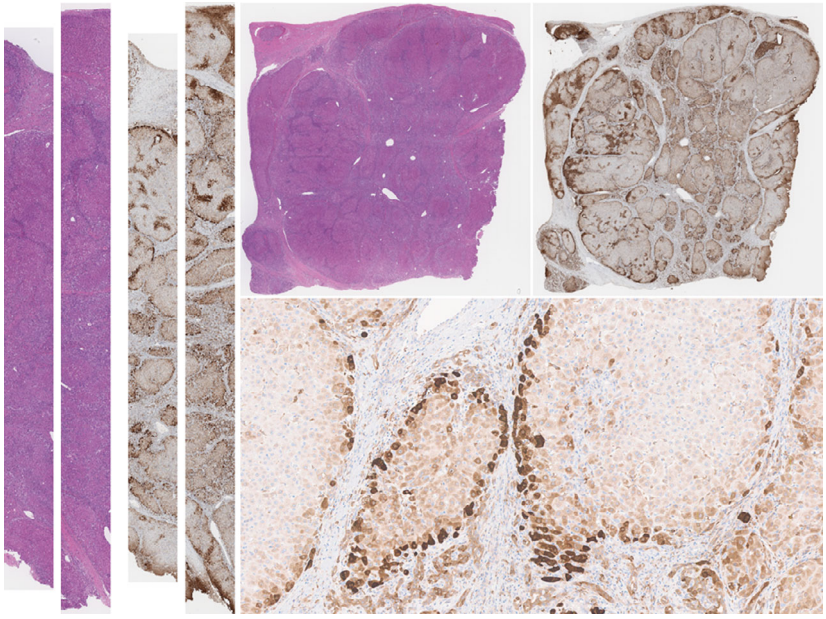

Figure 4. A focal nodular hyperplasia case showing a 'pseudo-map-like' glutamine synthetase (GS) staining pattern on needle biopsy. This pattern was characterised by less intense GS staining, more focal or narrower anastomosing areas, and expression limited to peripheral parts of the tumour nodules.

Central scar, ductular reaction, nodularity and abnormal vessels were all identified more frequently in true FNH cases than in non-FNH cases ($P = 0.008$, $P = 0.003$, $P = 0.003$, and $P = 0.006$, respectively). The two most specific histological features were central scar and ductular reaction, whereas the two most sensitive were ductular reaction and nodularity (Table 2). On the basis of these

findings, we propose that hepatocellular lesions with at least three of the above morphological features shown by H&E staining can be considered to be probable FNH, and those with two features can be considered to be possibly FNH. The presence of one or fewer features is non-diagnostic of FNH.

This study used virtual biopsies taken from resection specimens in order to have an optimal gold standard diagnosis. This study therefore allowed us to simulate a major clinical decision point at which a needle biopsy diagnosis of FNH would probably lead to non-operative management. With this approach, our data also provide information on an important quality indicator, which is how many cores are needed for a diagnosis of FNH. In this study, a higher number of tissue cores (from one to three) did not significantly correlate with an increase in diagnostic accuracy ($P = 0.76$). A study of a larger number of cases would be needed to examine and validate the effect of the number of biopsy cores. The fact that the virtually generated biopsies may be of higher quality (lacking crush artefact and fragmentation) than true 'real-life' needle biopsies is a limitation of this study. The virtual biopsies were randomly selected from the whole-slide image in order to more closely simulate the fact that liver biopsies may contain entirely lesional tissue, the edge of a lesion, or mostly perilesional tissue.

In summary, this study demonstrates that expert liver pathologists are reasonably accurate (83.1%) in the diagnosis of FNH on virtual needle biopsies. However, we reaffirmed that liver pathologists may not recognise 15–20% of FNH cases with H&E alone. The vast majority of cases (98.9%) showed one or more of the classic histological features, including ductular reaction (88.1%), nodularity (82.2%), abnormal vessels (66.2%), and central scar (54.1%). Furthermore, the presence of two or more of these histological features had 89.6% sensitivity and 86.2% specificity for a diagnosis of FNH, and thus the surgical pathologist can at least favour FNH in this setting. The overall diagnostic accuracy is improved with the use of GS IHC on needle biopsy specimens, with recognition of a map-like GS staining pattern having very high specificity (99.3%) for FNH. Therefore, we recommend the use of a GS stain in diagnostically challenging cases. However, we observed that there is significant interobserver variation among liver pathologists in the interpretation of GS staining on needle biopsy specimens in a minority of cases. Up to 15% of FNHs failed to show a map-like GS staining pattern, most often because of steatosis or pseudo-map-like GS staining, and this should not dissuade one from making a diagnosis of FNH in the appropriate H&E context.

Acknowledgements

The study was designed by RP Graham. Pathology review was conducted by DJ Rowan, DS Allende, AM Bellizzi, RM Gill, X Liu, CA McKenzie, RK Moreira, T Mounajjed, S Said, M Westerhoff, KP Batts, LJ Burgart, and LW Lamps. Data analysis and interpretation was performed by DJ Rowan, SM Jenkins, and RP Graham. The manuscript was written by DJ Rowan. All authors edited and approved the final draft of the manuscript.

Conflicts of interest

The authors state that they have no conflicts of interest.

Funding

No funding was received for this study.

Author contributions

The study was designed by R. P. Graham. Pathology review was performed by D. J. Rowan, D. S. Allende, A. M. Bellizzi, R. M. Gill, X. Liu, C. A. McKenzie, R. K. Moreira, T. Mounajjed, S. Said, M. Westerhoff, K. P. Batts, L. J. Burgart, and L. W. Lamps. Data analysis and interpretation was performed by D. J. Rowan, S. M. Jenkins, and R. P. Graham. The manuscript was written by D. J. Rowan. All authors edited and approved the final draft of the manuscript.

References

1. Nguyen BN, Flejou JF, Terris B, Belghiti J, Degott C. Focal nodular hyperplasia of the liver: a comprehensive pathologic study of 305 lesions and recognition of new histologic forms. *Am. J. Surg. Pathol.* 1999; **23**; 1441–1454.
2. Wanless IR, Mawdsley C, Adams R. On the pathogenesis of focal nodular hyperplasia of the liver. *Hepatology* 1985; **5**; 1194–1200.
3. Wanless IR, Albrecht S, Bilbao J *et al*. Multiple focal nodular hyperplasia of the liver associated with vascular malformations of various organs and neoplasia of the brain—a new syndrome. *Mod. Pathol.* 1989; **2**; 456–462.
4. Kerlin P, Davis GL, Mcgill DB, Weiland LH, Adson MA, Sheedy PF. Hepatic adenoma and focal nodular hyperplasia—clinical, pathologic, and radiologic features. *Gastroenterology* 1983; **84**; 994–1002.

5. Grazioli L, Morana G, Kirchin MA, Schneider G. Accurate differentiation of focal nodular hyperplasia from hepatic adenoma at gadobenate dimeglumine-enhanced MR imaging: prospective study. *Radiology* 2005; **236**; 166–177.
6. Demarco MP, Shen P, Bradley RF, Levine EA. Intraperitoneal hemorrhage in a patient with hepatic focal nodular hyperplasia. *Am. Surg.* 2006; **72**; 555–559.
7. Kumar P, Gill RM, Phelps A, Tulpule A, Matthay K, Nicolaides T. Surveillance screening in Li-Fraumeni syndrome: raising awareness of false positives. *Cureus* 2018; **10**; e2527.
8. Makhtouf HR, Abdul-Al HM, Goodman ZD. Diagnosis of focal nodular hyperplasia of the liver by needle biopsy. *Hum. Pathol.* 2005; **36**; 1210–1216.
9. Fabre A, Audet P, Vilgrain V *et al*. Histologic scoring of liver biopsy in focal nodular hyperplasia with atypical presentation. *Hepatology* 2002; **35**; 414–420.
10. Deniz K, Moreira RK, Yeh MM, Ferrell LD. Steatohepatitis-like changes in focal nodular hyperplasia, a finding to distinguish from steatohepatitic variant of hepatocellular carcinoma. *Am. J. Surg. Pathol.* 2017; **41**; 277–281.
11. Bryant BH, Zenali MJ, Swanson PE *et al*. Glutamine synthetase immunoreactivity in peritumoral hyperplasia in liver: case report of a metastatic paraganglioma with focal nodular hyperplasia-like changes and review of an additional 54 liver masses. *Am. J. Clin. Pathol.* 2016; **146**; 254–261.
12. Gill RM, Belt P, Wilson L, Bass NM, Ferrell LD. Centrizonal arteries and microvessels in nonalcoholic steatohepatitis. *Am. J. Surg. Pathol.* 2011; **35**; 1400–1404.
13. Bioulac-Sage P, Laumonier H, Rullier A *et al*. Over-expression of glutamine synthetase in focal nodular hyperplasia: a novel easy diagnostic tool in surgical pathology. *Liver Int.* 2009; **29**; 459–465.
14. Bioulac-Sage P, Cubel G, Taouji S *et al*. Immunohistochemical markers on needle biopsies are helpful for the diagnosis of focal nodular hyperplasia and hepatocellular adenoma subtypes. *Am. J. Surg. Pathol.* 2012; **36**; 1691–1699.
15. Tsai JH, Jeng YM, Pan CC, Lu SW, Kuo YJ. Immunostaining of glutamine synthetase is a sensitive and specific marker for diagnosing focal nodular hyperplasia in needle biopsy. *Pathology* 2012; **44**; 605–610.
16. Joseph NM, Ferrell LD, Jain D *et al*. Diagnostic utility and limitations of glutamine synthetase and serum amyloid-associated protein immunohistochemistry in the distinction of focal nodular hyperplasia and inflammatory hepatocellular adenoma. *Mod. Pathol.* 2014; **27**; 62–72.
17. Bioulac-Sage P, Sempoux C, Balabaud C. Immunohistochemical pitfalls in the diagnosis of focal nodular hyperplasia and inflammatory hepatocellular adenoma. *Clin. Res. Hepatol. Gastroenterol.* 2014; **38**; 245–249.
18. Bedossa P, Dargere D, Paradis V. Sampling variability of liver fibrosis in chronic hepatitis C. *Hepatology* 2003; **38**; 1449–1457.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Figure S1. Histological features of FNH.

Table S1. Definitions used for classification of reviewer diagnoses into diagnostic categories.

Table S2. Number of cases and reviews by round for each diagnostic category.

Table S3. Reviewer diagnoses for FNH cases in which the reviewer did not interpret the GS staining pattern as map-like.