DR DANIEL J ROWAN (Orcid ID : 0000-0001-7830-2525)

DR KENNETH BATTS (Orcid ID : 0000-0002-2963-3843)

Article type : Original Article

Diagnostic Challenges of Focal Nodular Hyperplasia: Interobserver Variability, Accuracy, and the Utility of Glutamine Synthetase Immunohistochemistry

Running Title: Diagnostic Challenges of Focal Nodular Hyperplasia

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

[6]Department of Tissue Pathology and Diagnostic Oncology, Royal Prince Alfred Hospital, Camperdown, Australia; New South Wales Health Pathology and Faculty of Medicine, University of Sydney;

[7]Department of Pathology, University of Michigan, Ann Arbor, MI;

[8]Division of Biomedical Statistics and Informatics, Mayo Clinic, Rochester, MN;

[9]Hospital Pathology Associates and MNGI, Minneapolis, MN

Corresponding Author: Rondell P. Graham, M.D.

Current Address:

Department of Laboratory Medicine and Pathology

Mayo Clinic

200 First Street SW

Rochester, MN 55905

Email: graham.rondell@mayo.edu

Word Count: 3570 words

Conflicts of Interest and Source of Funding: None

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.

Abstract

Aims: Diagnosis of focal nodular hyperplasia (FNH) and interpretation of glutamine synthetase (GS) stains 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 histologic features are most useful for diagnosis of FNH.

Methods and Result: The study included virtual needle biopsies generated from 75 resection specimens (30 FNHs, 15 hepatocellular adenomas, 15 hepatocellular carcinomas, and 15 non-lesional liver). Pathologists were reasonably accurate (83.1%) in diagnosis of FNH by H&E alone. Ductular reaction and nodularity had the highest sensitivity for a diagnosis of FNH (88.1% and 82.2%, respectively), while

central scar was the most specific feature (90.6%). The presence of ≥2 of the classic histologic features had 89.6% sensitivity and 86.2% specificity for diagnosis of FNH. Diagnostic accuracy was significantly higher with the addition of a GS stain. Map-like GS staining pattern was highly specific (99.3%) for FNH. However, GS was interpreted as non-map-like in 14.4% of reviews of true FNH cases and overall interobserver agreement for interpretation of GS staining pattern was only moderate (Kappa=0.42).

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

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

Introduction

Focal nodular hyperplasia (FNH) is a benign, non-neoplastic liver lesion thought to be a hyperplastic response to abnormalities in hepatic blood flow.[1-3] It is the second most common benign liver lesion (after hemangioma),[3] and is most commonly diagnosed in female patients under the age of 40 years.[4] The diagnosis can often be made by imaging when the characteristic features of central scar surrounded by a homogenous 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 cases of FNH 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 histologic 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 cases of FNH, the presence of macrovesicular steatosis[1], steatohepatitic features[10], or cholestatic features may contribute to diagnostic confusion. The histologic differential diagnosis of FNH, particularly in biopsy specimens, includes hepatocellular adenoma (particularly the inflammatory subtype), well-differentiated hepatocellular carcinoma, cirrhosis, nodular regenerative hyperplasia, liver adjacent to other mass lesions (FNH-like response has been described

adjacent to other types of tumors[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 histologic features and interpretation of the GS stain 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" to assess the accuracy of biopsy interpretation would be to also examine paired resection specimens, yet most studies that have investigated the utility of GS immunohistochemistry (IHC) are from 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 using virtual needle biopsy specimens; to evaluate the effect of GS immunohistochemistry on FNH diagnosis and the reproducibility of GS interpretation; to examine which histologic features are most useful for 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 Immunohistochemistry

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 (HNF1A)-inactivated subtype (n=3), and unclassified subtype (n=4)], and well-differentiated hepatocellular carcinoma (n=15). GS was performed on a representative block from each case. Glutamine Synthetase (Millipore, Temecula, CA, USA, catalog# MAB302, clone GS-6) was used at a 1/2,000 titer diluted with Ventana Antibody Diluent with Casein on the Ventana Benchmark Ultra.

Preparation of "Virtual Needle Biopsies" and Review

One representative hematoxylin and eosin (H&E) slide and the corresponding GS slide from each resection specimen were scanned at 40x magnification on the Aperio ScanScope AT2 brightfield instrument (Leica Biosystems) at a resolution of 0.25 microns per pixel. The scan output of the ScanScope AT2 was 24-bit contiguous pyramid tiled TIFFs, with the digital slide file (.svs) format being

standard pyramid tiled TIFFs having JPEG compression at a compression quality setting of 70. Virtual 18G needle biopsies were randomly generated from the scanned images using digital imaging software. Using the grid overlay and Extract Region tools in ImageScope (Leica Biosystems), random full resolution strip images (virtual needle biopsies), approximately 0.1cm in width and no longer than 3.1cm, were manually created from the scans obtained via 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 and y axes in the original scans were used for positional reference of the horizontal and vertical strips, respectively. Three liver pathologists (LWL, LJB and KPB) confirmed the diagnosis based on whole slides from the resection specimen for each case, and confirmed that the virtual biopsies sampled the lesion in all cases. All 3 reviewers agreed with the original diagnosis on 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, 2 separate cases contained 1 core with lesion and 1 core without lesion sampled. In 2 separate FNH cases 1 core contained lesional material and the other core only contained 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: focal nodular hyperplasia, 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 favored. Subsequently, the diagnoses and descriptions were reviewed and the responses were grouped into categories as shown in Supplemental Table 1 to allow for statistical analysis.

For each case in round 1, reviewers recorded whether each of the following histologic 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, based on review of the H&E slide.

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 to the true diagnosis) was compared between rounds, as well as between specific characteristics within rounds, using logistic regression models with generalized estimating equations (GEE) to account for correlated data within reviewer. These models were performed among all cases, as well as among the subset of true FNH cases. In round 1, the percentage of cases classified with histologic features (central scar, ductular reaction, nodularity, and abnormal thick-walled vessels) was compared between true FNH vs non-FNH cases using logistic regression models with GEE. Among the true FNH cases, the nodular features were also compared between cases reviewers classified as definite FNH, probable FNH, or non-FNH using logistic regression models with GEE. Sensitivity (among the true FNH cases) and specificity (among the true non-FNH cases) of the histologic features, number of histologic features present in each review, and for map-like GS staining pattern 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. Chi-square tests were used to test for differences between proportions. P-values less than 0.05 were considered statistically significant. All analyses were performed using SAS version 9.4 (SAS Institute Inc., Cary, NC).

Post Review Analysis of FNH Histology and GS Immunohistochemistry

After the four study rounds, the FNH cases and associated GS stains were separately reviewed by two authors (DJR and RPG) to identify histologic or immunohistochemical features which 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 by imaging or biopsy (n=8), and incidental resection because the patient was undergoing an unrelated surgery (n=3) (1 patient had a separate hepatocellular adenoma, 1 patient had a hepatic hemangioma, and 1 patient had a low-grade appendiceal mucinous neoplasm). The diagnostic impression on imaging for the 30 FNH cases was: consistent with FNH (n=18), favor 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 tumor recurrence.

Diagnostic Accuracy of FNH by H&E

There were a total of 2,475 reviews in the study by 9 liver pathologists over four rounds (Supplemental Table 2). FNH cases were accurately diagnosed as definite FNH in 78.4% of reviews and as probable or definite FNH in 83.1% of reviews by H&E alone (rounds 1, 3, and 4 combined) (Table 1). The 127 (16.9%) reviews of true FNH cases that were not recognized as definite or probable FNH on 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 HCC (1 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 diagnostic accuracy for true FNH cases. The diagnostic accuracy for FNH cases with one tissue core was 86.7% and for those with 3 tissue cores, the accuracy was 87.5% (p=0.76). Intraobserver agreement on H&E diagnosis of FNH cases ranged from 40.0% to 96.7% among reviewers.

In FNH cases where the radiologic findings were diagnostic of FNH, the diagnostic accuracy of reviewers on H&E was significantly higher (86.9%) compared to FNH cases where the radiology was indeterminate or favored a neoplastic lesion (72.5%) (p=0.00005).

Histologic Features of Focal Nodular Hyperplasia

Central scar, ductular reaction, nodularity, and abnormal vessels were all identified more frequently in true FNH cases than non-FNH cases (P= 0.008, 0.003, 0.003, and 0.006 respectively) (Supplemental Figure 1). In reviews correctly diagnosed as FNH (definite or probable FNH), each of the four histologic features were identified more frequently than in reviews not recognized as FNH (central scar 61.2% vs 26.7%, p=<0.00001; ductular reaction 90.2% vs 80.4%, p=0.04; nodularity 87.9% vs 60.7%, p=<0.00001; and abnormal vessels 75.7% vs 30.4%, p=<0.00001).

Among the four most commonly missed FNH cases by H&E alone, central scar and nodularity were identified significantly less frequently compared with the remaining FNH cases (central scar 22.2% vs 59.0%, p=0.0004; nodularity 36.1% vs 89.3%, p=<0.00001); while there was no significant difference

between the frequencies of ductular reaction (80.6% vs 89.3%, p=0.13) or abnormal vessels (52.7 % vs 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 on H&E alone in 96.1% of reviews; in 84.3% with three, 65.2% with two, and 44.0% with one feature. Ductular reaction and nodularity were the histologic 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 diagnosis of FNH (specificity 90.6%). The presence of ≥2 of the above histologic features had an 89.6% sensitivity and 86.2% specificity for a diagnosis of FNH on biopsy.

Reviewer Requests for Glutamine Synthetase Stains in True FNH Cases

Based on 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 by H&E alone, in 92.3% of cases that were diagnosed as probable FNH, and in 94.6% of cases that were not recognized as FNH.

Among the four reviewers who did not request a GS 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 where no GS was requested and in 40.2% of reviews when GS was requested (p=0.04). Abnormal vessels were identified in 94.7% of reviews where no GS was requested and in 58.5% of reviews when GS 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.

Utility and Reproducibility of Glutamine Synthetase Stain Interpretation

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

GS IHC was interpreted as map-like in 85.6% of reviews of true FNH cases. Reviewers unanimously agreed on interpretation of the GS stain in 20 of the 30 FNH cases. Twenty-six (14.4%) reviews of GS stains in true FNH cases were interpreted as non-map-like. The reviewer interpretations could be classified into three groups: 1) increased/diffuse pattern (10 reviews), 2) patchy staining pattern (8 reviews), and 3) normal/non-specific pattern (2 reviews). The reviewer diagnoses for these cases are shown in Supplemental Table 3. There was one FNH case in which the GS stain failed, as evidenced by no perivenular staining in the adjacent normal liver, but was interpreted as negative (6 reviews), rather than failed. There was a moderate level of interobserver agreement on the interpretation of GS stains for map-like staining pattern 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 interpretation proffered. This case was an inflammatory type hepatocellular adenoma and the GS stain was interpreted as map-like by two separate reviewers (Figure 2). All other non-FNH cases were interpreted to have a non-map-like GS staining pattern.

Discussion

The diagnosis of FNH on needle biopsy specimens can be challenging, even for experienced liver pathologists, because the characteristic histologic features are not always sampled, and, as discussed above, the differential diagnosis is fairly broad. Several studies have shown that map-like GS expression is a sensitive and specific finding supporting a diagnosis of FNH, primarily on resection specimens.[13-15] However, this has not been studied extensively on needle biopsy specimens, nor on 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 where future surgical intervention can be avoided. Additionally, prior studies on FNH and GS stain interpretation have relied on expert consensus opinion as the gold standard diagnosis, rather than comparison of needle biopsies to resection specimens. Thus, we sought to determine the reproducibility of FNH diagnosis and GS stain interpretation among a multi-institutional group of pathologists with experience in liver pathology.

This study was carried out using virtual needle biopsies that were digitally generated using whole slide images of slides from resection specimens to mimic 18G needle biopsy specimens. This approach had the major advantage of allowing for 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 by H&E alone on many needle biopsy specimens, but also reinforce that the diagnosis is difficult in approximately 15-20% of cases. 83.1% of reviews of FNH needle biopsies were correctly diagnosed as definite or probable FNH by H&E alone (78.4% were called definite FNH and 4.7% were called probable FNH). Interestingly, the diagnostic accuracy by H&E was significantly lower among FNH cases that were indeterminate based on radiologic features (despite reviewers being blinded to the radiology findings).

When reviewers had access to both H&E and GS stained 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 GS staining pattern was only moderate (Kappa=0.42). While there was unanimous agreement among reviewers on the interpretation of the GS stain in 20 FNH cases, there was disagreement on the interpretation of GS among reviewers in the remaining 10 FNH cases. Among these cases, the non-map-like GS staining patterns described by reviewers were: patchy (8 reviews), increased but not definitely map-like (7 reviews), negative (6 reviews, all the same case), diffuse (3 reviews) and non-specific pattern (1 review). In the four true FNH cases where the majority opinion was that the GS staining pattern was non-map like, there was a lower diagnostic accuracy (Round 2 [with GS] accuracy was 59.4%; Round 1 and 3 accuracy [without GS] was 83.2%) showing that in some FNH cases, the lack of a clear map-like staining pattern may mislead pathologists away from a diagnosis of FNH.

A review of GS stains from 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% or more macrovesicular steatosis which we believe contributed to varied interpretations due 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 4 cases. FNH cases with a pseudo map-like pattern on needle biopsy had less intense GS staining, more

focal or narrower anastomosing areas, or expression limited to peripheral parts of the tumor 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 pattern in 4 of 24 FNH biopsy cases and we noted that pattern in a somewhat similar proportion of our cases (4 of 30 cases). The cause of this variant pattern of expression is unknown. It is also important to note that the pseudo map-like 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.

Central scar, ductular reaction, nodularity, and abnormal vessels were all identified more frequently in true FNH cases than non-FNH cases (P= 0.008, 0.003, 0.003, and 0.006 respectively). The two most specific histologic features were central scar and ductular reaction, whereas the two most sensitive were ductular reaction and nodularity (Table 2). Based on these findings, we propose that hepatocellular lesions with at least three of the above morphologic features on H&E stain can be considered likely FNH and those with two features are possibly FNH. One or fewer features is non-diagnostic of FNH.

The design of 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 wherein a needle biopsy diagnosis of FNH would likely lead to non-operative management. From 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 artifact 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 peri-lesional tissue.

In summary, this study demonstrates that expert liver pathologists are reasonably accurate (83.1%) in the diagnosis of FNH on virtual needle biopsy. However, we reaffirmed that liver pathologists may not recognize 15-20% of FNH cases on H&E alone. The vast majority of cases (98.9%) demonstrate one or more of the classic histologic 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 histologic

features had an 89.6% sensitivity and 86.2% specificity for a diagnosis of FNH, and thus the surgical pathologist can at least favor FNH in this setting. Overall diagnostic accuracy is improved with the use of GS IHC on needle biopsy specimens with recognition of map-like GS staining having a 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 stains on needle biopsy specimens in a minority of cases. Up to 15% of FNHs failed to demonstrate map-like GS staining, most often due to 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.

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;23:1441-54.

2. Wanless IR, Mawdsley C, Adams R. On the pathogenesis of focal nodular hyperplasia of the liver. Hepatology 1985;5:1194-200.

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. Modern Pathol 1989;2:456-62.

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;236:166-77.

6. Demarco MP, Shen P, Bradley RF, Levine EA. Intraperitoneal hemorrhage in a patient with hepatic focal nodular hyperplasia. Am Surg 2006;72:555-9.

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;36:1210-6.

9. Fabre A, Audet P, Vilgrain V, et al. Histologic scoring of liver biopsy in focal nodular hyperplasia with atypical presentation. Hepatology 2002;35:414-20.

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. American Journal of Surgical Pathology 2017;41:277-81.

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;146:254-61.

12. Gill RM, Belt P, Wilson L, Bass NM, Ferrell LD. Centrizonal arteries and microvessels in nonalcoholic steatohepatitis. Am J Surg Pathol 2011;35:1400-4.

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;29:459-65.

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;36:1691-9.

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-

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;38:245-9.

18. Bedossa P, Dargere D, Paradis V. Sampling variability of liver fibrosis in chronic hepatitis C. Hepatology 2003;38:1449-57.

Figure Legends

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.

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

Figure 3: Focal nodular hyperplasia case with prominent steatosis and high interobserver variation in the interpretation of glutamine synthetase stain. 3 reviewers called the GS map-like and 3 reviewers interpreted the GS stain as patchy. By H&E alone, 22 of 25 reviewers called definite or probable FNH. In round 2 with H&E and GS stain, all reviewers diagnosed the case as probable or definite FNH.

Figure 4: Focal nodular hyperplasia case showing a "pseudo map-like" glutamine synthetase staining patter on needle biopsy. This pattern was characterized by less intense GS staining, more focal or narrower anastomosing areas, and expression limited to peripheral parts of the tumor nodules.

Supplemental Figure 1: Histologic features of FNH: abnormal vessels (A), ductular reaction (B), nodularity (C), and central scar (D).

Tables

Table 1: Reviewer diagnoses for 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%

Table 2: Sensitivity and specificity by histologic feature and combination of histologic features (using round 1 data, Round 2 for GS).

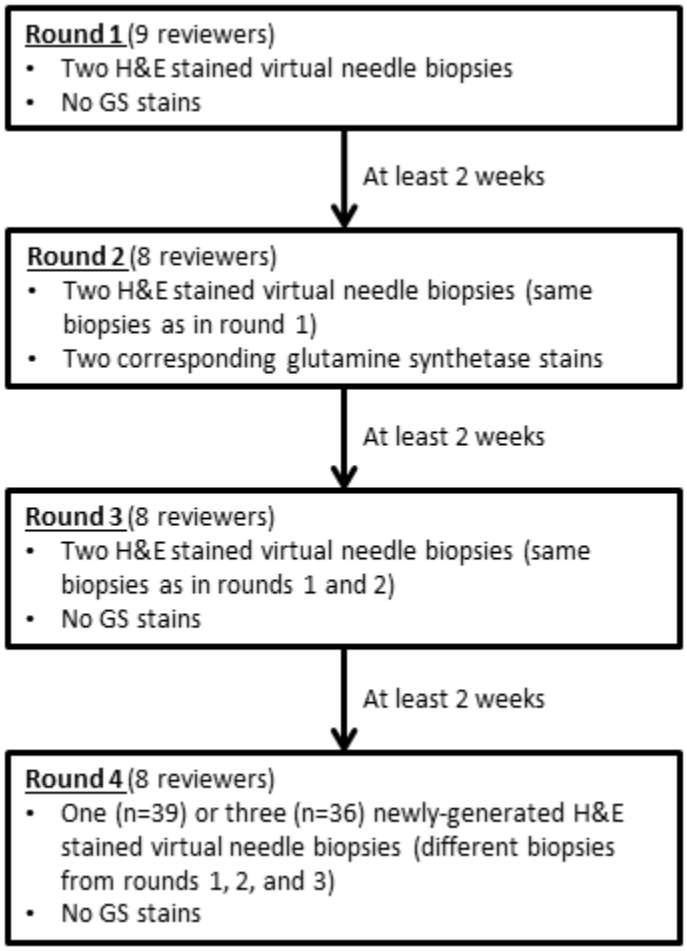
Feature(s)	**Sensitivity (%)**	**Specificity (%)**
Central scar	54.1	90.6
Ductular reaction	88.1	88.9
Nodularity	82.2	84.2
Abnormal vessels	66.2	76.8
All 4 histologic features	38.2	99.8
≥3 histologic features	64.1	96.5
≥2 histologic features	89.6	86.2
≥1 histologic feature	98.9	58.0
Map-like GS staining pattern	85.6	99.3

Abbreviations: GS- glutamine synthetase

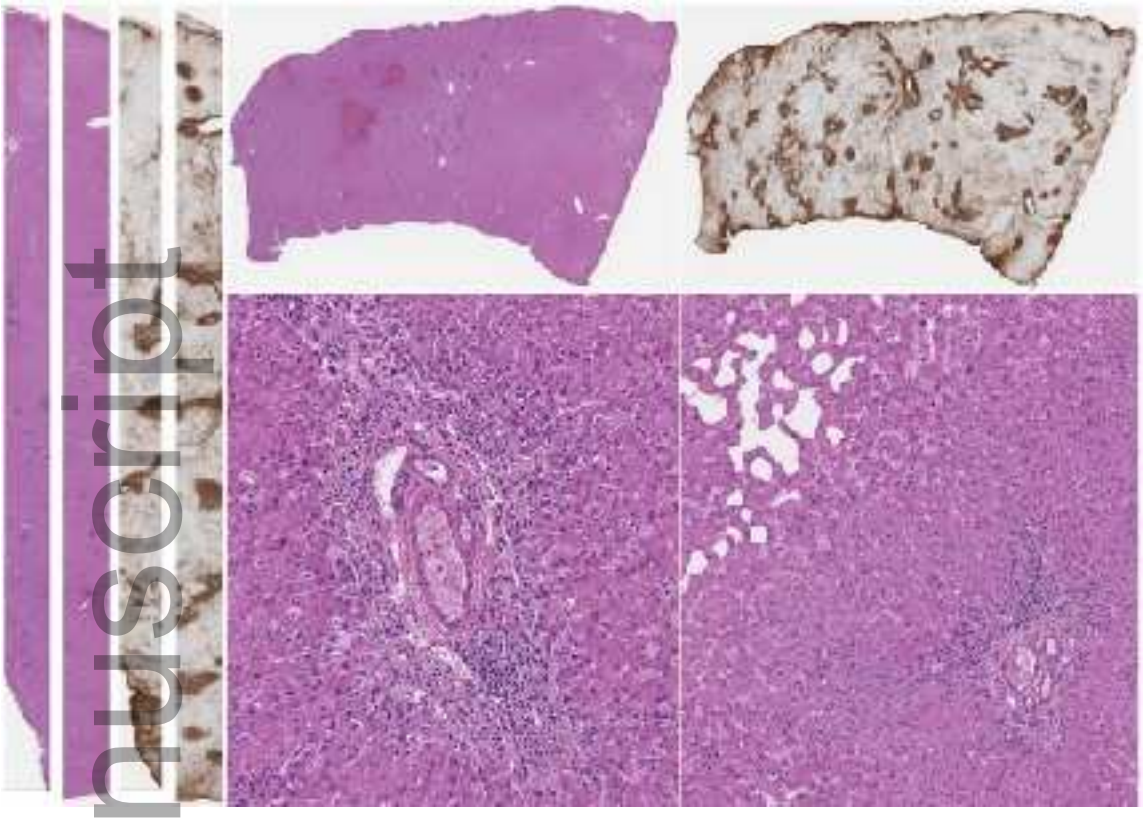
Table 3: Diagnostic accuracy by round.

	Round 1 H&E 2 cores	**Round 2 H&E, GS 2 cores**	**Round 3 H&E 2 cores**	**Round 4 H&E All cases**	**Round 4 H&E 1 core**	**Round 4 H&E 3 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%

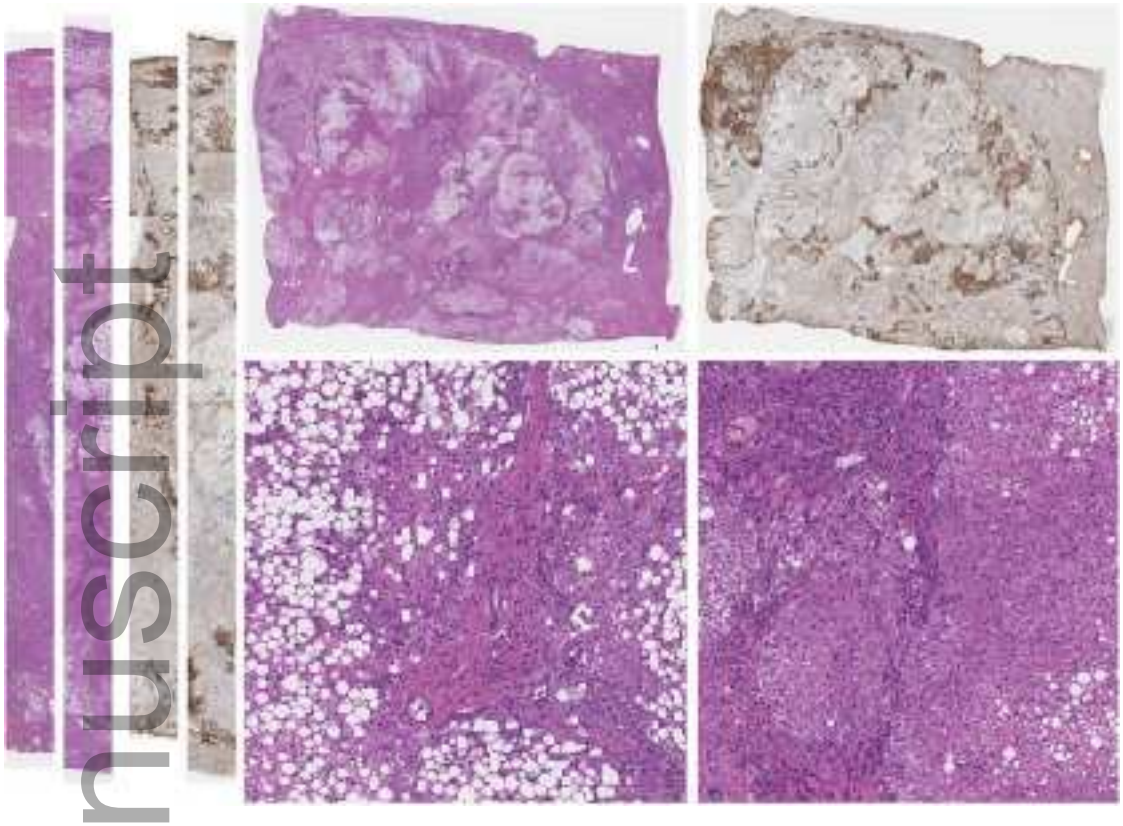
Abbreviations: FNH- focal nodular hyperplasia; GS- glutamine synthetase; H&E- hematoxylin and eosin



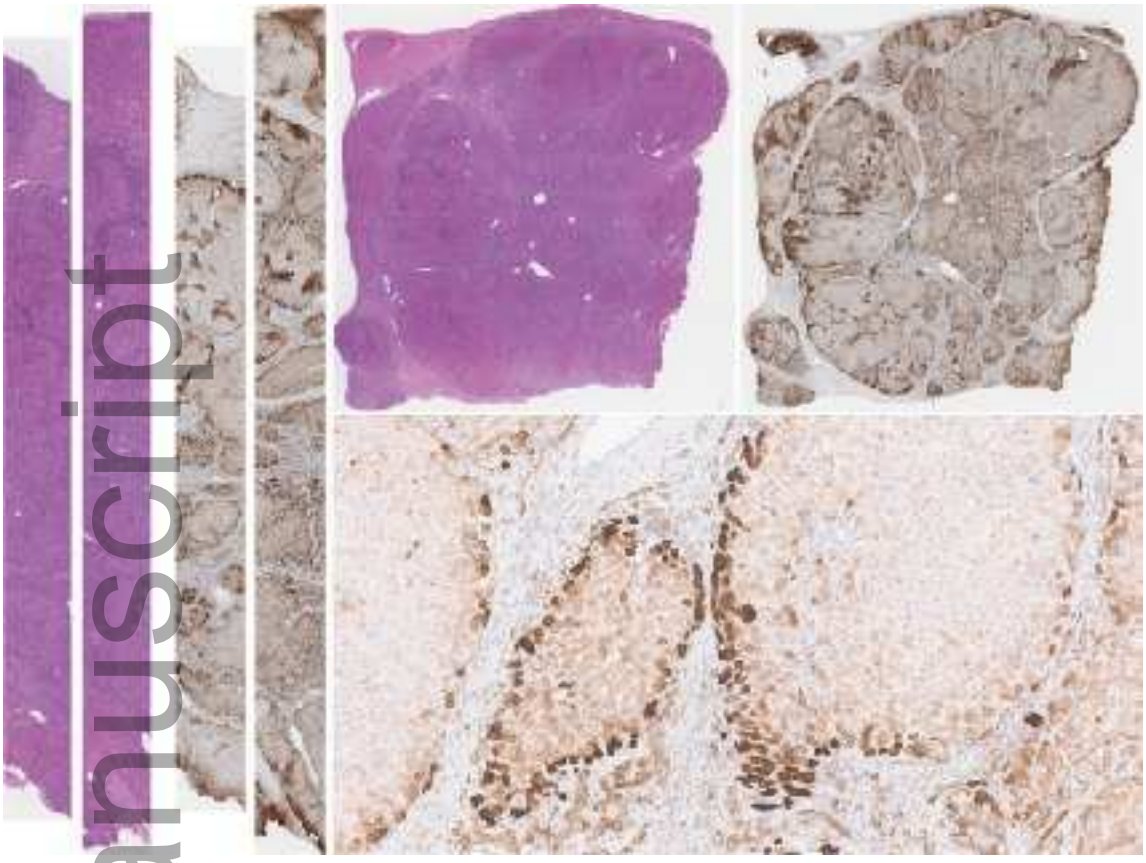
his_14424_f1.tif



his_14424_f2.tif



his_14424_f3.tif



his_14424_f4.tif