

Digital Pathology AI methods based on supervised ML and annotations provided by pathologists yield to moderate performance results

Made with FibroNest

Evaluation of the performance of a novel Digital Pathology method for the continuous quantification of Steatosis, Ballooning and Inflammation in liver biopsies and its correlation with NASH-CRN scores in patients with NASH

Louis PETITJEAN¹, Li CHEN¹, Arras N. MATTIS², Mojgan HOSSEINI³, Michael GOEDKEN⁴, Cynthia BEHLING⁵, Arun J. SANYAL⁶, Mathieu PETITJEAN¹

1 Pharnanest, Princeton, NJ, USA 2 Department of Pathology, University of California San Francisco, CA, USA 3 Department of Pathology, University of California San Diego, CA, USA 4 Rutgers Research Pathology Services, New Brunswick, NJ USA, 5 Department of Pediatrics University of California, San Diego, CA 4 Virginia Commonwealth University, Richmond, VA, USA

PharmaNest



1 Introduction

Manual histological evaluation of liver biopsy is the gold standard for fibrosis and steatosis staging in Non-Alcoholic Steatohepatitis (NASH), but it is limited by its inter and intra-reader variability. Quantitative Digital Pathology image analysis and AI (FibroNest™) methods have the potential to overcome the limitation of these standards

2 Aim

We have previously shown that the Phenotypic Fibrosis Composite Score calculated by the FibroNest methods correlate with the NASH-CRN histological fibrosis stages and steatosis grades established from collagen-stained histology slides. Here, we report the performance of the FibroNest method to quantify **Ballooning, Inflammation** and **Steatosis** from digital images of H&E human liver biopsy sections.

3 Method

1. Image Dataset:

- Retrospective cohort of 85 patient with NASH diagnosed by histologic assessment of liver biopsy according to NASH-CRN.
- 20X digital images of H&E stained FFPE sections of liver biopsies

2. Machine Learning Training:

- Cohort of 21 selected images
- Annotations by four expert pathologists:
 - Positive features
 - False or "Look-Alike" features
 - Healthy Tissue landmarks
 - "one-click" annotation / point per feature

3. Predictive ML Model:

- Topographical probability maps for Lobular, Portal inflammation and hepatocellular ballooning
- Macro-vesicular steatosis feature confidence

4. Composite scores calculation:

The selection of a model probability cut off defines real-estate "objects" that can be quantified for count, density, morphology at the tissue level or for 200X equivalent FOV (total 64 components). Principal components are combined into composite scores.

5 Conclusions

As reported by other teams, liver biopsy Digital Pathology AI methods based on supervised ML and annotations provided by pathologists (H&E stains, Steatosis, Inflammation, Ballooning) result in continuous quantification methods of moderate performance.

6 Contact information

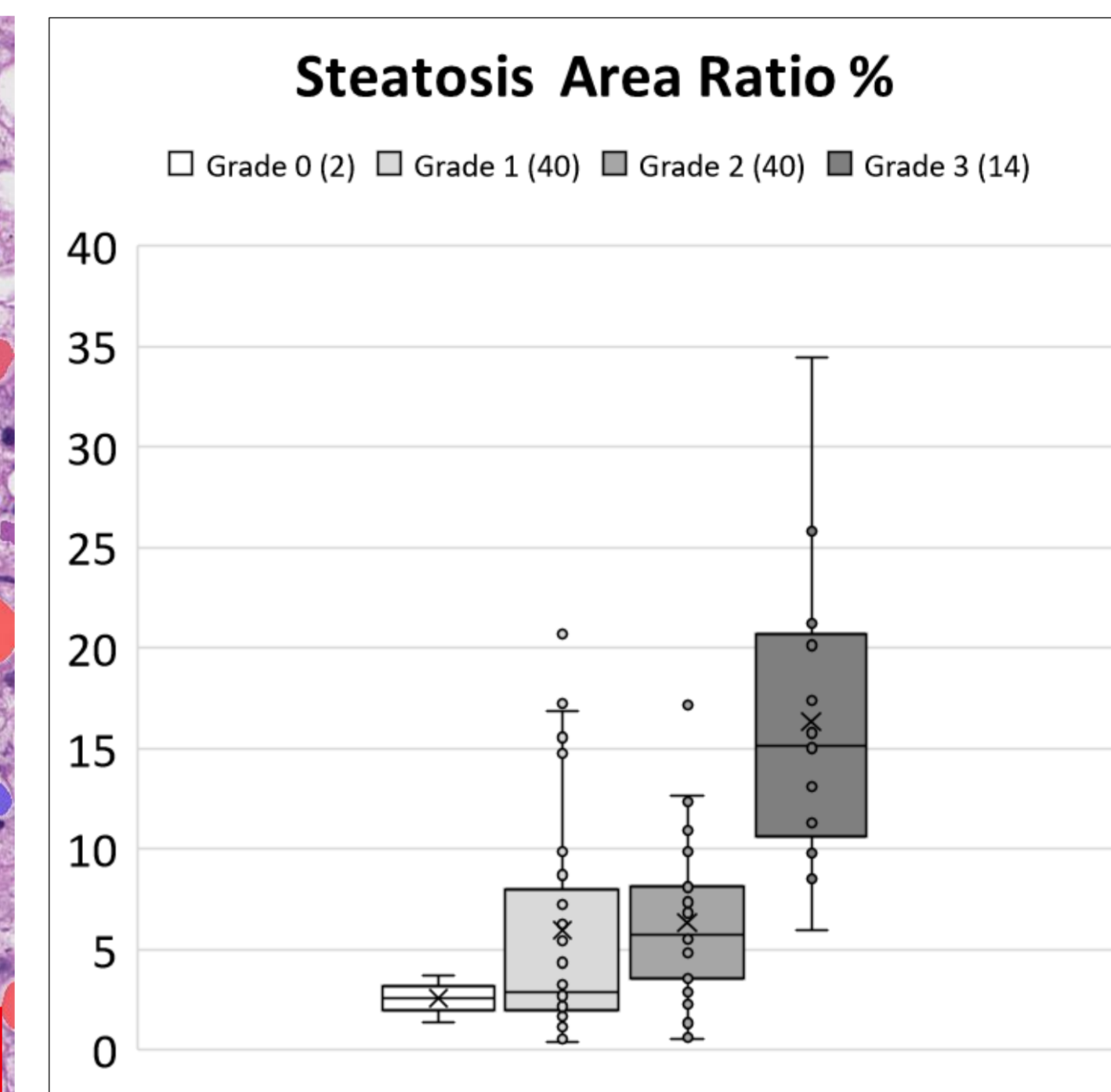
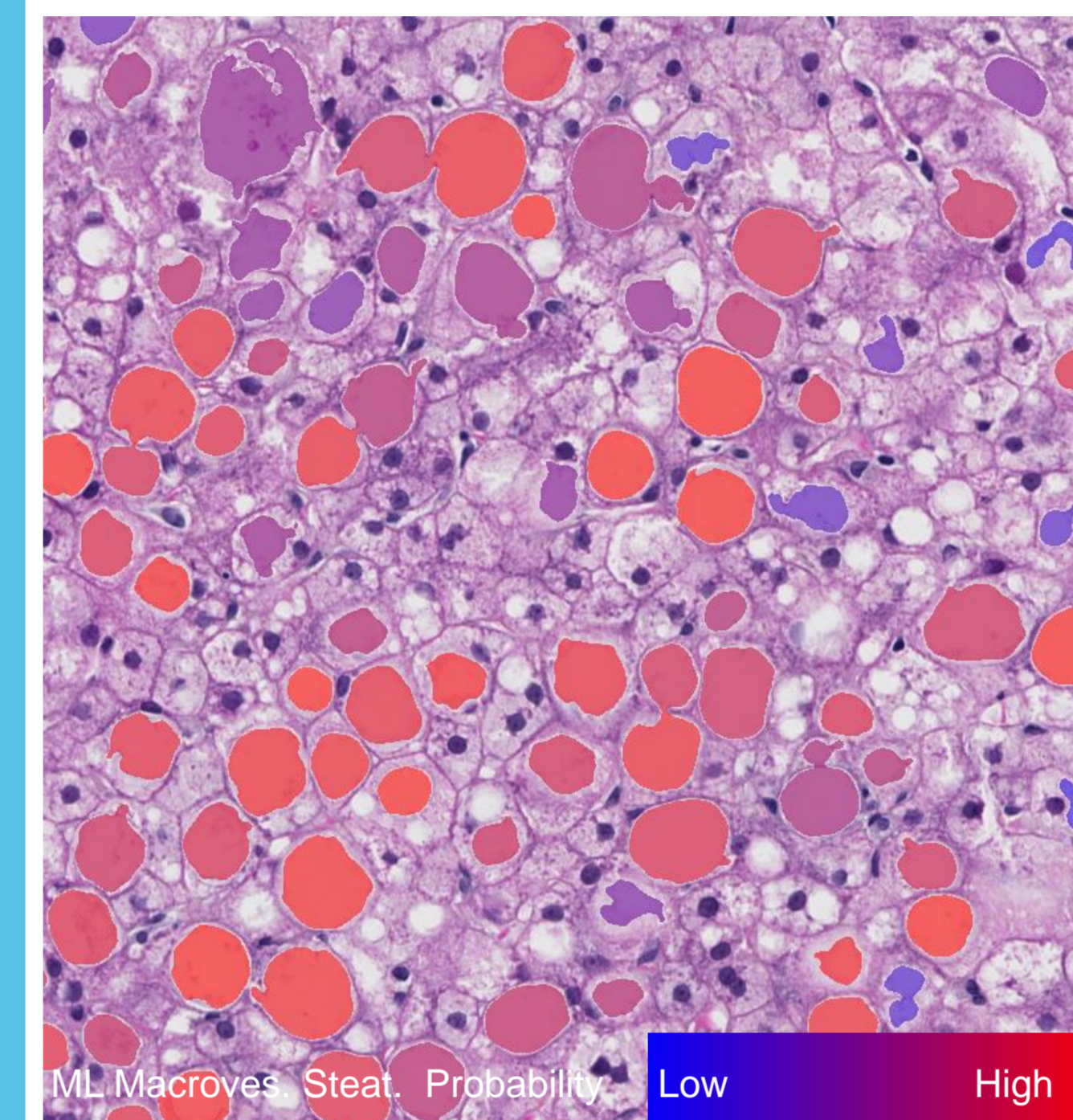
Louis.Petitjean@pharnanest.com

Histologic Assesment (NASH-CRN)	Steatosis		Lobular Inflammation		Hepatocyte ballooning	
	N	%	N	%	N	%
Training Cohort						
Grade 0	1	4.8%	4	19.0%	5	23.8%
Grade 1	6	28.6%	8	38.1%	8	38.1%
Grade 2	8	38.1%	9	42.9%	8	38.1%
Grade 3	6	28.6%	0	0.0%	0	0.0%
Total	21		21		21	
Training Annotations						
Lobular Inflammation Foci			2.30K	24.3%		
Portal and Peri Inflammation Foci			5.62K	59.5%		
Sinusoidal Nuclei Alignment			611	6.5%		
Ductal Foci			701	7.4%		
Other Nuclei Cluster			217	2.2%		
Hepatocellular Ballooning					1.08K	57.8%
"Look-alike" / False Hep. Ballooning					787	42.2%
Macrovesicular Steatosis	8.31K	90.4%				
Anatomical Features (several)	240	2.6%				
Other Non-Macrovesicular features	640	7.0%				
Total Project Annotations	9.19 K		9.45 K		1.87 K	
Total Annotations (all included)					24.79K	
Validation Cohort						
Grade 0	2	2%	7	8%	31	36%
Grade 1	39	46%	66	78%	41	48%
Grade 2	29	34%	12	14%	13	15%
Grade 3	15	18%	0	0%	0	0%
Total	85		85		85	

5. Validation Cohort

- Training cohort is re-incorporated in the Validation Cohort
- Validation cohort is process and the agreement of the composite score with NASH-CRN grades is evaluated.
- Iterations (steps 1 to 5) are performed adjust and enrich the predictive model and the performance of the scores (not reported here)

4 Steatosis Results

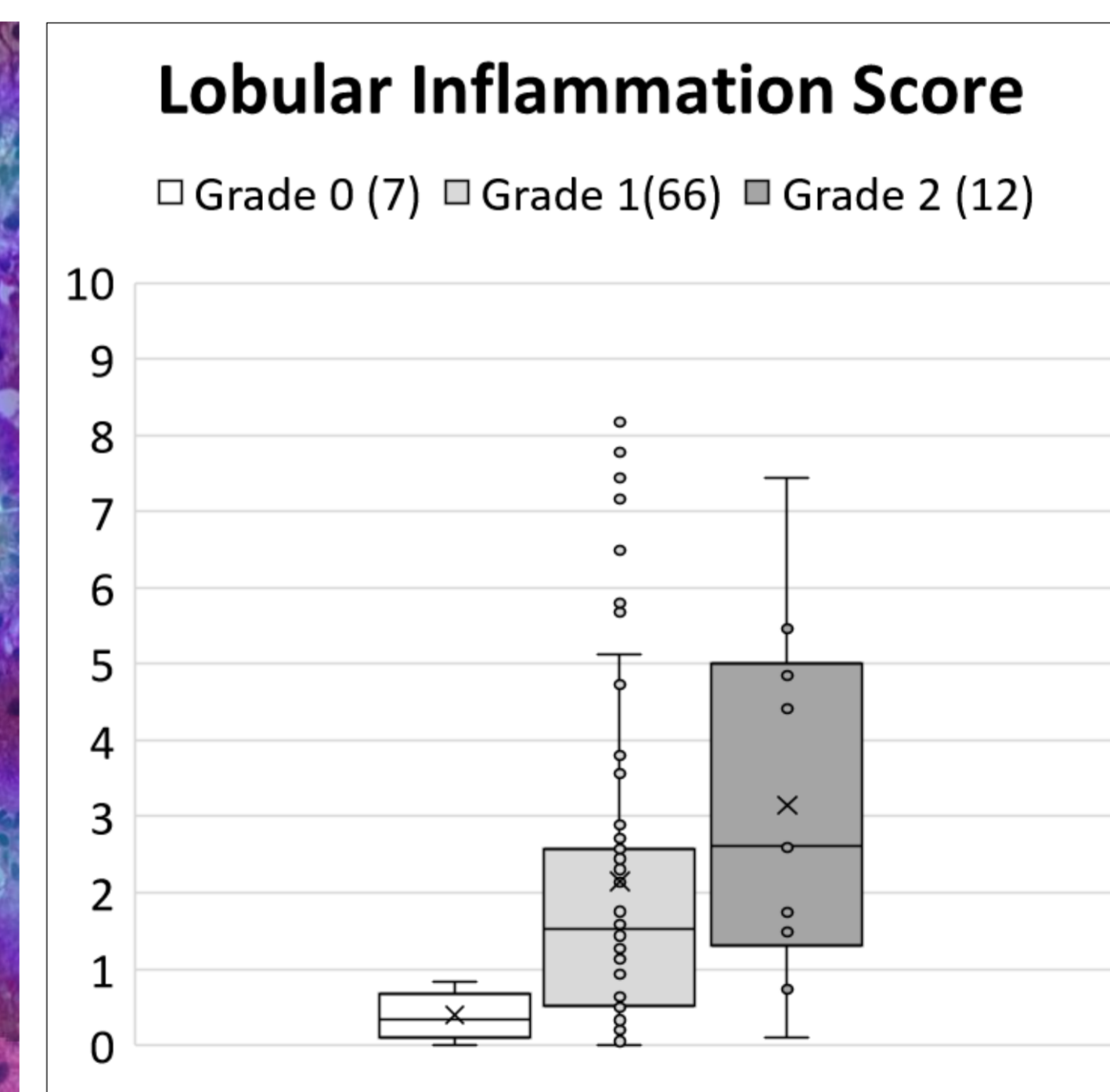
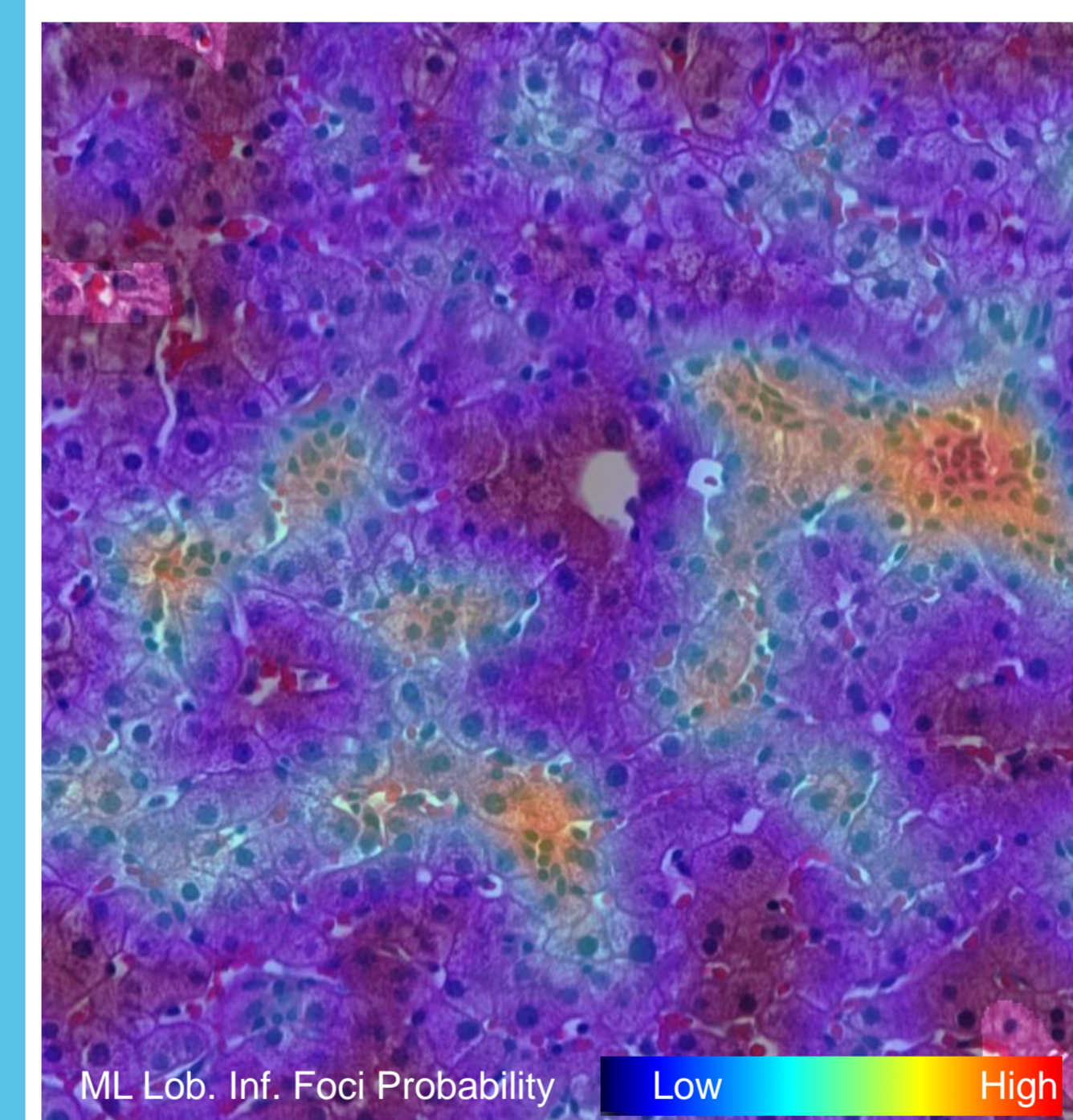


p-Values (Student t-Test)

Steatosis Area Ratio	Grade 0 (2)	Grade 1 (39)	Grade 2 (29)	Grade 3 (15)
Grade 0 (2)		0.1169	0.1190	0.0002
Grade 1 (39)			0.7747	0.0001
Grade 2 (29)				0.0001

- Effective exclusion of micro vesicular steatosis and glycogenotic hepatocytes.
- Model translates to Pre-clinical tissues
- Grade 1 vs 2 confusion is the combination of ML accuracy and Pathologists annotation.
- This approach does not meet FDA outcomes definitions ("either percent of steatosis hepatocytes or non-fibrotic tissue fat area ratio")
- New FibroNest release resolves these issues and results and results higher performance (see EASL2002 poster # FRI098)

Nuclear Inflammation Results

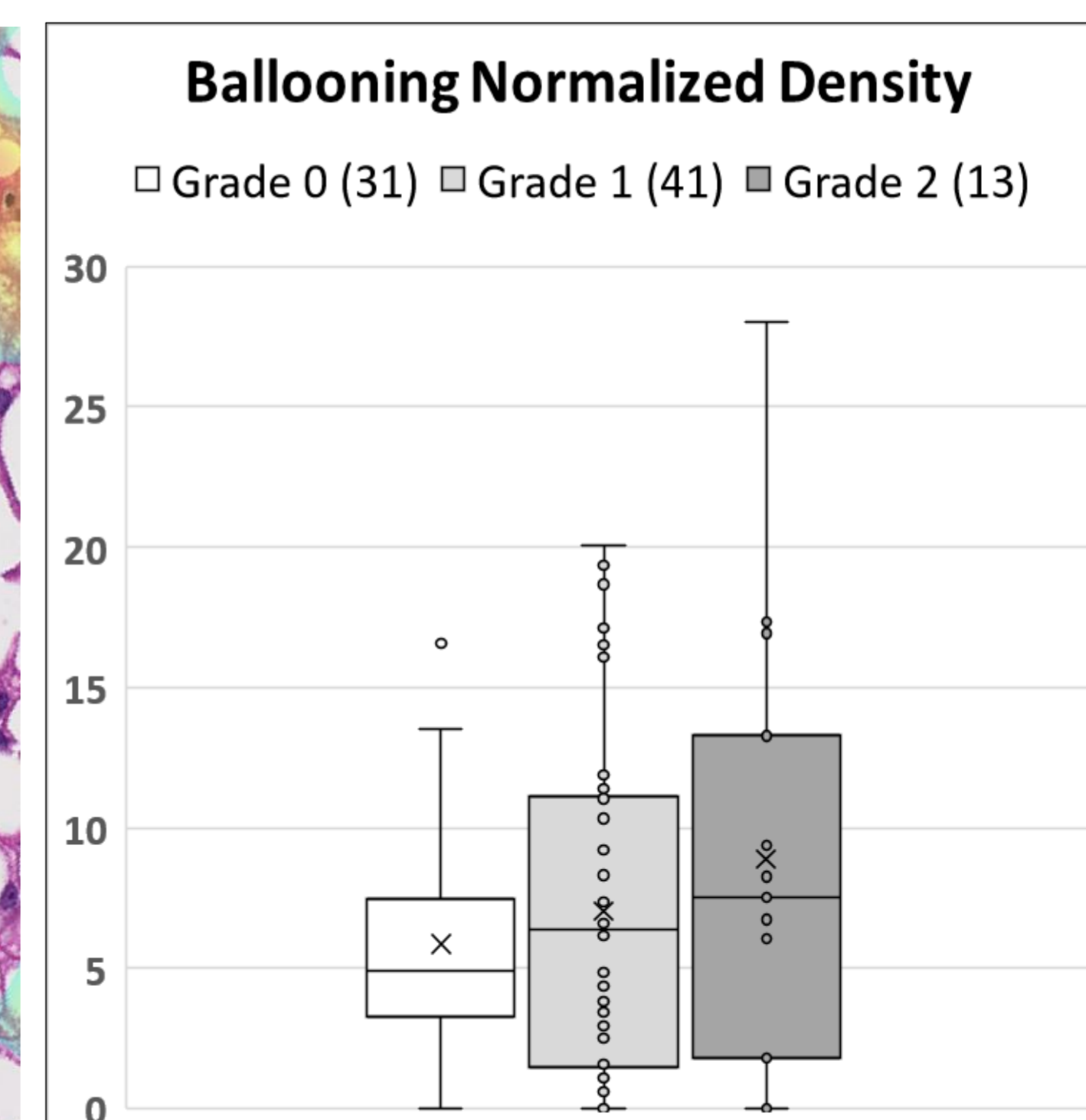
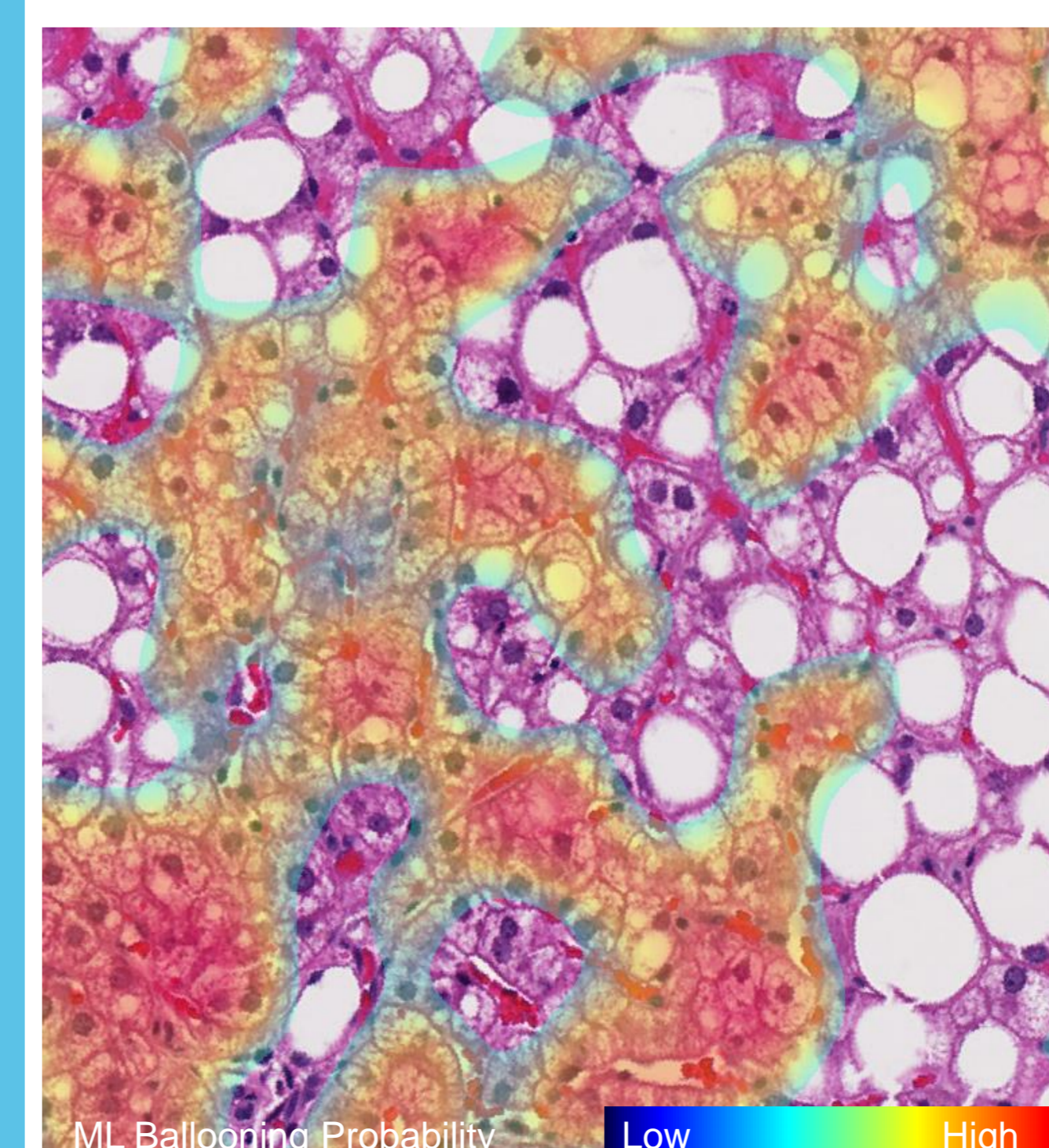


p-Values (Student t-Test)

Lobular Inflammation Score	Grade 0 (7)	Grade 1 (66)	Grade 2 (12)
Grade 0 (7)		0.0000	0.0018
Grade 1 (66)			0.1874

- ML Model accuracy (F1) for Lobular foci is 56%
- Further improvements are possible if the histological definition of a "lobular foci" is improved
- The intrinsic automated and systematic quantification method is attractive as an investigational endpoint in NASH studies
- The performance of the method is reduced for significantly hematoxylin overstained tissues
- Quantitative Image Analysis ("single Nuclei" analysis, FibroNest V3.1) resolves these issues.

Hepatocellular Ballooning Results



p-Values (Student t-Test)

Hepatocyte Ballooning Density	Grade 0 (7)	Grade 1 (66)	Grade 2 (12)
Grade 0 (7)		0.2266	0.8483
Grade 1 (66)			0.4676

- Significant disagreement between pathologists' annotations as reported elsewhere), due to a poor definition of (a) ballooned cells, "definite" vs "diagnostically borderline" and (b) accounting methods "none / few / many"
- The ML Model accuracy (F1) for Ballooned Hepatocytes is 37%
- Thresholding the "ballooned Hepatocyte" topographical probability maps at high probability (>65%) enables the detection of a Density scores that correlates poorly with Pathology grades.
- Further improvements are possible if the histological definition of "ballooning" is improved