Innovative Tools for Quantitative Analysis of NAFLD Histology

Samer Gawrieh, MD
Associate Professor of Clinical Medicine
Indiana University School of Medicine
Genetic Modifiers

NAFLD Histological Phenotypes

NAFL

Borderline

NASH

Steatosis

Lobular Inflammation

Ballooning

Portal Inflammation

Fibrosis

Pathogenesis-
Animal Studies

Therapeutic Trials

Why Innovate?
Liver Biopsy-Based NAFLD Phenotyping:
Factors Affecting Biopsy Sample Quality and Diagnostic Yield

Ratziu V et al. Gastroenterology 2005
Merriman RB et al. Hepatology 2006

Biopsy sample analysis
Current State:
Manual and Semi-quantitative Grading of NAFLD Lesions

- **NAFLD Activity Score (NAS)** = MS + LI + HB
  - 0-3 for MS and LI
  - 0-2 for HB
  - 0-8 for NAS
  - 0-4 for FS

Kleiner DE et al. Hepatology 2005
Observer Agreement (k) on Reading NAFLD Histological Features
(65 biopsies, 2 pathologists, 260 readings)

Reported Inter-observer Agreement (κ) on Interpretation of NAFLD Histology

<table>
<thead>
<tr>
<th>Feature</th>
<th>Inter-Observer Agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cases (n)</td>
<td>53</td>
</tr>
<tr>
<td>Observers (n)</td>
<td>4</td>
</tr>
<tr>
<td>SG</td>
<td>0.64</td>
</tr>
<tr>
<td>FS</td>
<td>0.60</td>
</tr>
<tr>
<td>LI</td>
<td>0.21</td>
</tr>
<tr>
<td>PI</td>
<td>0.18</td>
</tr>
<tr>
<td>HB</td>
<td>0.50</td>
</tr>
<tr>
<td>Diagnosis</td>
<td>0.50</td>
</tr>
</tbody>
</table>
### Interventions

**Education / training**
- Slide review
- Discussion of diagnostic criteria

### Scoring sheet

#### Fatty liver disease scoring sheet

<table>
<thead>
<tr>
<th>Patient</th>
<th>S#</th>
<th>MRN</th>
</tr>
</thead>
</table>

#### Definitions

**FATTY LIVER** diagnosed when **ONLY STEATOSIS (> 5%)** is present

**Steatohepatitis (SH)** is **steatosis + 2 of the following zone 3, centric features**
- Hepatocellular ballooning
- Inflammatory infiltrate
- Pericellular/Perisinusoidal fibrosis

**Possible/borderline SH** is **steatosis + 1 of the above zone 3, centric features**

---

### Simplified written criteria for diagnosis

1. **Fatty liver**
2. Possible/borderline **SH**
3. **Steatohepatitis (SH)**
4. **Normal**

Additional or other diagnosis

---

### Photo case

---
### Pre- and Post-Intervention Observer Agreement (k) on NAFLD Histological Features and Phenotype

(65 biopsies, 2 pathologists, 520 readings)

<table>
<thead>
<tr>
<th>Feature</th>
<th>Intra-observer</th>
<th>Inter-observer</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pre</td>
<td>Post</td>
</tr>
<tr>
<td>Steatosis grade</td>
<td>0.72</td>
<td>0.75</td>
</tr>
<tr>
<td>Fibrosis stage</td>
<td>0.64</td>
<td>0.75</td>
</tr>
<tr>
<td>Lobular inflammation</td>
<td>0.37</td>
<td>0.48</td>
</tr>
<tr>
<td>Portal inflammation</td>
<td>0.46</td>
<td>0.58</td>
</tr>
<tr>
<td>Hepatocellular ballooning</td>
<td>0.32</td>
<td>0.56*</td>
</tr>
<tr>
<td>Diagnostic classification</td>
<td>0.51</td>
<td>0.54</td>
</tr>
</tbody>
</table>

* $p$ for pre- versus post comparisons was significant only for intra-observer k on HB (0.009)

NAFLD epidemic is on the rise, new approaches to decrease interobserver variability in interpretation of NAFLD histology are urgently needed. Refined histopathologic criteria and training on assessment of HB and LI would likely have the largest impact on reproducibility of NAFLD phenotyping and staging. Automation of assessment of NAFLD histologic features to improve accuracy and reproducibility of the interpretation is another consideration. This can be envi-
What to Detect and Quantify: Cardinal NAFLD Lesions

- Macrosteatosis
- Lobular Inflammation
- Hepatocyte Ballooning
- Fibrosis
- Portal Inflammation
What to Detect and Quantify: Cardinal NAFLD Lesions

- Macrosteatosis
- Lobular Inflammation
- Hepatocyte Ballooning
- Fibrosis Architectural Pattern
- Collagen proportionate area (CPA)
- Portal Inflammation
## How to Quantify: Automation Approaches and Requirements

<table>
<thead>
<tr>
<th></th>
<th>Machine learning</th>
<th>Algorithm-based</th>
<th>Other (Adobe- or color-based tools)</th>
</tr>
</thead>
<tbody>
<tr>
<td>What’s quantified</td>
<td>NAFLD lesions</td>
<td>Correlates with NAFLD lesions</td>
<td>Usually fibrosis and/or steatosis assessment</td>
</tr>
<tr>
<td>Stains</td>
<td>H&amp;E, Mason trichrome, Sirius red</td>
<td>Unstained/stained slides</td>
<td>H&amp;E, Mason trichrome, Sirius red</td>
</tr>
<tr>
<td>Digital Slide Scanner</td>
<td>√</td>
<td>√</td>
<td>+/-</td>
</tr>
<tr>
<td>Equipment additional to software</td>
<td>None</td>
<td>SHG/TPEF Microscope*</td>
<td>None</td>
</tr>
</tbody>
</table>

* SHG: Second harmonic generation microscopy, PTEF: Two-photon excitation fluorescence microscopy
Machine Learning (ML)/ Artificial Intelligence (AI)

• ML: algorithms and statistical models that learn from labelled training data, from which they are able to recognize and infer patterns

• General AI: ability of a machine to communicate, reason and operate independently in both familiar and novel scenarios in a similar manner to a human

• Commonly, ML interchangeable with AI

General Approach to Developing ML Models for NAFLD Histology Analysis

Scientists Team
(Computer Scientists, Pathologists, Hepatologist)

Digital images of NAFLD liver biopsies

Pathologist’s annotations

- Model development and internal validation (Labeled data)
- Correlation with pathologist scores

External Validation (unlabeled data)
Supervised Machine Learning

Candidate Region

Feature Vector:
- Size
- Shape
- Color
- Texture
- Etc.

Learning Data

MACHINE LEARNING CLASSIFIER

Model Prediction
- Steatosis
- Central Vein
- Bile Duct.
- etc.
Deep Machine Learning and Neural Networks

Pathologist Annotations Software
Annotation Attributes (Feature Vectors) with Supervised ML

2-Jet

Scaled Representation

(a) sigma=0  (b) sigma=1  (c) sigma=2  (d) sigma=4  (e) sigma=8

2-Jet features

(a) Original Image  (b) Gradient Magnitude  (c) Laplacian  (d) Determinant of Hessian  (e) Level Curvature

Vanderbeck S et al. Hum Pathol 2014
Internal Testing/Validation

• 10 fold cross validation

Experiment 1

Experiment 2

Experiment 10

• Leave one out approach
Hepatocyte Ballooning Example

Original Tile
Probability
Threshold
Result
Classification Approach

• Tile image into equal size pieces
• Classify each tile as either containing or NOT hepatocyte ballooning
• Calculate the total percent of tissue with hepatocyte ballooning

\[
% \text{ Ballooning} = \frac{\text{total area ballooning tiles}}{\text{total tissue area}}
\]

• Similar procedure for other lesions

Vanderbeck S et al. Hum Pathol 2015
White Regions on H&E Stained Liver Biopsy Images
Automated Detection of Liver Microscopic Anatomic Landmarks

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Precision</td>
<td>Recall</td>
<td>ROC area</td>
<td>Precision</td>
<td>Recall</td>
<td>ROC area</td>
</tr>
<tr>
<td>Bile duct</td>
<td>0.92</td>
<td>0.87</td>
<td>0.98</td>
<td>1.00</td>
<td>0.56</td>
<td>0.93</td>
</tr>
<tr>
<td>Central vein</td>
<td>0.64</td>
<td>0.79</td>
<td>0.92</td>
<td>0.67</td>
<td>0.64</td>
<td>0.83</td>
</tr>
<tr>
<td>Macrosteatosis</td>
<td>0.98</td>
<td>0.99</td>
<td>0.98</td>
<td>0.92</td>
<td>0.94</td>
<td>0.96</td>
</tr>
<tr>
<td>Other</td>
<td>1.00</td>
<td>0.86</td>
<td>0.99</td>
<td>0.76</td>
<td>0.69</td>
<td>0.90</td>
</tr>
<tr>
<td>Portal artery</td>
<td>0.85</td>
<td>0.77</td>
<td>0.97</td>
<td>0.67</td>
<td>0.18</td>
<td>0.94</td>
</tr>
<tr>
<td>Portal vein</td>
<td>0.91</td>
<td>0.77</td>
<td>0.97</td>
<td>0.81</td>
<td>0.88</td>
<td>0.97</td>
</tr>
<tr>
<td>Sinusoid</td>
<td>0.90</td>
<td>0.89</td>
<td>0.96</td>
<td>0.80</td>
<td>0.79</td>
<td>0.92</td>
</tr>
</tbody>
</table>

Vanderbeck S et al. Hum Pathol 2014
Correlation and Relationship of Model Calculated Percent Steatosis with the Average of Pathologists Grade

Vanderbeck S et al. Hum Pathol 2014
Automated Continuous Quantification of Macrosteatosis

<table>
<thead>
<tr>
<th>Level</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>&lt;5%</td>
</tr>
<tr>
<td>1</td>
<td>6-33%</td>
</tr>
<tr>
<td>2</td>
<td>33-66%</td>
</tr>
<tr>
<td>3</td>
<td>&gt;66%</td>
</tr>
</tbody>
</table>
Correlation of Model Calculated Percent *Hepatocyte Ballooning* with the Average of Pathologists Grade

R² = 0.4905

<table>
<thead>
<tr>
<th>Feature</th>
<th>Precision (PPV)</th>
<th>Recall (Sensitivity)</th>
<th>ROC Area</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hepatocyte Ballooning</td>
<td>0.912</td>
<td>0.542</td>
<td>0.983</td>
</tr>
<tr>
<td>Not-Hepatocyte Ballooning</td>
<td>0.990</td>
<td>0.999</td>
<td>0.983</td>
</tr>
<tr>
<td>OVERALL</td>
<td>0.989</td>
<td>0.989</td>
<td>0.983</td>
</tr>
</tbody>
</table>

Vanderbeck S et al. Hum Pathol 2015
Correlation of Model Calculated Percent *Lobular Inflammation* with the Average of Pathologists Grade

\[ R^2 = 0.1724 \]

<table>
<thead>
<tr>
<th>Feature</th>
<th>Precision (PPV)</th>
<th>Recall (Sensitivity)</th>
<th>ROC Area</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lobular Inflammation</td>
<td>0.696</td>
<td>0.489</td>
<td>0.946</td>
</tr>
<tr>
<td>Not-Lobular Inflammation</td>
<td>0.968</td>
<td>0.986</td>
<td>0.946</td>
</tr>
<tr>
<td><strong>OVERALL</strong></td>
<td><strong>0.952</strong></td>
<td><strong>0.956</strong></td>
<td><strong>0.946</strong></td>
</tr>
</tbody>
</table>

Vanderbeck S et al. Hum Pathol 2015
Refined Model for **Lobular** Inflammation and New Model for **Portal** Inflammation in Human NAFLD

AUROC – 97.4%, Precision – 79.3%, Sensitivity – 81.3%,

Unpublished data

AUROC – 97.9%, Precision – 82.1%, Sensitivity – 88.3%
Automated Fibrosis Assessment in Human NAFLD

CPA Correlation with Pathologist DEK

\[ y = 73.031x^2 + 10.28x - 0.7224 \]

\[ R^2 = 0.6071 \]

CPA Correlation with Pathologist OWC

\[ y = 310.57x^2 - 44.386x + 1.4567 \]

\[ R^2 = 0.867 \]

Automated Identification of Architectural Type of Liver Fibrosis in Human NAFLD Liver Biopsies

<table>
<thead>
<tr>
<th>Fibrosis Type</th>
<th>Precision (%)</th>
<th>Recall (%)</th>
<th>AUROC (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>85.6</td>
<td>83.3</td>
<td>91.9</td>
</tr>
<tr>
<td>Pericellular</td>
<td>76.6</td>
<td>82.7</td>
<td>83.3</td>
</tr>
<tr>
<td>Periportal</td>
<td>72.1</td>
<td>76.9</td>
<td>78.6</td>
</tr>
<tr>
<td>Portal</td>
<td>77</td>
<td>84.4</td>
<td>86.4</td>
</tr>
<tr>
<td>Bridging</td>
<td>84.9</td>
<td>91.7</td>
<td>93</td>
</tr>
<tr>
<td>Nodule</td>
<td>89.8</td>
<td>91.6</td>
<td>95.4</td>
</tr>
</tbody>
</table>

Machine Learning for Automated NAFLD Histology Assessment

246 patients with NAFLD (190 with NASH, 56 with simple steatosis)

Correlation (rho) with Pathologist’s scores
- Steatosis: 0.66
- Inflammation: 0.36
- Ballooning: 0.52
- Fibrosis: 0.57

AUROC for CPA for F≥3: 0.82

Forlano R et al. Clin Gastroenterol Hepatol. 2020
Deep Machine Learning for Automated NASH Histology Assessment

- 834 liver biopsies from subjects screened for a phase 3 trial of selonsertib (STELLAR-4) (advanced cirrhotic NASH)
- CNN with over 20 layers and 8 million parameters using over 68,000 annotations collected from 75 board-certified pathologists

Correlations (rho) with average of two pathologists scores
- Steatosis: 0.86
- Inflammation: 0.56
- Ballooning 0.68
- Fibrosis 0.83 (CRN), 086 (Ishak)
Phenotypic Fibrosis Composite Score by FibroNest Image Analysis Platform

77 NASH Biopsies; unstained-SHG/TPEF or stained by Sirius red or trichrome-digitally scanned

Chen L et al. EASL abstract 2020
External Validation of Steatosis Classifiers in Murine NAFLD

Macrosteatosis

Microsteatosis

Accuracy of labels: 100%

Accuracy of labels: 63%

Sethunath D et al. Plos1 2018
Automated algorithm- and SHG microscopy-based assessment of NAFLD histology (qFIBS)

Correlation (r) with Pathologist’s scores
- qSteatosis: 0.802
- qInflammation: 0.557
- qBallooning: 0.533
- qFibrosis: 0.776

<table>
<thead>
<tr>
<th></th>
<th>AUROC</th>
<th>95% CI</th>
<th>P Value</th>
<th>Cutoff</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>PPV</th>
<th>NPV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>(Youden’s</td>
<td></td>
<td>Index)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>qFibrosis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F0 vs. F21</td>
<td>0.870</td>
<td>0.787-0.953</td>
<td>&lt;0.001</td>
<td>0.761</td>
<td>94%</td>
<td>63%</td>
<td>84%</td>
<td>83%</td>
</tr>
<tr>
<td>F51 vs. F22</td>
<td>0.801</td>
<td>0.804-0.959</td>
<td>&lt;0.001</td>
<td>0.882</td>
<td>97%</td>
<td>58%</td>
<td>65%</td>
<td>96%</td>
</tr>
<tr>
<td>F52 vs. F33</td>
<td>0.945</td>
<td>0.891-0.999</td>
<td>&lt;0.001</td>
<td>1.491</td>
<td>96%</td>
<td>76%</td>
<td>66%</td>
<td>97%</td>
</tr>
<tr>
<td>F53 vs. F4</td>
<td>0.961</td>
<td>0.906-0.996</td>
<td>&lt;0.001</td>
<td>2.395</td>
<td>87%</td>
<td>91%</td>
<td>72%</td>
<td>96%</td>
</tr>
<tr>
<td>qInflammation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 vs. ≥1</td>
<td>0.838</td>
<td>0.752-0.924</td>
<td>0.105</td>
<td>1.261</td>
<td>83%</td>
<td>100%</td>
<td>100%</td>
<td>14%</td>
</tr>
<tr>
<td>≤1 vs. ≥2</td>
<td>0.820</td>
<td>0.726-0.913</td>
<td>&lt;0.001</td>
<td>1.357</td>
<td>93%</td>
<td>58%</td>
<td>56%</td>
<td>93%</td>
</tr>
<tr>
<td>≤2 vs. 3</td>
<td>0.831</td>
<td>0.729-0.933</td>
<td>0.112</td>
<td>1.503</td>
<td>100%</td>
<td>79%</td>
<td>12%</td>
<td>100%</td>
</tr>
<tr>
<td>qBallooning</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 vs. ≥1</td>
<td>0.844</td>
<td>0.731-0.957</td>
<td>0.011</td>
<td>1.086</td>
<td>71%</td>
<td>100%</td>
<td>100%</td>
<td>20%</td>
</tr>
<tr>
<td>≤1 vs. 2</td>
<td>0.813</td>
<td>0.708-0.918</td>
<td>&lt;0.001</td>
<td>1.266</td>
<td>60%</td>
<td>89%</td>
<td>67%</td>
<td>85%</td>
</tr>
<tr>
<td>qSteatosis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 vs. ≥1</td>
<td>0.986</td>
<td>0.959-1.000</td>
<td>&lt;0.001</td>
<td>0.796</td>
<td>99%</td>
<td>100%</td>
<td>100%</td>
<td>50%</td>
</tr>
<tr>
<td>≤1 vs. 2</td>
<td>0.948</td>
<td>0.903-0.993</td>
<td>&lt;0.001</td>
<td>1.572</td>
<td>91%</td>
<td>86%</td>
<td>83%</td>
<td>92%</td>
</tr>
<tr>
<td>≤2 vs. 3</td>
<td>0.939</td>
<td>0.867-1.000</td>
<td>&lt;0.001</td>
<td>2.210</td>
<td>67%</td>
<td>98%</td>
<td>86%</td>
<td>95%</td>
</tr>
</tbody>
</table>

Liu F et al. Hepatology 2020
Considerations for Development of Automated Methods for NAFLD Histology Analysis

• Minimum acceptable standards for liver biopsy sample
• Derivation biopsy cohort
  – Minimizing bias in biopsy selection: Representation of the entire histological spectrum of NAFLD
  – Large number of annotations by expert NAFLD pathologists to train the model (Different severity/typical/atypical variety of each lesion)
  – Number of expert pathologists involved? 1 vs more, observer agreement, other factors...
• Trade offs of how thresholds are set {High sensitivity/High specificity/Optimal(Youden)}
Considerations for Development of Automated Methods for NAFLD Histology Analysis

• External validation
  – Verification by experts pathologists of accuracy of lesion identification on unseen biopsy images
  – Validation of performance in different cohorts

• ? Weight of strength of correlation with semi-quantitative assessments/scores

• Lack of explanability factor for deep learning networks and regulatory approvals:
  • Unknown what attributes of a lesion are used/contribute to decision making process in nodes/networks (Black box factor)

• These tools may be viewed as complimentary decision aids/guides, not replacements, to pathologists
Current and Future States of NAFLD Histology Analysis

Current

• Manual
• Semi-quantitative data
• Limited scale
• Intra- & inter-observer variability
• Limited pool of experienced NAFLD pathologists
• Limited access to experienced NAFLD pathologists

Future: Optimize/Maximize Extracted Data

• Automated
• Continuous data
• Large scale
• Precise
• Reproducible
• Available
• Accessible

Pathologists are key partners in leading us through this transformation of the field
Acknowledgements

• Pathologists
  – David Kleiner, MD, PhD (National Cancer Institute)
  – Richard Komorowski, MD (Medical College of Wisconsin)
  – Oscar W. Cummings, MD (Indiana University School of Medicine)

• Computer Engineers
  – Scott Vanderbeck, M.S. and Joe Bockhurst, PhD (University of Wisconsin – Milwaukee)
  – Mihran Tuceryan, PhD (Indiana University-Purdue University-Indianapolis)