Comprehensive characterization of granular fibrotic and cellular features in liver tissue enabled by deep learning models

Adam Stanford-Moore[1*#], Neel Patel[1*#], Ylaine Gerardin[1*], Yibo Zhang[1#], Jun Zhang[1#], Pratik Mistry[1#], Deeksha Kartik[1], Yi Liu[1#], Nicholas Indorf[1#], Darren Fahy[1], Geetika Singh[1], Jonathan Glickman[1#], Murray Resnick[1#], Lily Windholz[1], Andrew Billin[2^], Tim Watkins[2^], Jacqueline Brosnan-Cashman[1], Christina Jayson[1], Justin Lee[1#], Ben Glass[1], Andrew H. Beck[1], Janani Iyer[1#], Michael G. Drage[1#], Lara Murray[1], Robert Egger[1]

1) PathAI, Boston, MA
2) Gilead Sciences, Foster City, CA

* Contributed equally to study; # Employee of PathAI at the time of study; ^ Employee of Gilead Sciences at the time of study

Corresponding author:

Robert Egger
PathAI, Inc.
1325 Boylston Street, Suite 10000
Boston, MA 02215
robert.egger@pathai.com

Study Funding: This work was funded by PathAI, Inc. and Gilead Sciences.

Author Contributions: Conceptualization: AS-M, NP, YG, AB, TW, JB-C, CJ, JI, MGD, LM, RE; Methodology: AS-M, YZ, JZ, PM, DK, YL, NI, DF, GS, JG, MR, JL, RE; Investigation: NP, YG; Visualization: NP, YG, LM; Funding acquisition: N/A; Project administration: LW; Supervision: BG, AHB, JI, MGD, LM, RE; Writing – original draft: AS-M, NP, YG, JB-C, LM, RE; Writing – review & editing: AS-M, NP, YG, YZ, JZ, PM, DK, YL, NI, DF, GS, JG, MR, LW, AB, TW, JB-C, CJ, JL, BG, AHB, JI, MGD, LM, RE

Data and Code Availability: Model parameters for Liver Explore, and codes for model training and inference are not disclosed. Access requests for such a code will not be considered to safeguard PathAI's intellectual property. All source code for reproducing statistical analyses will be posted to a github repository prior to publication. Access to Liver Explore prediction heatmaps is available upon reasonable request to academic investigators without relevant conflicts of interest for non-commercial use who agree not to distribute the data. Access requests can be made to robert.egger@pathai.com.

Abstract

Background & Aims: Histologic staging of metabolic dysfunction-associated steatohepatitis (MASH) requires semiquantitative assessment of hepatocellular ballooning, steatosis, lobular inflammation, and fibrosis. We hypothesize that quantitative histologic analysis will better reflect the continuous distribution of histologic features, and thus the disease biology.

Methods: We developed an AI-powered digital pathology tool, Liver Explore, consisting of a suite of machine learning models that detect and classify liver tissue regions, lobular zones, cell types, and fibrosis subtypes from hematoxylin and eosin-stained whole slide images. Human interpretable features (HIFs) were extracted and computed that correspond to predicted substances. The correlation of Liver Explore HIFs with pathologist-provided MASH CRN grades and fibrosis stages, AIM-MASH-generated continuous CRN grades and stages, non-invasive biomarkers, transcriptomics, and outcomes was assessed in participants of the STELLAR-3 and STELLAR-4 trials (NCT04052516).

Results: Liver Explore predictions were consistent with manual pathologist annotations. Steatosis, lobular inflammation, hepatocellular ballooning, and fibrosis Liver Explore HIFs were significantly correlated with pathologist CRN grades/stages, while model-derived tissue and cell features revealed quantitative changes in the disease microenvironment as MASH progressed. Pathological and advanced fibrosis HIFs were correlated with non-invasive metrics of fibrosis and a gene signature associated with hepatic stellate cells. HIFs associated with nodular or advanced fibrosis and inflammation were associated with an increased risk of liver-related events in patients from STELLAR-3 and STELLAR-4.

Conclusions: The quantitative characterization of the liver disease microenvironment by Liver Explore delivers context relevant to MASH progression beyond the resolution afforded by categorical CRN scoring, highlighting the promise of this tool for broad applications in drug development, from enhancing understanding of mechanisms of action of novel MASH therapeutics to identifying histologic biomarkers for use in clinical trials.

Background

The liver is a critical and complex organ. Lobules of normal hepatocytes are interspersed between central veins and portal tracts, resulting in a gradient of oxygen levels manifesting as functional zones in the lobules [1]. Improvements in liver histology have improved our understanding of the microscopic patterns associated with liver diseases broadly [2]. Visual evidence that the normal liver microarchitecture has been disrupted by inflammation, injury, steatosis, and fibrosis within certain hepatic regions, such as foam cell arteriopathy and ductopenia in cases with chronic liver transplant rejection, guide a pathologist's decision-making [3,4].

However, manual pathologist assessment of liver histology is not without limitations. It is infeasible to exhaustively count cells or measure the area of histologic substances. While prior work has suggested that changes in liver composition not captured by standard scoring guidelines (lipid-associated macrophages and antigen-dependent CD8+ lymphocytes [5,6]) are relevant to metabolic dysfunction-associated steatohepatitis (MASH), these changes cannot be quantified by the human eye. Fibrosis evaluation represents a particular challenge for pathologists due to its reliance on special staining (e.g., Masson trichrome (MT) or picrosirius Red [7]) or imaging of adjacent unstained sections [8], rather than in tissue stained with hematoxylin and eosin (H&E). The inability to accurately assess fibrosis on H&E-stained slides has complicated the assessment of fibrosis in conjunction with other histologic features.

Opportunities remain for more in-depth understanding of disease-specific liver microarchitecture alterations. In MASH, levels of steatosis, inflammation, hepatocellular injury, and fibrosis within a biopsy inform a pathologist's assessment. However, patients present across a complex and multifaceted spectrum that can be challenging to capture with manual pathologist review, traditionally reliant on categorical staging [9–11]. While non-invasive tests (NITs), such as blood-based tests and biomarkers and imaging-based measurements [12] have demonstrated utility for MASH and other liver diseases, these tests do not offer a detailed mechanistic understanding of underlying tissue- and cell-level events and can be confounded by unrelated extrahepatic processes [13]. Moreover, the relationship of these metrics with histological changes is unclear. While sampling variability contributes to this shortcoming, the current categorical histological scoring paradigm renders it difficult to correlate histologic features with quantitative biomarker data. Thus, the ability to quantify liver histology at scale would greatly aid our understanding of liver diseases.

Artificial intelligence (AI)-powered digital pathology (DP) algorithms have the ability to provide unbiased, quantitative, and reproducible evaluation of whole slide images (WSIs) of histology specimens. AI tools to segment and classify cell types and tissue regions in WSIs have been developed for categorizing disease microenvironments [14,15]. Notably, algorithms have been developed to measure the MASH clinical research network (CRN) criteria (hepatocellular ballooning, steatosis, lobular inflammation, and fibrosis), including

qFIBS [16] and AIM-MASH [17]. Both algorithms have been demonstrated to improve inter-pathologist concordance for scoring fibrosis and recapitulate manually-scored fibrosis clinical trial endpoints, while AIM-MASH has been shown to improve pathologist reproducibility and corroborate manual endpoints for the additional CRN criteria, leading to its qualification by the European Medicines Agency [17–20]. However, both tools necessitate additional slides than H&E-stained biopsies for fibrosis detection, either unstained (qFIBS) [16] or MT-stained (AIM-MASH) [17]. The ability to assess fibrosis on a H&E-stained slide would potentially streamline liver pathology workflows.

Here, we report an approach for the comprehensive and quantitative characterization of liver microarchitecture: a suite of AI-powered algorithms designed to detect, classify, and quantify tissue regions, cell types, and collagen directly from a H&E-stained WSI of liver biopsies. This tool, Liver Explore (for research use only; not for use in diagnostic procedures), measures histologic features in a manner that correlates with established biomarkers of liver disease as well as with disease progression and regression.

Patients and Methods

Ethics

This study was conducted in accordance with the Declaration of Helsinki and Good Clinical Practice guidelines. Anonymized WSIs of H&E-stained liver biopsies were obtained from randomized controlled trials of MASH therapeutics, PathAI Diagnostics, Cleveland Clinic, Precision for Medicine, and Dr. Fabio Tavora as part of data sharing agreements between PathAI and each institution (**Supplementary Table S1**). All patients provided informed consent for future research and tissue histology, and approval by central institutional review boards was granted for each clinical trial [20–26].

Datasets

Model development datasets consisted of H&E-stained images (N=3,092) from clinical, commercial, academic, and diagnostic sources (**Supplementary Table S1-S2**). Model performance was evaluated on a subset of a held-out clinical cohort (**Supplementary Tables S1&S3**). Samples were scanned using AT2 or GT450 slide scanners (Leica Biosystems, Vista, CA) at 40x or 20x magnification (Tables S8-S9). Development and evaluation samples covered the spectrum of MASH severity to maximize model performance generalization. Model development datasets were randomly divided into training (~75%) and validation (~25%) at the patient level in a manner to ensure balance across relevant metadata, including MASH severity and scanner type. Each model used a subset of the full development set (**Supplementary Table S2**) to ensure split alignment across models while also allowing the flexibility to tailor splits per model.

Liver Explore features were further assessed on a held-out dataset consisting of WSIs of baseline and post-treatment biopsies WSIs from two completed MASH clinical trials of selonsertib, STELLAR-3 and STELLAR-4 (NCT04052516) [26], for which gene expression [27], NIT metrics, and clinical outcomes were available.

Model development

Annotations

Board-certified expert MASH pathologists (N=82) provided hand-drawn annotations (N=636,646) of histologic substances after completing a proficiency assessment.

Annotations were produced on a proprietary WSI viewing platform. Pathologists either entered point clicks on a cell nucleus or provided hand-drawn regions surrounding specific tissue-level substances (e.g., lobular inflammation, nodular fibrosis, portal tract). Annotation types per model are listed in **Supplementary Table S4**. A subset of each pathologist's annotations were reviewed after every annotation job for quality control. Definitions for each annotated substance were provided to pathologists (**Supplementary Table S4**).

Course annotations were used for developing the fibrosis classification model. Pathologists were instructed to draw circles around each fibrosis subtype, ensuring that there was only a single fibrosis subtype within the annotated region; surrounding liver parenchyma was also included in these annotations. Pathologists were provided a neighboring Masson Trichrome-stained section and, in some cases, a collagen overlay derived from iQMAI to assist.

Model Training

Liver Explore utilized previously trained artifact and tissue models [17]. Convolutional neural network models were trained for nucleus detection and cell and fibrosis subtype classification using pathologist annotations as previously described [14,17].

Image augmentations were performed prior to model training as previously described to ensure robustness to variations in tissue preparation [17,28,29]. Models were developed in an iterative manner. Based on an initial set of pathologist-provided annotations of cell and fibrosis subtypes, a first set of models was trained. At each iteration, model outputs were evaluated quantitatively and qualitatively, with additional annotations provided by pathologists in areas where the model made incorrect predictions. This cycle was repeated until performance converged to a satisfactory level. During this process, new slides would be chosen from the overall split to focus annotation collection on underperforming or underrepresented substances.

iQMAI

iQMAI training was previously described in cancer specimens [30]. A similar training procedure was used for iQMAI in liver biopsies. Briefly, MASH biopsies (N=88) from commercial and diagnostics laboratories (**Supplementary Table S1**) were digitized (Aperio AT2; Leica Biosystems, Vista, CA) at 40x magnification and 0.2522 microns per pixel and also imaged using quantitative multimodal anisotropy imaging (QMAI) [30]. Slides were randomized at the patient level into training (N=59, 67.0%), validation (N=19, 21.6%) and testing (N=10, 11.4%) sets in a manner to balance CRN fibrosis stage (0-4) and slide origin and ensuring no data leakage between development and test sets. Brightfield and QMAI images were registered for pixel-level alignment prior to model training [30].

The iQMAI architecture is modified from the U-Net architecture proposed by Ronneberger and colleagues [31]. In each double convolution layer, each convolution is followed by a SiLU activation and a batch normalization for stable training. Additionally, 2x bilinear upsampling with batch norm is used in the upsampling blocks instead of transposed convolution, because transposed convolution tends to lead to gridding artifacts, which bilinear upsampling effectively mitigates. During training, the input patch size is 572 x 572 pixels with three channels, representing the H&E RGB image; the output patch size is 196 x 196 pixels (due to using valid padding) with a single channel, representing the predicted intensity of the QMAI signal. A sigmoid activation is used before the output to

convert the logits to values within [0,1]. Prior to training, each input H&E RGB patch was augmented with random flipping and rotation, stain color augmentation with S-DOTA and ContriMix] [28,29], followed by brightness, hue, and noise augmentations. A linear combination of perceptual loss and mean squared error loss was used as the objective, as previously described [30,32].

Liver Explore deployment

Liver Explore deployment consists of a series of chained computational steps including model inference, overlay generation and HIF generation.

First, the artifact model identifies all regions of slide background, artifact (tissue folds, marker ink, blur) and evaluable tissue within the WSI. All subsequent quantification is carried out in areas of evaluable tissue only.

The output of the fibrosis classification model output is passed through a series of transformations that 1) smooth the heatmap to remove pixelated or gridded prediction edges, 2) consolidate model predictions in connected components consisting of: nodular fibrosis and complete septal fibrosis; complete septal fibrosis and incomplete septal fibrosis; portal tract and structural collagen; portal tract and central vein into the class with the majority of pixels in the connected component; and consolidate connected components consisting of structural collagen, portal tract, central vein, incomplete septal fibrosis, complete septal fibrosis and nodular fibrosis into nodular fibrosis only, 3) remove portal tract and central vein predictions below a minimum area threshold, and 4) fill holes within portal tracts and central veins.

The processed portal tract and central vein predictions are then extracted into a new heatmap forming the basis of the Zonal Overlay. Each portal tract and central vein is dilated by 150 μm to form Zones 1 and 3, respectively, and Zone 2 is formed by another 150um dilation around the outside of each zone. The amount of dilation was chosen based on prior work and [33] qualitative review of different dilations with several liver pathologists. Zonal dilation occurs only within the evaluable tissue region and does not extend into regions of structural collagen or advanced fibrosis.

To form the fibrosis overlay, the transformed fibrosis classification heatmap is further processed by the zonal heatmap, such that periportal fibrosis only occurs in zone 1 and fibrosis outside of this region is classified as perisinusoidal fibrosis. Finally, the iQMAI collagen detection heatmap is used to mask out all non-collagen pixels, leaving the granular fibrosis subtyping overlay on H&E.

The cell overlay is produced by assigning cell classification model predictions to detected nuclei from the nuclei detection model. Ballooned hepatocyte cells are enforced to only exist within the hepatocellular ballooning tissue region.

All model overlays are used to generate human interpretable features, or HIFs (N>1,000) which describe the liver architecture through the quantification of tissue regions, cell

types, zonal regions, and fibrosis subtypes using the process previously described [34]. Briefly, features were calculated through the combination of heatmap overlays, yielding relative areas of zonal regions or fibrosis types, relative cell densities, cell density ratios, and substances within a certain distance of others.

Model Evaluation

Model predictions were verified on a subset of a clinical trial cohort completely held-out from development (N=50; **Supplementary Table S3**).

Steatosis, lobular inflammation, and hepatocellular ballooning predictions were validated previously [18]. Predictions of portal inflammation, interface hepatitis, normal interface, bile ducts, blood vessels, and normal hepatocyte regions were qualitatively reviewed by a pathologist on the held-out dataset. To verify zonal segmentation accuracy, the number of portal tracts and central veins identified in the overlay were compared to the number counted by pathologists across the held-out set, as follows. Each contiguous region of portal tract- or central vein-labelled pixels was considered a separate entity, and contiguous regions were summed. Three pathologists independently annotated the center of every portal tract and central vein in the entire WSI, and the median number of the annotations was calculated. Overlay and pathologist counts were compared using ICC(2,1) [35].

Cell and fibrosis classification models were evaluated using a nested pairwise framework [36], which iteratively compares model predictions to pathologist labels in a consensus-free manner and calculates the desired metric based on all (prediction, label) pairs across all annotators. Each pathologist (X) is compared to the labels from another pathologist (Y) for all pairs of pathologists X and Y. 95% confidence intervals were derived by bootstrapping. To assess cell detection and classification accuracy, image frames (75μm x 75μm; n=400 across 50 WSIs) were selected to represent a diversity of disease, cell densities, and underlying tissue regions. To ensure representation for ballooned hepatocytes, 50 frames were specifically selected by an internal pathologist for having a high likelihood of containing at least 1 ballooned hepatocyte. Five pathologists independently exhaustively annotated every nucleus in every frame using only the underlying H&E image.

Fibrosis predictions were evaluated for collagen detection accuracy and for classification accuracy. iQMAI-derived collagen heatmaps were compared to QMAI images of 10 slides held-out from model development. First, a pathologist provided manual annotations of regions of early-stage fibrosis (perisinusoidal/periportal), septal fibrosis (incomplete/complete septa), and nodular fibrosis. These annotations were used to directly label the polarization output and iQMAI prediction heatmaps. Next, the total area of each fibrosis subtype was determined using the iQMAI labelled heatmaps as well as

the polarization images, which were compared to measure HIF similarity. Furthermore, to ensure that the iQMAI and polarization collagen signals aligned at the pixel level (not just HIF level), labelled images were also aligned and compared using Intersection over Union.

To assess collagen classification, frames (1mm x 1mm; n=125 across 50 WSIs) were selected across fibrotic and non-fibrotic regions to ensure representation across a range of fibrosis subtypes and disease severity. Exhaustive annotations on every frame were collected from three pathologists (**Supplementary Figure S1**). Given the difficulty in annotating fine fibers of fibrosis on H&E, pathologists were asked to coarsely label fibrosis types with circles around homogenous fibers and were given unclassified iQMAI heatmaps and MT serial sections for reference. Pixel-wise comparison between pathologist labels and predicted classes was performed using nested pairwise analysis.

Clinical utility analysis

WSI with <4 mm2 tissue, <1000 total hepatocytes, and <200 hepatocytes per mm^2 tissue were excluded, resulting in n=3199 WSI across STELLAR-3 and STELLAR-4. Continuous CRN fibrosis scores were derived using AIM-MASH+ [17] deployed on paired MT slides. All analyses were performed in Python.

Expression of transcriptomic modules

Gene expression data for baseline and post-treatment biopsies in the STELLAR-3 and STELLAR-4 datasets were processed into log-normalized counts per million as previously described [27], then used to calculate the HSC-2 score based on the published list of genes and corresponding weights [37].

Correlation with outcomes

Liver Explore features in baseline biopsies were independently tested for association with liver-related events (LRE) in STELLAR-3 and in STELLAR-4. Features with >3% missing values across the baseline sample set or relating to artifact or background areas, nuclear morphology features, as well as raw areas and raw counts were excluded. Area proportion and count proportion features were logit-transformed after imputing zeros to half the minimal non-zero value and imputing ones to the average between 1 and the maximal non-one value. All other features were log-transformed after imputing zeros to the half-minimal value. The entire feature set was then normalized using the robust Z-score method.

Cox proportional hazards models for each transformed baseline feature in each data set were fitted to LRE data (N=440 for STELLAR-3; N=629 for STELLAR-4), including progression to cirrhosis for STELLAR-3, with treatment arms as covariates. Samples with missing feature values were excluded. A significance threshold of p=0.05 after false discovery rate correction was used.

Results

Model Outputs

Liver Explore generates interpretable overlays that visualize model-derived segmentations of input WSI into biologically relevant regions: artifact, tissue, cell, lobular zones, and fibrosis subtype (**Figure 1**). The artifact model identifies slide background, as well as regions of tissue that may interfere with downstream feature analysis (e.g., blur, tissue folds, debris). Predicted artifact regions are excluded from further analysis. The tissue model detects and displays lobular inflammation, hepatocellular ballooning, steatosis, portal inflammation, and interface hepatitis, tissue regions pertinent to liver function (bile ducts and blood vessels), and normal hepatocyte regions. The cell model detects and exhaustively segments all nuclei and classifies each as one of eleven cell classes, including various immune and inflammatory cells (e.g., plasma cells, lymphocytes, macrophages) and relevant hepatocyte subclasses (e.g., ballooned hepatocytes). The zonal model identifies portal tracts, central veins, and the surrounding lobular zones (zones 1, 2, and 3). Lastly, the fibrosis model detects and exhaustively segments all collagen into non-pathological collagen and categories of fibrosis associated with different degrees of liver disease severity (e.g., perisinusoidal and nodular fibrosis).

Liver Explore overlays enable the calculation of over one thousand quantitative human-interpretable features (HIFs; **Supplementary Figures S2–S5, Supplementary Tables S5-S6**) [34], which capture the granularity and complexity of the liver microarchitecture. The simplest HIFs quantify the area of predicted regions or total or relative cell type counts across the entire WSI. Combining Liver Explore-derived predictions of different regions (e.g., the cell and fibrosis subtype overlays), yields novel spatial information that is difficult to evaluate by manual review (e.g., the proximity of hepatocellular ballooned cells to perisinusoidal fibrosis). Overall, these HIFs provide a quantitative and comprehensive characterization of liver microarchitecture.

Model Performance

To assess model prediction accuracy, we compared the cell, collagen, and zonal model outputs to pathologist manual annotations or other state-of-the-art approaches; tissue model performance was previously described [17].

Cell Detection and Classification

Cell model performance accuracy was assessed using nested pairwise comparison [36] to pathologist annotations on frames sampled from a held-out dataset (N=50 WSI). Sensitivity and positive predictive values (PPVs) for model-derived and pathologist-derived predictions are shown in **Table 1**.

Predictions of hepatocyte classes are comparable to manual pathologist annotations (**Table 1**). Ballooned hepatocyte performance was slightly below pathologist performance with model sensitivity of 0.54 [0.24, 0.66] and PPV 0.52 [0.26, 0.63] compared to

pathologist sensitivity and PPV of 0.67 [0.51, 0.75]. Model immune cell predictions showed much stronger sensitivity (0.84 [0.81, 0.86] compared to pathologist sensitivity of 0.54 [0.52, 0.56]), while having slightly lower PPV than pathologists (0.39 [0.37, 0.41] compared to 0.54 [0.52, 0.56]). Of the immune cell types, lymphocytes were detected with greatest accuracy (sensitivity 0.63 [0.59, 0.66] and PPV 0.43 [0.39, 0.46]).

Collagen Detection and Fibrosis Classification

Liver Explore leverages inferred quantitative multimodal anisotropic imaging (iQMAI), which predicts collagen structures on H&E-stained WSI [30]. To evaluate iQMAI-based fibrosis predictions on H&E-stained WSI, we directly compared Liver Explore-derived collagen predictions with direct collagen detection via QMAI, a polarization-based collagen imaging modality [30], on 10 held-out slides. Collagen-related features derived from both techniques were highly correlated, with Pearson r values >0.9 (**Supplementary Table S7**). Furthermore, a direct comparison of collagen signal between the two modalities using intersection over unions (IOU) revealed highly similar pixel-level predictions: overall collagen had an IOU of 0.51 [0.40, 0.62], which was similar for thicker collagen types, while thinner early-stage fibrosis (perisinusoidal/periportal) had an IOU of 0.25 [0.13, 0.38] (**Supplementary Table S7**). IOU was influenced by the quality of alignment between model-derived and QMAI heatmaps. These results support the accuracy of Liver Explore's iQMAI-based collagen detection.

Fibrosis subtype predictions were then compared to pathologist annotations of fibrosis subtypes. Model-identified fibrosis types relevant for MASH CRN scoring demonstrated sensitivity and PPV similar to manual pathologist annotation (**Table 2**). Perisinusoidal fibrosis, the very thin strands that are critical for distinguishing CRN stages 0–2, has sensitivity 0.60 [0.53, 0.66] and PPV 0.37 [0.28, 0.44] compared to pathologist sensitivity of 0.51 [0.43, 0.59] and PPV of 0.51 [0.43, 0.59]. The most accurate model predictions were pathological fibrosis and advanced fibrosis groupings, while periportal fibrosis was difficult to accurately identify for both model and pathologists.

Zonal Segmentation

To assess the accuracy of the zonal segmentation model in detecting portal tracts and central veins, we compared model-predicted counts of these features against counts derived by a panel of three pathologists in a held-out dataset (N=50). For portal tracts, the ICC(2,1) between model-predicted and median pathologist (n=3) was 0.65 [0.49, 0.78], comparable to the average inter-pathologist ICC(2,1) of 0.67 [0.59, 0.74] (**Supplementary Table S8**). Portal tract prediction accuracy was higher in early stages (0.76 [0.59, 0.87]; F-stages 0–2) compared to inter-pathologist ICC (0.56 [0.44,0.67]). However, model portal tract accuracy was lower in advanced fibrotic stages (0.54 [0.29, 0.75]; F-stages 3–4) than inter-pathologist correlation (0.71 [0.55,0.82]). For central veins, the overall correlation between predicted and consensus counts was 0.60 [0.48, 0.71], compared to inter-pathologist correlation of 0.88 [0.84, 0.91], with minimal variation between early and advanced fibrosis.

Clinical utility of histologic features

We next sought to examine Liver Explore's potential to support evaluations relevant to clinical trials. Using baseline and post-treatment liver biopsies from STELLAR-3 and STELLAR-4 [26], which enrolled patients with stage 3 and 4 fibrosis, respectively, we investigated Liver Explore-derived features for understanding and predicting MASH disease progression.

Association between Liver Explore histological features and CRN scores

MASH CRN scoring involves semi-quantitative grading/staging criteria for lobular inflammation, hepatocellular ballooning, steatosis, and fibrosis and are used for histologic endpoint measurements [9][38]. We identified a Liver Explore feature quantifying the relevant histology of each CRN component and assessed the distributions of these features across pathologist-assessed grades/stages (**Figure 2**). Features demonstrated strong correlation to manual scores, further confirming the ability of Liver Explore to accurately capture histologic features relevant for clinical trials.

Given the importance of fibrosis for determining MASH severity, we assessed changes in liver biopsy composition accompanying fibrosis progression. Using Liver Explore, we quantified tissue, fibrosis, and cellular composition in WSIs of STELLAR-3 and STELLAR-4 biopsies. We also derived sub-ordinal fibrosis stages using AIM-MASH+, which predicts continuous MASH CRN grades and stages [17], to granularly characterize changes occurring during fibrosis progression. The proportionate area of normal hepatocyte regions, representing the amount of healthy tissue, decreased as a function of fibrosis stage (**Figure 3A**), consistent with the notion that healthy tissue decreases as disease progresses. Interestingly, at continuous fibrosis scores above 3.0, large expansions of portal inflammation, blood vessels (**Figure 3A**), nodular fibrosis (**Figure 3B**), immune cell density, and fibroblast density were observed (**Figure 3C**), progressively increasing with fibrosis severity. Intermediate disease stages were characterized by elevated complete septal (bridging) fibrosis, incomplete septal fibrosis, and perisinusoidal fibrosis, which appeared first in early-stage disease with maximum percentages occurring near continuous scores of 4.0. Steatosis tissue and steatotic hepatocyte density both peaked in biopsies with continuous fibrosis scores between 1.0 and 3.0 (**Figure 3A,C**).

We next assessed spatial relationships between liver components that cannot be gleaned from manual microscopic examination. Tissue regions surrounding perisinusoidal fibrosis were strongly enriched for ballooned hepatocytes (8.82x density compared to other tissue regions) and moderately enriched for infiltrating immune cells (macrophages=1.85x, neutrophils=1.64x, lymphocytes=1.47x and eosinophils=1.37x density ratio; **Figure 3D**). However, steatotic hepatocytes and normal hepatocytes showed little to no spatial bias towards fibrosis regions. These results suggest that the enrichment of ballooned hepatocytes and certain inflammatory cells near perisinusoidal fibrosis is due to a cell-type specific mechanism.

Association of HIFs with non-invasive MASH biomarkers

NITs and blood-based biomarkers are widely used for diagnosis [39–41] and assessment of MASH severity [12], as they assess fibrosis (FIB-4, ELF, FibroScan, NFS, APRI), inflammation (C-reactive protein [CRP]), cell death (CK-18-M30, and -M65), liver enzyme levels (aspartate transaminase [AST] and alanine transaminase [ALT]), blood triglyceride, and cholesterol levels. We performed correlation analysis of Liver Explore HIFs with available NITs from patients enrolled in STELLAR-3 and STELLAR-4 as additional validation (**Figure 4A**).

Comparisons of fibrosis NITs with fibrosis HIFs revealed strong positive associations between all fibrosis NITs and model-derived total fraction of pathological fibrosis area. Additionally, these NITs were more strongly correlated with the fraction of advanced fibrosis area than with early-stage fibrosis area, potentially indicating a higher sensitivity of fibrosis NITs for advanced fibrosis and cirrhosis [13]. Similarly, Aspartate Transaminase Platelet Ratio Index (APRI), which detects advanced fibrosis and cirrhosis, more strongly correlated with advanced than early-stage fibrosis. At the cellular level, fibrosis NITs correlated positively with the fraction of fibroblasts and negatively with normal hepatocyte area and counts. The association of Liver Explore fibrosis HIFs with fibrosis NITs provides orthogonal validation of model-predicted features.

Having observed increased portal inflammation and immune cells with increasing fibrosis score (**Figure 3A**), we also sought to confirm this association using NITs. Indeed, the fraction of inflammation area in tissue was strongly correlated with fibrosis NITs, suggesting that inflammation could accompany advanced fibrosis development [42].

Of the liver function tests, AST demonstrated stronger correlations with most HIFs than ALT, including positive correlations with early-stage fibrosis, hepatocellular ballooning area, and ballooned hepatocyte counts. CK-18-M30 and -M65 also demonstrated strong positive correlation with hepatocellular ballooning features (area and cell count) and negative correlation with normal hepatocyte features (area and cell count).

Finally, we compared Liver Explore HIFs with transcriptomic data. The association of Liver Explore HIFs with single cell-based transcriptomic modules related to MASH severity [37] is shown in **Figure 4B**. Interestingly, despite Liver Explore not directly quantifying hepatic stellate cells, pooled Liver Explore HIFs by predicted substance were observed to strongly correlate with the hepatic stellate cells-2 (HSC-2) transcriptomic module, representing hepatic stellate cells and containing genes related to collagen and matrix remodeling [37]. To further examine this association, we compared individual HIFs to the HSC-2 module score. HSC-2 was positively correlated with the fractions of total pathological fibrosis area and immune cells and negatively correlated with the fraction of normal hepatocytes.

Liver Explore HIFs are prognostic of MASH progression

Given the changes in Liver Explore-predicted features accompanying MASH progression, we hypothesized that these features may be prognostic of outcome. To address this question, we tested individual HIFs in baseline biopsies for association with liver-related events (LREs) in STELLAR-3, where 95% of LREs were progression to cirrhosis (ref) and STELLAR-4. Several groups of features demonstrated statistically significant prognostic capabilities (**Figure 5**). In both trials, features quantifying inflammation and nodular or advanced fibrosis were associated with higher risk of LRE, consistent with the enrichment of nodular fibrosis and high densities of immune cells and inflammation in high fibrosis stages (**Figure 3**). Conversely, features relating to incomplete septal fibrosis and areas of landmark zones in STELLAR-3, and features relating to complete septal fibrosis and densities of hepatocytes in proximity to fibrosis in STELLAR-4, were associated with lower LRE risk. These results are likely driven by the distinct histologies within each study (e.g., breakdown of landmark regions in F3, progression to nodular fibrosis in F4).

Discussion

Liver disorders, including MASH, involve complex changes in hepatic microarchitecture. While liver pathologists are tasked with recognizing specific histologic patterns associated with these diseases, manual evaluations are semiquantitative and difficult to scale. AI-based DP approaches, such as Liver Explore, provide an opportunity to improve our understanding of liver diseases, broadly, compared to manual histologic evaluation alone. While we focused this study on MASH, the histologic substances identified are applicable to pathology studies across hepatology.

The application of DP to liver diseases has been actively studied, yielding automated predictions of the MASH CRN steatosis, hepatocellular ballooning, lobular inflammation, and fibrosis scores [16–20,40,41]. Still, additional insights into liver disease can be gained through quantifying histological structures and their relationships. Liver Explore addresses this outstanding need via highly accurate, scalable, pixel-level classifications of tissue regions, cell types, zonal regions, and fibrosis subtypes in H&E-stained WSIs. The cell model performance for immune cells and lymphocytes is excellent and in line with what we would expect to observe in an H&E-stained image. Ballooned hepatocytes, difficult for pathologists to agree upon [45], are detected by Liver Explore with greater accuracy and sensitivity than published metrics from other models [46]. The correlation of Liver Explore HIFs with related NIT and/or transcriptomic metrics further confirms its specificity. The performance and scalability of Liver Explore allows for well-powered statistical comparisons across patient populations that can reveal novel insights into the liver disease microenvironment.

The ability to capture information spanning fibrosis subtypes and cell types from a single H&E-stained image spatial relationships between these features to be characterized, an untenable task across separate images (e.g., characterizing cell types and fibrosis from H&E- and MT-stained specimens, respectively). Liver Explore simultaneously predicts tissue regions, cell types, and fibrosis subtypes from an H&E-stained WSI, enabling detailed examination of spatial relationships between these substances and allowing novel tissue composition analyses. Here, we found that ballooned hepatocytes were specifically enriched in regions near perisinusoidal fibrosis, a previously unquantifiable fibrosis subtype indicative of early disease progression [47]. Additionally, we described histological features characterizing the transition from steatosis-dominated to fibrosis-dominated pathology with increasing disease severity, as well as stage-specific and -independent features prognostic of progression to LRE.

The insights gained from such granular analyses have the potential to inform MASH therapeutic development. A deeper understanding of the relationship between processes manifesting in distinct histologic events (e.g., inflammation and fibrosis) may enable the therapeutic design, enhance the understanding of the mechanism(s) of action for existing therapeutics, and inform clinical trial design. As trials incorporating combination therapies become more common in hepatology, knowledge of the interplay between disease

processes will inform selection of the agents to be assessed. Additionally, the ability to more thoroughly detect histologic changes may lead to a better understanding of the features associated with disease progression and regression and inform biomarker development for precision medicine approaches.

While this work provides an exciting path forward in liver pathology, it is not without limitations. The analysis cohort used here only included MASH subjects from F3 and F4-enrolled trials. As such, other liver disease populations were not assessed herein. Future studies of Liver Explore will assess its features in a broader spectrum of liver diseases and severities. In addition, while Liver Explore accurately predicts cell types for which annotations were collected, cells with similar morphology cannot be distinguished. As such, the model cannot accurately predict these cells – such as hepatic stellate cells – without accurate annotations as input. Other techniques remain necessary for the quantitation of these liver cell types.

In conclusion, Liver Explore provides a comprehensive, ML-enabled approach to fully quantify granular histopathologic features in liver disease. The accuracy of this approach, combined with its ability to yield spatial insights into liver histology not feasible with traditional analyses, has the potential to yield critical insights for the treatment of patients with MASH and other liver diseases.

Acknowledgements

The authors would like to thank the pathologists who contributed to this study, as well as the software engineering and ML operations teams at PathAI for developing the systems and pipelines used for model development and feature extraction.

References

[1] Plesec T, Cruise M. Liver Pathology. Pediatric Gastrointestinal and Liver Disease, Elsevier; 2021, p. 717–32.e3.

[2] Torbenson M, Washington K. Pathology of liver disease: advances in the last 50 years. Hum Pathol 2020;95:78–98.

[3] Liu G, Butany J, Wanless IR, Cameron R, Greig P, Levy G. The vascular pathology of human hepatic allografts. Hum Pathol 1993;24:182–8.

[4] Rubin R, Muñoz SJ. Clinicopathologic features of late hepatic dysfunction in orthotopic liver transplants. Hum Pathol 1993;24:643–51.

[5] Ganguly S, Rosenthal SB, Ishizuka K, Troutman TD, Rohm TV, Khader N, et al. Lipid-associated macrophages' promotion of fibrosis resolution during MASH regression requires TREM2. Proc Natl Acad Sci U S A 2024;121:e2405746121.

[6] Burtis AEC, DeNicola DMC, Ferguson ME, Santos RG, Pinilla C, Kriss MS, et al. Ag-driven CD8 + T cell clonal expansion is a prominent feature of MASH in humans and mice. Hepatology 2025;81:591–608.

[7] Standish RA, Cholongitas E, Dhillon A, Burroughs AK, Dhillon AP. An appraisal of the histopathological assessment of liver fibrosis. Gut 2006;55:569–78.

[8] Astbury S, Grove JI, Dorward DA, Guha IN, Fallowfield JA, Kendall TJ. Reliable computational quantification of liver fibrosis is compromised by inherent staining variation. J Pathol Clin Res 2021;7:471–81.

[9] Kleiner DE, Brunt EM, Van Natta M, Behling C, Contos MJ, Cummings OW, et al. Design and validation of a histological scoring system for nonalcoholic fatty liver disease. Hepatology 2005;41:1313–21.

[10] Bedossa P, FLIP Pathology Consortium. Utility and appropriateness of the fatty liver inhibition of progression (FLIP) algorithm and steatosis, activity, and fibrosis (SAF) score in the evaluation of biopsies of nonalcoholic fatty liver disease. Hepatology 2014;60:565–75.

[11] Pai RK, Kleiner DE, Hart J, Adeyi OA, Clouston AD, Behling CA, et al. Standardising the interpretation of liver biopsies in non-alcoholic fatty liver disease clinical trials. Aliment Pharmacol Ther 2019;50:1100–11.

[12] Tincopa MA, Loomba R. Noninvasive tests to assess fibrosis and disease severity in metabolic dysfunction-associated steatotic liver disease. Semin Liver Dis 2024;44:287–99.

[13] Patel K, Sebastiani G. Limitations of non-invasive tests for assessment of liver fibrosis. JHEP Rep 2020;2:100067.

[14] Abel J, Jain S, Rajan D, Padigela H, Leidal K, Prakash A, et al. AI powered quantification of nuclear morphology in cancers enables prediction of genome instability and prognosis. Npj Precision Oncology 2024;8:134.

[15] Markey M, Kim J, Goldstein Z, Gerardin Y, Brosnan-Cashman J, Javed SA, et al. Spatial mapping of immunosuppressive CAF gene signatures in H&E-stained images using additive multiple instance learning. bioRxiv 2024. https://doi.org/10.1101/2024.08.12.607604.

[16] Liu F, Goh GB-B, Tiniakos D, Wee A, Leow W-Q, Zhao J-M, et al. QFIBS: An automated technique for quantitative evaluation of fibrosis, inflammation, ballooning, and steatosis in patients with nonalcoholic steatohepatitis. Hepatology 2020;71:1953–66.

[17] Iyer JS, Juyal D, Le Q, Shanis Z, Pokkalla H, Pouryahya M, et al. AI-based automation of enrollment criteria and endpoint assessment in clinical trials in liver diseases. Nat Med

2024:1–10.
[18] Pulaski H, Harrison SA, Mehta SS, Sanyal AJ, Vitali MC, Manigat LC, et al. Clinical validation of an AI-based pathology tool for scoring of metabolic dysfunction-associated steatohepatitis. Nat Med 2025;31:315–22.
[19] Abdurrachim D, Lek S, Ong CZL, Wong CK, Zhou Y, Wee A, et al. Utility of AI digital pathology as an aid for pathologists scoring fibrosis in MASH. J Hepatol 2025;82:898–908.
[20] Harrison SA, Alkhouri N, Ortiz-Lasanta G, Rudraraju M, Tai D, Wack K, et al. A phase IIb randomised-controlled trial of the FFAR1/FFAR4 agonist icosabutate in MASH. J Hepatol 2025. https://doi.org/10.1016/j.jhep.2025.01.032.
[21] Loomba R, Sanyal AJ, Nakajima A, Neuschwander-Tetri BA, Goodman ZD, Harrison SA, et al. Pegbelfermin in patients with nonalcoholic steatohepatitis and stage 3 fibrosis (FALCON 1): A randomized phase 2b study. Clin Gastroenterol Hepatol 2024;22:102–12.e9.
[22] Abdelmalek MF, Sanyal AJ, Nakajima A, Neuschwander-Tetri BA, Goodman ZD, Lawitz EJ, et al. Pegbelfermin in patients with nonalcoholic steatohepatitis and compensated cirrhosis (FALCON 2): A randomized phase 2b study. Clin Gastroenterol Hepatol 2024;22:113–23.e9.
[23] Loomba R, Cejvanovic V, Iyer JS, Kjær MS, Krarup N, Sejling A, Pericas JM. Comparison of the effects of semaglutide on liver histology in patients with non-alcoholic steatohepatitis cirrhosis between machine learning model assessment and pathologist evaluation 2022.
[24] Harrison SA, Bashir MR, Guy CD, Zhou R, Moylan CA, Frias JP, et al. Resmetirom (MGL-3196) for the treatment of non-alcoholic steatohepatitis: a multicentre, randomised, double-blind, placebo-controlled, phase 2 trial. Lancet 2019;394:2012–24.
[25] Hirschfield GM, Shiffman ML, Gulamhusein A, Kowdley KV, Vierling JM, Levy C, et al. Seladelpar efficacy and safety at 3 months in patients with primary biliary cholangitis: ENHANCE, a phase 3, randomized, placebo-controlled study. Hepatology 2023;78:397–415.
[26] Harrison SA, Wong VW-S, Okanoue T, Bzowej N, Vuppalanchi R, Younes Z, et al. Selonsertib for patients with bridging fibrosis or compensated cirrhosis due to NASH: Results from randomized phase III STELLAR trials. J Hepatol 2020;73:26–39.
[27] Conway J, Pouryahya M, Gindin Y, Pan DZ, Carrasco-Zevallos OM, Mountain V, et al. Integration of deep learning-based histopathology and transcriptomics reveals key genes associated with fibrogenesis in patients with advanced NASH. Cell Rep Med 2023;4:101016.
[28] Nguyen TH, Juyal D, Li J, Prakash A, Nofallah S, Shah C, et al. ContriMix: Scalable stain color augmentation for domain generalization without domain labels in digital pathology. arXiv [eessIV] 2023.
[29] Gullapally SC, Zhang Y, Mittal NK, Kartik D, Srinivasan S, Rose K, et al. Synthetic DOmain-Targeted Augmentation (S-DOTA) improves model generalization in digital pathology. arXiv [csCV] 2023. https://doi.org/10.48550/ARXIV.2305.02401.
[30] Nguyen TH, Zhang J, Hipp J, Chhor G, Griffin M, Le N, et al. Quantification of collagen and associated features from H&E-stained whole slide pathology images across cancer types using a physics-based deep learning model. bioRxiv 2025. https://doi.org/10.1101/2025.03.17.643273.
[31] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. arXiv [csCV] 2015.
[32] Zhang R, Isola P, Efros AA, Shechtman E, Wang O. The unreasonable effectiveness of

deep features as a perceptual metric. arXiv [csCV] 2018.
[33] Ho H, Zhang E. Virtual lobule models are the key for multiscale biomechanical and pharmacological modeling for the liver. Front Physiol 2020;11:1061.
[34] Diao JA, Wang JK, Chui WF, Mountain V, Gullapally SC, Srinivasan R, et al. Human-interpretable image features derived from densely mapped cancer pathology slides predict diverse molecular phenotypes. Nat Commun 2021;12:1613.
[35] Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979;86:420–8.
[36] Gerardin Y, Shamshoian J, Shen J, Le N, Prezioso J, Abel J, et al. Improved statistical benchmarking of digital pathology models using pairwise frames evaluation. arXiv [csCV] 2023. https://doi.org/10.48550/ARXIV.2306.04709.
[37] Fred RG, Steen Pedersen J, Thompson JJ, Lee J, Timshel PN, Stender S, et al. Single-cell transcriptome and cell type-specific molecular pathways of human non-alcoholic steatohepatitis. Sci Rep 2022;12:13484.
[38] Loomba R, Ratziu V, Harrison SA, NASH Clinical Trial Design International Working Group. Expert Panel Review to Compare FDA and EMA Guidance on Drug Development and Endpoints in Nonalcoholic Steatohepatitis. Gastroenterology 2022;162:680–8.
[39] Loomba R, Amangurbanova M, Bettencourt R, Madamba E, Siddiqi H, Richards L, et al. MASH Resolution Index: development and validation of a non-invasive score to detect histological resolution of MASH. Gut 2024;73:1343–9.
[40] Cathcart J, Barrett R, Bowness JS, Mukhopadhya A, Lynch R, Dillon JF. Accuracy of non-invasive imaging techniques for the diagnosis of MASH in patients with MASLD: A systematic review. Liver Int 2025;45:e16127.
[41] Gbadamosi SO, Evans KA, Brady BL, Hoovler A. Noninvasive tests and diagnostic pathways to MASH diagnosis in the United States: a retrospective observational study. J Med Econ 2025;28:314–22.
[42] Koyama Y, Brenner DA. Liver inflammation and fibrosis. J Clin Invest 2017;127:55–64.
[43] Taylor-Weiner A, Pokkalla H, Han L, Jia C, Huss R, Chung C, et al. A Machine Learning Approach Enables Quantitative Measurement of Liver Histology and Disease Monitoring in NASH. Hepatology 2021;74:133–47.
[44] Loomba R, Noureddin M, Kowdley KV, Kohli A, Sheikh A, Neff G, et al. Combination Therapies Including Cilofexor and Firsocostat for Bridging Fibrosis and Cirrhosis Attributable to NASH. Hepatology 2021;73:625–43.
[45] Brunt EM, Clouston AD, Goodman Z, Guy C, Kleiner DE, Lackner C, et al. Complexity of ballooned hepatocyte feature recognition: Defining a training atlas for artificial intelligence-based imaging in NAFLD. J Hepatol 2022;76:1030–41.
[46] Kleiner DE, Clouston AD, Goodman Z, Guy CD, Brunt EM, Lackner C, et al. FRI-208 qBallooning: AI-based ballooned hepatocyte detection and quantification by second harmonic generation/two photon excitation microscopy. J Hepatol 2024;80:S467.
[47] Bataller R, Brenner DA. Liver fibrosis. J Clin Invest 2005;115:209–18.

Tables

Table 1. Model sensitivity and positive predictive values (PPV) for model-derived and pathologist-derived cell type predictions. *indicates where the model-derived predictions pass non-inferiority test, two-sided alpha=0.05 with margin 0.1

Cell class	Model Sensitivity [95% CI]	Model PPV [95% CI]	Pathologist Sensitivity [95% CI]	Pathologist PPV [95% CI]
all hepatocytes*	0.81 [0.77, 0.85]*	0.83 [0.80, 0.85]*	0.86 [0.85, 0.87]	0.86 [0.85, 0.87]
normal hepatocytes*	0.52 [0.42, 0.62]	0.57 [0.52, 0.61]*	0.59 [0.54, 0.62]	0.59 [0.54, 0.62]
other hepatocytes*	0.54 [0.46, 0.60]*	0.41 [0.37, 0.45]*	0.44 [0.40, 0.48]	0.44 [0.40, 0.48]
steatotic hepatocytes*	0.64 [0.59, 0.68]*	0.66 [0.63, 0.71]*	0.67 [0.64, 0.70]	0.67 [0.64, 0.70]
ballooned hepatocytes	0.54 [0.24, 0.66]	0.52 [0.26, 0.63]	0.67 [0.51, 0.75]	0.67 [0.51, 0.75]
all immune cells*	0.84 [0.81, 0.86]*	0.39 [0.37, 0.41]	0.54 [0.52, 0.56]	0.54 [0.52, 0.56]
eosinophils	0.54 [0.38, 0.68]	0.45 [0.33, 0.60]	0.62 [0.53, 0.72]	0.62 [0.53, 0.72]
Neutrophils	0.66 [0.53, 0.80]*	0.18 [0.12, 0.25]	0.45 [0.35, 0.55]	0.45 [0.35, 0.55]
lymphocytes*	0.63 [0.59, 0.66]*	0.43 [0.39, 0.46]	0.51 [0.49, 0.54]	0.51 [0.49, 0.54]
plasma cells*	0.29 [0.14, 0.43]*	0.20 [0.11, 0.31]	0.27 [0.15, 0.39]	0.27 [0.15, 0.39]
macrophages*	0.55 [0.51, 0.59]*	0.16 [0.14, 0.18]*	0.24 [0.22, 0.26]	0.24 [0.22, 0.26]
all other cells	0.35 [0.30, 0.40]	0.61 [0.58, 0.65]*	0.57 [0.54, 0.60]	0.57 [0.54, 0.60]
other cells	0.21 [0.17,	0.73 [0.68,	0.54 [0.51, 0.57]	0.54 [0.52, 0.57]

	0.26]	0.77]*		
fibroblasts*	0.55 [0.47, 0.62]*	0.27 [0.22, 0.31]	0.37 [0.31, 0.42]	0.37 [0.31, 0.42]

Table 2. Model sensitivity and positive predictive values (PPV) for model-derived and pathologist-derived fibrosis subtype predictions. *indicates where the model-derived predictions pass non-inferiority test, two-sided alpha=0.05 with margin 0.2

Fibrosis Subtype	Model Sensitivity [95% CI]	Model PPV [95% CI]	Pathologist Sensitivity [95% CI]	Pathologist PPV [95% CI]
pathological fibrosis*	0.92 [0.84, 0.96]*	0.76 [0.65, 0.84]	0.87 [0.80, 0.91]	0.87 [0.80, 0.91]
early-stage fibrosis*	0.58 [0.50, 0.64]*	0.40 [0.32, 0.45]*	0.49 [0.42, 0.56]	0.49 [0.42, 0.56]
periportal fibrosis*	0.14 [0.07, 0.27]*	0.18 [0.11, 0.27]*	0.22 [0.15, 0.31]	0.22 [0.15, 0.31]
perisinusoidal fibrosis*	0.60 [0.53, 0.66]*	0.37 [0.28, 0.44]	0.51 [0.43, 0.59]	0.51 [0.43, 0.59]
advanced fibrosis*	0.87 [0.79, 0.92]*	0.72 [0.59, 0.81]	0.84 [0.77, 0.89]	0.84 [0.77, 0.89]
nodular fibrosis*	0.81 [0.53, 0.94]*	0.45 [0.25, 0.58]	0.63 [0.44, 0.74]	0.63 [0.44, 0.74]
complete septal fibrosis	0.28 [0.15, 0.44]	0.50 [0.33, 0.68]*	0.55 [0.43, 0.68]	0.55 [0.43, 0.68]
incomplete septal fibrosis	0.26 [0.15, 0.43]	0.40 [0.26, 0.54]	0.45 [0.31, 0.59]	0.45 [0.31, 0.59]
non-pathological collagen	0.67 [0.45, 0.81]	0.81 [0.66, 0.90]*	0.85 [0.75, 0.93]	0.85 [0.75, 0.93]
portal collagen	0.50 [0.34, 0.69]	0.79 [0.66, 0.90]	0.78 [0.66, 0.88]	0.78 [0.66, 0.88]
perivenular fibrosis	0.59 [0.45, 0.77]	0.50 [0.30, 0.66]	0.65 [0.48, 0.76]	0.65 [0.48, 0.76]
structural collagen	0.74 [0.28, 0.93]	0.80 [0.41, 0.95]	0.85 [0.58, 0.96]	0.85 [0.58, 0.96]
nonspecific collagen*	0.28 [0.14, 0.52]	0.09 [0.06, 0.12]	0.37 [0.23, 0.49]	0.37 [0.23, 0.49]

Figures

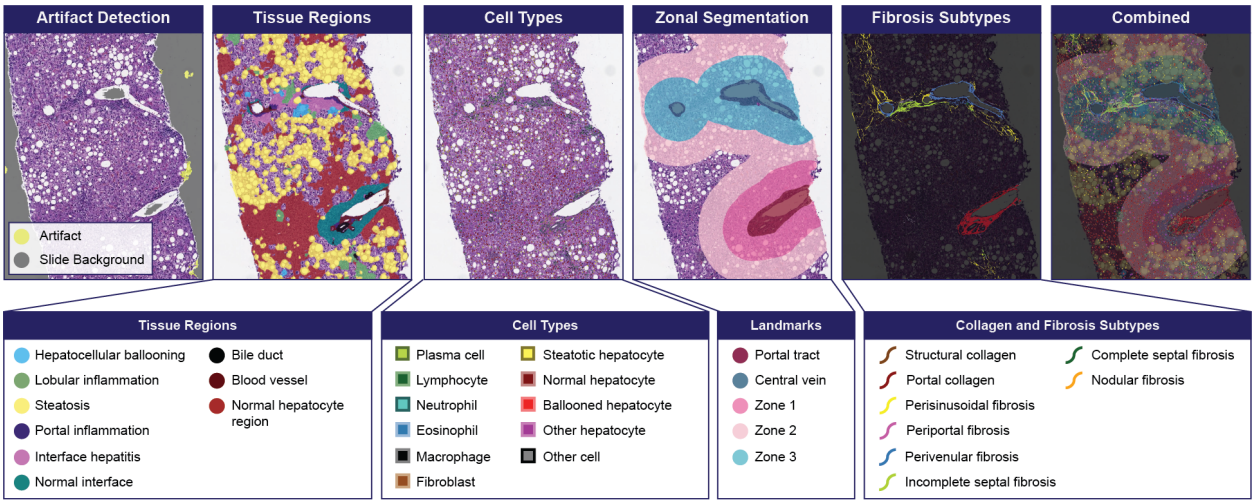


Figure 1. Liver Explore overlays and detected classes.

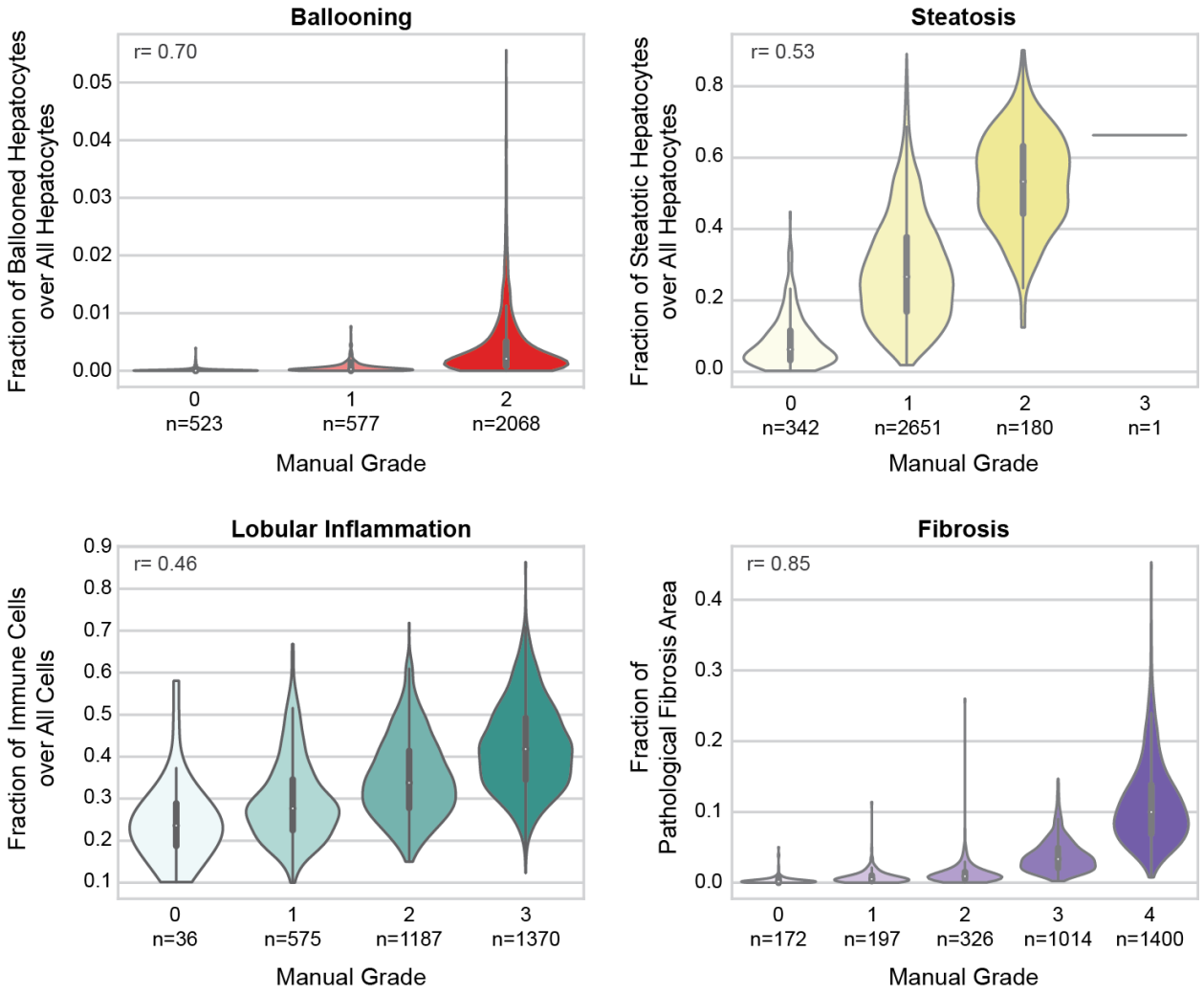


Figure 2. Distribution of Liver Explore-quantified histological substances within pathologist-assessed MASH CRN grades and fibrosis stages

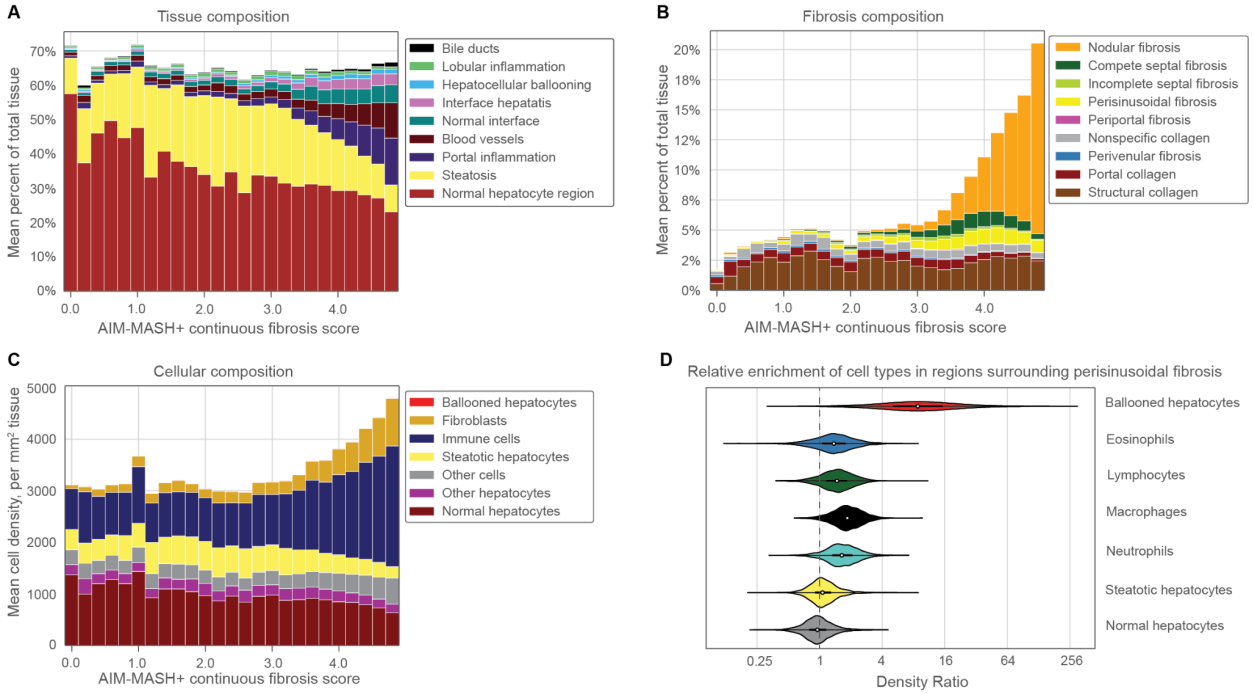


Figure 3. Granular changes in MASH liver tissue composition revealed at sub-ordinal and spatially differentiated resolution with increasing disease severity. Continuous fibrosis scores were binned into 5 equal-sized intervals per integer score, and bars show feature values averaged across all samples falling within each interval for **A)** predicted tissue types, **B)** predicted fibrosis subtypes, and **C)** predicted cell types. **D**) Relative enrichment of each cell type in regions surrounding perisinusoidal fibrosis. Enrichment for each biopsy image is defined by the ratio of cell density within 40 μm of perisinusoidal fibrosis divided by cell density in tissue excluding those proximal regions.

A

	Noninvasive markers of fibrosis					Liver function tests		Markers of inflammation/apoptosis			Circulating Lipids	
Fibrosis subtypes	FIB-4	NFS	ELF	FibroScan	APRI	AST	ALT	CRP	CK-18(M30)	CK-18(M65)	Cholesterol	Triglycerides
Area of PATHOLOGICAL FIBROSIS	0.54	0.49	0.57	0.62	0.44	0.28	0.017	0.12	0.045	0.072	-0.086	-0.073
Area of ADVANCED FIBROSIS	0.55	0.52	0.57	0.61	0.42	0.23	-0.023	0.09	-0.0012	0.02	-0.1	-0.11
Area of EARLY STAGE FIBROSIS	0.27	0.12	0.34	0.29	0.35	0.42	0.27	0.17	0.27	0.3	0.034	0.11
Area of INCOMPLETE SEPTAL FIBROSIS	-0.11	-0.22	-0.085	-0.19	-0.003	0.14	0.22	-0.047	0.095	0.093	0.087	0.18
Inflammatory tissue												
Area of LOBULAR INFLAMMATION	0.027	-0.12	0.076	-0.024	0.2	0.37	0.35	0.12	0.26	0.28	0.032	0.044
Area of INFLAMMATION	0.5	0.41	0.53	0.5	0.46	0.35	0.1	0.17	0.11	0.13	-0.039	-0.11
Fibroblasts												
Counts of FIBROBLASTS	0.41	0.36	0.43	0.44	0.31	0.17	0.0015	-0.0022	0.03	0.055	-0.062	-0.048
Immune cells												
Counts of LYMPHOCYTES	0.33	0.32	0.34	0.3	0.26	0.17	-0.016	0.18	-0.036	-0.021	-0.049	-0.085
Counts of PLASMA CELLS	0.33	0.35	0.36	0.34	0.26	0.18	-0.00063	0.19	-0.03	-0.0081	-0.077	-0.057
Counts of MACROPHAGES	0.25	0.21	0.31	0.31	0.28	0.27	0.13	0.28	0.14	0.14	0.044	-0.053
Ballooning												
Area of HEPATOCELLULAR BALLOONING	0.29	0.19	0.33	0.31	0.37	0.44	0.29	0.22	0.28	0.29	0.075	0.077
Counts of BALLOONED HEPATOCYTES	0.27	0.17	0.31	0.28	0.35	0.44	0.29	0.23	0.32	0.33	0.069	0.11
Steatosis												
Area of STEATOSIS	-0.28	-0.28	-0.21	-0.14	-0.066	0.14	0.32	0.062	0.22	0.22	0.084	0.18
Counts of STEATOTIC HEPATOCYTES	-0.33	-0.33	-0.28	-0.25	-0.13	0.063	0.25	0.032	0.17	0.16	0.11	0.16
Normal hepatocytes												
Area of NORMAL HEPATOCYTE REGION	-0.068	-0.059	-0.18	-0.19	-0.23	-0.32	-0.3	-0.25	-0.27	-0.28	-0.14	-0.11
Counts of NORMAL HEPATOCYTES	-0.22	-0.22	-0.31	-0.34	-0.3	-0.31	-0.19	-0.29	-0.17	-0.18	-0.11	-0.025

B

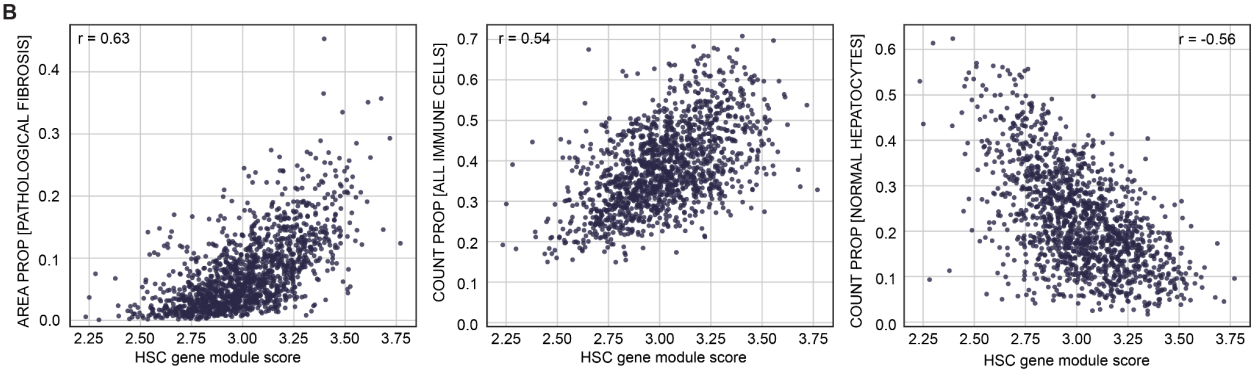


Figure 4. Correlation of Liver Explore features with orthogonal data modalities. **A**) Liver Explore features representing MASH-relevant disease features correlated against blood-based biomarkers of liver function and metabolism (Spearman r). 'Area' features are proportionate areas out of all evaluable tissue; 'count' features are cell proportions out of all cells. Data includes baseline and post-treatment time points from both studies. **B**) Liver Explore features correlated with gene expression signature of hepatic stellate cells (Spearman r).

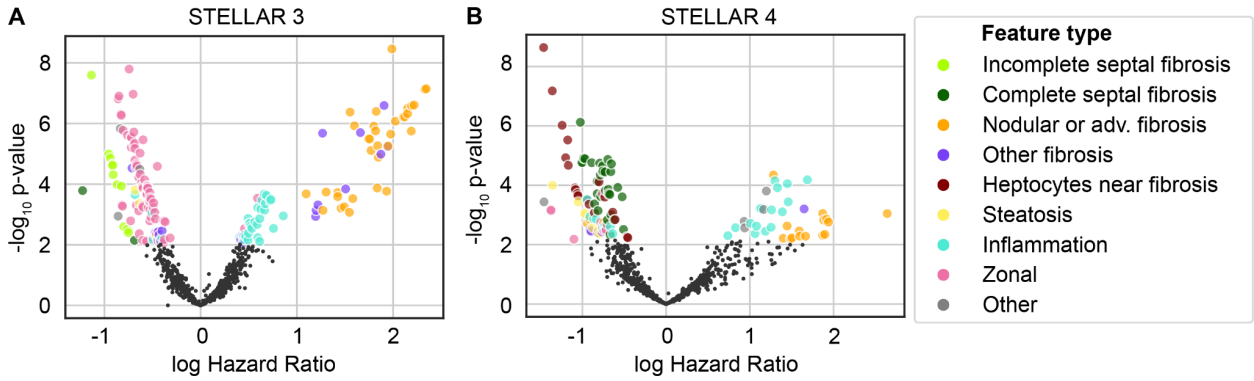


Figure 5. Association between Liver Explore features in baseline samples and clinical outcomes from **A**) STELLAR-3 and **B**) STELLAR-4. Each point represents a single feature's nominal p-value and effect size from Cox regression. Large colored markers indicate features exceeding FDR-adjusted significance.

Supplementary Tables

Supplementary Table S1. Development and evaluation dataset characteristics. The evaluation dataset was used for verifying all models while the development datasets were used for development of cell classification, iQMAI collagen detection, and fibrosis classification models. Nuclei, tissue, and artifact models are described elsewhere.

Sample Source Type	Sample Source	NCT Number	Number of WSIs
Development Dataset			
Diagnostic samples	PathAI Diagnostics	N/A	663
Diagnostic samples	PathAI Diagnostics	N/A	546
Diagnostic samples	PathAI Diagnostics	N/A	368
Clinical (Ph2)	BMS	NCT03486899	321
Clinical (Ph2)	BMS	NCT03486912	222
Clinical (Ph2)	Novo Nordisk	NCT03884075	180
Commercial	Cleveland Clinic	N/A	160
Commercial	Precision for Medicine	N/A	146
Clinical (Ph2)	Madrigal	NCT02912260	122
Commercial	Precision for Medicine	N/A	116
Academic samples	Cleveland Clinic	N/A	90
Diagnostic Samples	Fabio Tavora's Lab	N/A	50
Diagnostic samples	PathAI Diagnostics	N/A	39
Diagnostic samples	PathAI Diagnostics	N/A	35
Clinical (Ph2)	CymaBay	NCT03551522	14
Commercial	Precision for Medicine	N/A	10
Commercial	Precision for Medicine	N/A	10
Evaluation Dataset			
Clinical (Ph2)	Northsea	NCT04052516	100

Supplementary Table S2. Summary of the model development dataset for Liver Explore. Datasets were all aligned so that each model's train/val/test slides were drawn from the overall train/val/test split. When available, MASH CRN fibrosis scores and NAS component grades were used to ensure train/val/internal-test sets held representation of all disease activity.

	Group	**Cell Classification Model** N_{WSI} = 2,201 Development set (%)	**Fibrosis Classification Model** N_{WSI} = 1,070 Development set (%)	**iQMAI Collagen Detection Model** N_{WSI} = 88 Development set (%)
Data split	Training Validation Internal-Test	73% 25% 2%	77% 18% 5%	67% 22% 11%
Scanner	Aperio AT2 Aperio GT450	52% 48%	85% 15%	100% 0%
Data Source	Clinical Sources Commercial Sources	73% 27%	42% 58%	0% 100%

Supplementary Table S3. Summary of the Liver Explore evaluation dataset. This held-out set was used for verification of the Cell overlay, Zonal overlay, secondary substances of the tissue model, qualitative collagen detection model, fibrosis classification model.

	Group	**Evaluation set (%)** N_{WSI} = 50
Scanner	Aperio AT2	100%
Collection Type	Biopsy	100%
Data Source	Clinical	100%
Fibrosis Stage	0 1 2 3 4	20% 20% 20% 20% 20%
Ballooning Grade	0 1 2	66% 30% 4%
Steatosis Grade	0 1 2 3	0% 36% 48% 16%
Lobular Inflammation Grade	0 1 2 3	2% 72% 26% 0%

Supplementary Table S4. Overlay class definitions

Artifact Overlay	
Class	**Definition/Clarification**
Artifact	Image artifact from tissue preparation, staining, and scanning which includes regions of blur, tissue folds, and debris.
Evaluable Tissue	Tissue that is free from artifact. All quantitative model results are generated from this region.
Background	Slide background
Tissue Overlay	
Class	**Definition/Clarification**
Steatosis	Regions of macrovesicular steatosis
Lobular inflammation	Inflammatory clusters consisting of at least 3 inflammatory cells. The inflammation should involve the hepatic lobule and may include sinusoidal inflammatory cells, however, sinusoidal inflammatory cells alone should be interpreted with caution.
Hepatocellular ballooning	Enlarged hepatocytes (1.5-2x at least) with reticulated cytoplasm with or without Mallory's hyaline
Portal inflammation	Increased inflammatory cells of at least a mild degree within the portal areas.
Interface hepatitis	Inflammation and hepatocyte destruction occurring at the limiting plate
Normal Interface	No evidence of interface hepatitis
Bile ducts	Bile ducts and ductules not including the lumen
Blood vessels	Any discernible blood vessel, capillary, venule, vein, arteriole, artery not including the lumen
Normal hepatocyte region	Normal liver with no evidence of inflammation, ballooning, swelling, glycogenosis, steatosis (macrovesicular, small droplet, microvesicular)
Cell Overlay	

Class	Definition/Clarification
Normal hepatocytes	Normal hepatocytes including pigmented hepatocytes.
Ballooned hepatocytes	Ballooned hepatocytes including those with and without mallory denk bodies
Steatotic hepatocytes	Nuclei of macrovesicular steatosis as well as small droplet macrovesicular steatosis.
Other hepatocytes	All other hepatocytes should be in this category including: ● Hepatocellular glycogenosis ● Hepatocellular swelling ● Microvesicular steatosis
Macrophages	Macrophages including pigmented and foamy macrophages and Kupffer cells that are clearly distinguishable
Lymphocytes	Normal lymphocytes
Eosinophils	Normal eosinophils
Neutrophils	Normal neutrophils
Plasma cells	Annotated cells should have a plasmacytoid appearance and the presence of at least one of the following: ● Perinuclear hof ● 'Clock Face' chromatin
Fibroblasts	Normal fibroblasts
Other Cells	This category is for any other cell nuclei that does not fit in the prior categories. Most of them are one of: ● Vascular non-sinusoidal endothelial cells ● Sinusoidal endothelial cells ● Normal bile duct epithelial cells ● Smooth muscle cells
Zonal Overlay	
Class	**Definition/Clarification**

Portal tracts	This does not include residual portal tracts within septae which should be either complete septal fibrosis or nodular fibrosis.
Central veins	Central vein predictions should include the lumen and the vein wall and may include circumferential perivenular fibrosis.
Zone 1	The 150um hepatocyte region around portal tracts. Regions of nodular fibrosis, complete septal fibrosis, and structural collagen are not included in any zone.
Zone 2	The 150um hepatocyte region adjacent to Zone 1 or Zone 3. Regions of nodular fibrosis, complete septal fibrosis, and structural collagen are not included in any zone.
Zone 3	The 150um hepatocyte region around central veins. Regions of nodular fibrosis, complete septal fibrosis, and structural collagen are not included in any zone.
Fibrosis Overlay	
Class	**Definition/Clarification**
Portal collagen	Collagen within the approximate original outline of the portal tract. This does not include residual portal tracts within septae which should be labeled either complete septal fibrosis or nodular fibrosis. Large portal tracts (definition: bile ducts that have columnar biliary epithelium) should be annotated as “structural collagen”. Structural collagen can also be used to annotate collagen from the perihilar connective tissue, etc.
Perivenular fibrosis	Circumferential fibrosis around the rim of a central vein. This does not include chicken-wire perisinusoidal fibrosis
Periportal fibrosis	Pathological fibrosis extending from a portal tract that does not connect to another portal tract or central vein.
Perisinusoidal fibrosis	This is the “chicken-wire” fibrosis that is not connected to a portal tract. This also includes thin fibrosis extending away from septal/nodular fibrosis.

Incomplete septal fibrosis	Linear fibrosis that does not form a complete bridge, or extends off the biopsy.
Complete septal fibrosis	Fibrosis that bridges two vascular structures.
Nodular fibrosis	Fibrosis bordering/surrounding a nodule. A cirrhotic nodule might not be completely encircled by fibrosis if the nodule extends beyond the biopsy border.
Structural collagen	This includes subcapsular fibrosis and large non-pathologic fibrous septae which may include large blood vessels. Large biliary structures with columnar epithelium and likely intrahepatic septae. Large vessels like intrahepatic septae with thick walls, small bile ducts, and/or large dilated veins
Nonspecific collagen	Any collagen that is detected but does not clearly belong in any of the above categories. This category tends to include collagen predictions on non-liver tissue and is largely ignored in analysis.

Supplementary Table S5. Human-Interpretable Feature Types in Liver Explore

	Feature Type	Feature Property	Relevant classes	Examples
Whole slide features	Individual region overlay features: Fibrosis, Tissue, Zones	Areas Area proportions	See Fibrosis, Tissue, Zonal Ontologies	AREA (MM2) OF [PATHOLOGICAL FIBROSIS]
	Individual Cell features	Counts Count proportions Densities Density ratios	See Cell Ontology	COUNT [ALL HEPATOCYTES] IN [TISSUE]
	Cell-Fibrosis proximity features	Counts Count proportions	Cell classes/groups within "TISSUE SURROUNDING (40UM) <fibrosis type>"	COUNT [ALL HEPATOCYTES] IN [TISSUE SURROUNDING (40UM) ADVANCED FIBROSIS]
	Cross-model region features: Fibrosis in Zones, Fibrosis in Tissue regions, Tissue in Zones	Areas Area proportions	See fibrosis and tissue ontologies. Zones 1, Zone 2, Zone 3	AREA PROP [[BILE DUCTS] IN [ADVANCED FIBROSIS REGION]] OVER [ADVANCED FIBROSIS REGION]
Object-level features	Large object features	MEAN[]	Portal Tracts, Central Veins, Zone 1, Zone 3	MEDIAN [PERIMETER (MM) OF [PORTAL TRACTS]]
	Nuclei features	MEAN[], STD[]	Normal hepatocytes, ballooned hepatocytes, steatotic hepatocytes, fibroblasts, lymphocytes, macrophages	MEAN[STEATOTIC HEPATOCYTE_NUCLEUS_CIRCULARITY]

Supplementary Table S6. Human-Interpretable Feature Names in Liver Explore

Feature Name Format	Example Feature
AREA (MM2) OF []	AREA (MM2) OF [PATHOLOGICAL FIBROSIS]
AREA (MM2) OF [] IN []	AREA (MM2) OF [PERISINUSOIDAL FIBROSIS] IN [ZONE 3]
AREA PROP [] OVER []	AREA PROP [PERISINUSOIDAL FIBROSIS] OVER [PATHOLOGICAL FIBROSIS]
AREA PROP [[] IN []] OVER []	AREA PROP [[STEATOSIS] IN [ZONE 1]] OVER [STEATOSIS]
AREA PROP [[] IN []] OVER [[] IN []]	AREA PROP [[PERISINUSOIDAL FIBROSIS] IN [ZONE 3]] OVER [[PERISINUSOIDAL FIBROSIS] IN [PARENCHYMA]]
AREA RATIO [[] IN []] OVER [[] IN []]	AREA RATIO [[PERISINUSOIDAL FIBROSIS] IN [ZONE 3]] OVER [[PERISINUSOIDAL FIBROSIS] IN [ZONE 1]]
COUNT [] IN []	COUNT [ALL HEPATOCYTES] IN [TISSUE]
COUNT PROP [[] IN []] OVER []	COUNT PROP [[ALL IMMUNE CELLS] IN [PARENCHYMA]] OVER [ALL IMMUNE CELLS]
COUNT PROP [[] OVER []] IN []	COUNT PROP [[LYMPHOCYTES] OVER [ALL IMMUNE CELLS]] IN [ADVANCED FIBROSIS REGION]
COUNT PROP [[] IN []] OVER [[] IN []]	COUNT PROP [[BALLOONED HEPATOCYTES] IN [ZONE 1]] OVER [[BALLOONED HEPATOCYTES] IN [PARENCHYMA]]
DENSITY (COUNT/MM2) []	DENSITY (COUNT/MM2) [PORTAL TRACTS]
DENSITY (COUNT/MM2) [] IN []	DENSITY (COUNT/MM2) [ALL IMMUNE CELLS] IN [ADVANCED FIBROSIS REGION]
DENSITY RATIO [[] IN []] OVER [[] IN []]	DENSITY RATIO [[BALLOONED HEPATOCYTES] IN [ZONE 3]] OVER [[BALLOONED HEPATOCYTES] IN [ZONE 1]]
MEDIAN AREA (MM2) OF []	MEDIAN AREA (MM2) OF [PORTAL TRACTS]
MEDIAN [COUNT [] IN []]	MEDIAN [COUNT [ALL CELLS] IN [PORTAL TRACTS]]
MEDIAN [DENSITY (COUNT/MM2) [] IN []]	MEDIAN [DENSITY (COUNT/MM2) [ALL IMMUNE CELLS] IN [PORTAL TRACTS]]
MEDIAN [PERIMETER (MM) OF []]	MEDIAN [PERIMETER (MM) OF [PORTAL TRACTS]]
MEDIAN [COUNT PROP [[] OVER []] IN []]	MEDIAN [COUNT PROP [[BALLOONED HEPATOCYTES] OVER [ALL HEPATOCYTES]] IN [ZONE 3]]

MEDIAN [CENTROID DISTANCE (MM) FROM [] TO NEAREST []]	MEDIAN [CENTROID DISTANCE (MM) FROM [PORTAL TRACTS] TO NEAREST [CENTRAL VEINS]]
MEAN[]	MEAN[STEATOTIC HEPATOCYTE_NUCLEUS_CIRCULARITY]
STD[]	STD[LYMPHOCYTE_NUCLEUS_AREA_IN_UM2]

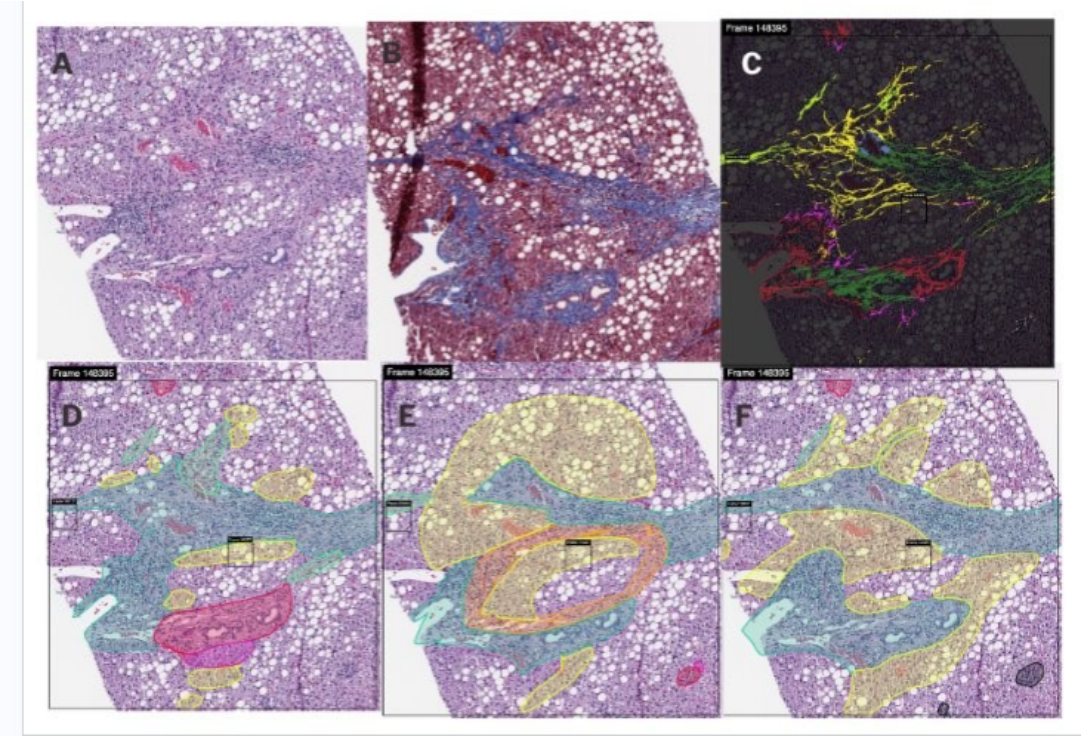
Supplementary Table S7. Comparison collagen and fibrosis features derived from iQMAI and QMAI.

	Slide #	1	2	3	4	5	6	7	8	9	10		
Measurement	**Fibrosis Stage:**	F0	F0	F1	F1	F2	F2	F3	F3	F4	F4	**Pearson Correlation**	**Mean IOU**
AREA (MM2) OF [ALL COLLAGEN]	**QMAI**	0.34	0.74	0.27	0.25	0.19	0.35	0.30	0.08	0.54	0.34	0.913	0.51 [0.40, 0.62] (N=10)
	iQMAI	0.38	0.69	0.24	0.28	0.19	0.32	0.28	0.20	0.75	0.40		
AREA (MM2) OF [EARLY STAGE FIBROSIS]	**QMAI**	0.01	0.01	0.05	0.02	0.07	0.00	0.03	0.01	0.11	0.03	0.917	0.25 [0.13, 0.38] (N=8)
	iQMAI	0.01	0.00	0.02	0.00	0.06	0.00	0.01	0.02	0.14	0.03		
AREA (MM2) OF [SEPTAL FIBROSIS]	**QMAI**	0.00	0.00	0.00	0.00	0.03	0.00	0.05	0.06	0.32	0.09	0.989	0.61 [0.49, 0.73] (N=8)
	iQMAI	0.00	0.00	0.00	0.00	0.03	0.00	0.04	0.13	0.44	0.10		
AREA (MM2) OF [NODULAR FIBROSIS]	**QMAI**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.08	0.992	0.55 [0.17, 0.93] (N=2)
	iQMAI	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.09		
AREA (MM2) OF [NON-PATHOLOGICAL COLLAGEN]	**QMAI**	0.32	0.73	0.21	0.23	0.09	0.35	0.23	0.01	0.09	0.14	0.991	0.65 [0.53, 0.78] (N=10)
	iQMAI	0.37	0.69	0.23	0.27	0.10	0.32	0.23	0.04	0.14	0.19		

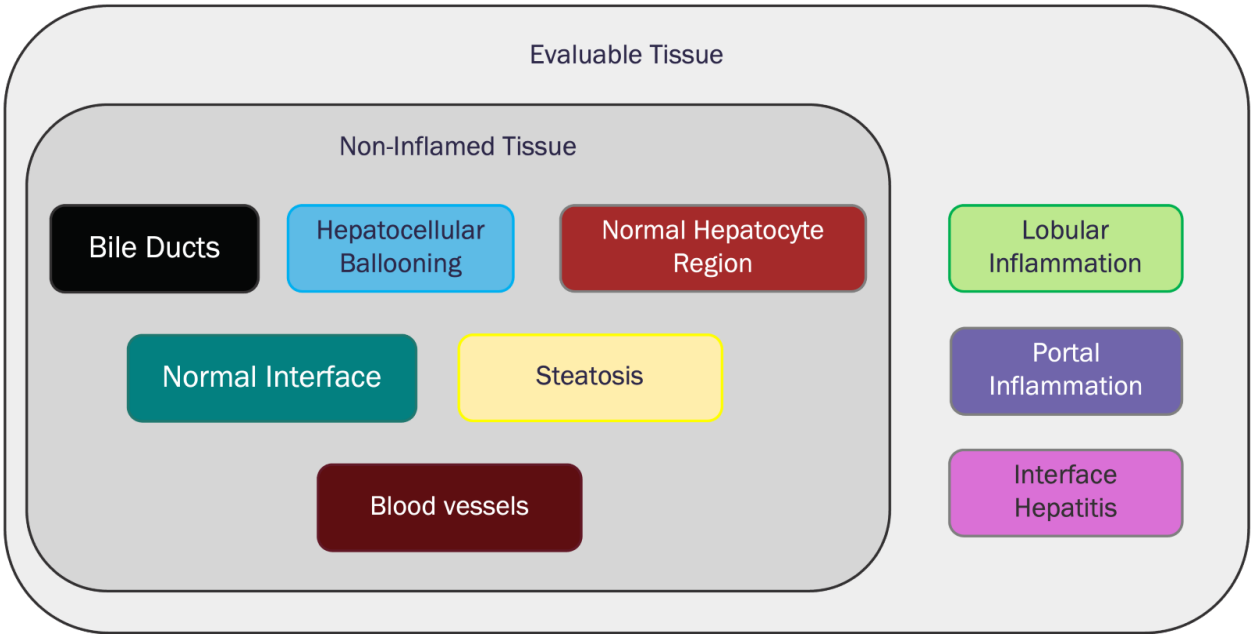
Supplementary Table S8. Evaluation of zonal model performance.

Component	Cohort	Model vs Consensus Correlation (ICC)	Average inter-pathologist correlation (ICC)
Portal tracts	All slides (N=50)	0.65 [0.49, 0.78]	0.67 [0.59, 0.74]
	F-stage 0–2 (N=30)	0.76 [0.59, 0.87]	0.56 [0.44, 0.67]
	F-stage 3–4 (N=20)	0.54 [0.29, 0.75]	0.71 [0.55, 0.82]
Central Veins	All slides (N=50)	0.60 [0.48, 0.71]	0.88 [0.84, 0.91]
	F-stage 0–2 (N=30)	0.58 [0.43, 0.73]	0.84 [0.76, 0.89]
	F-stage 3–4 (N=20)	0.50 [0.25, 0.69]	0.88 [0.82, 0.92]

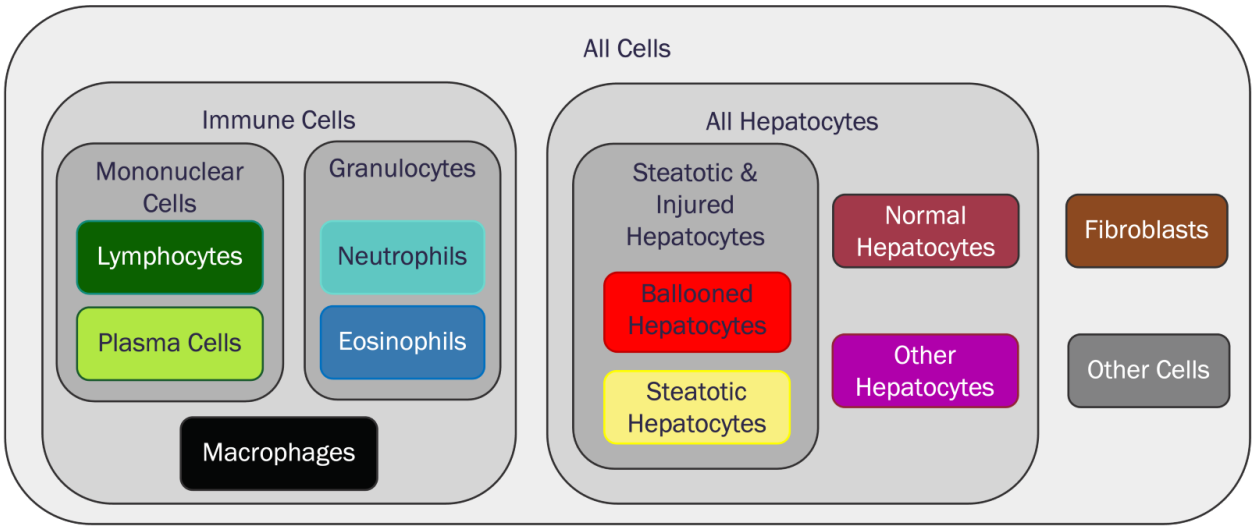
Supplementary Figures



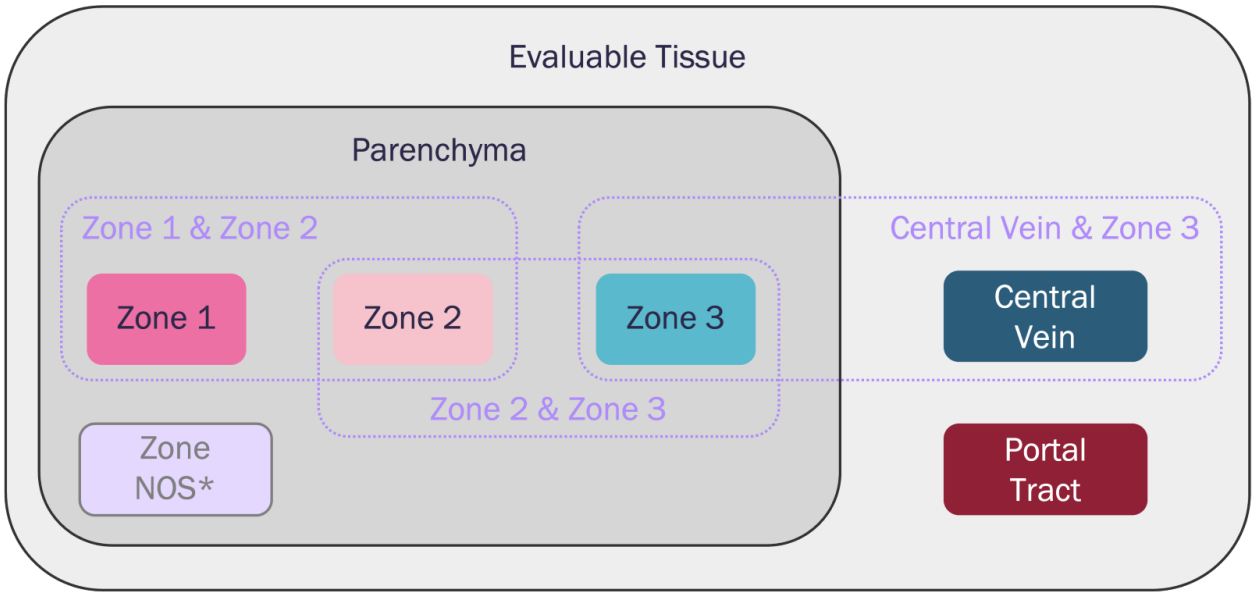
Supplementary Figure S1. Example of Fibrosis Subtype frame annotations. **A.** H&E image without any overlays. **B**. Corresponding trichrome image. **C.** Liver Explore Fibrosis Subtype model overlay. **D-F**. Examples of fibrosis subtype annotations from three different pathologists.



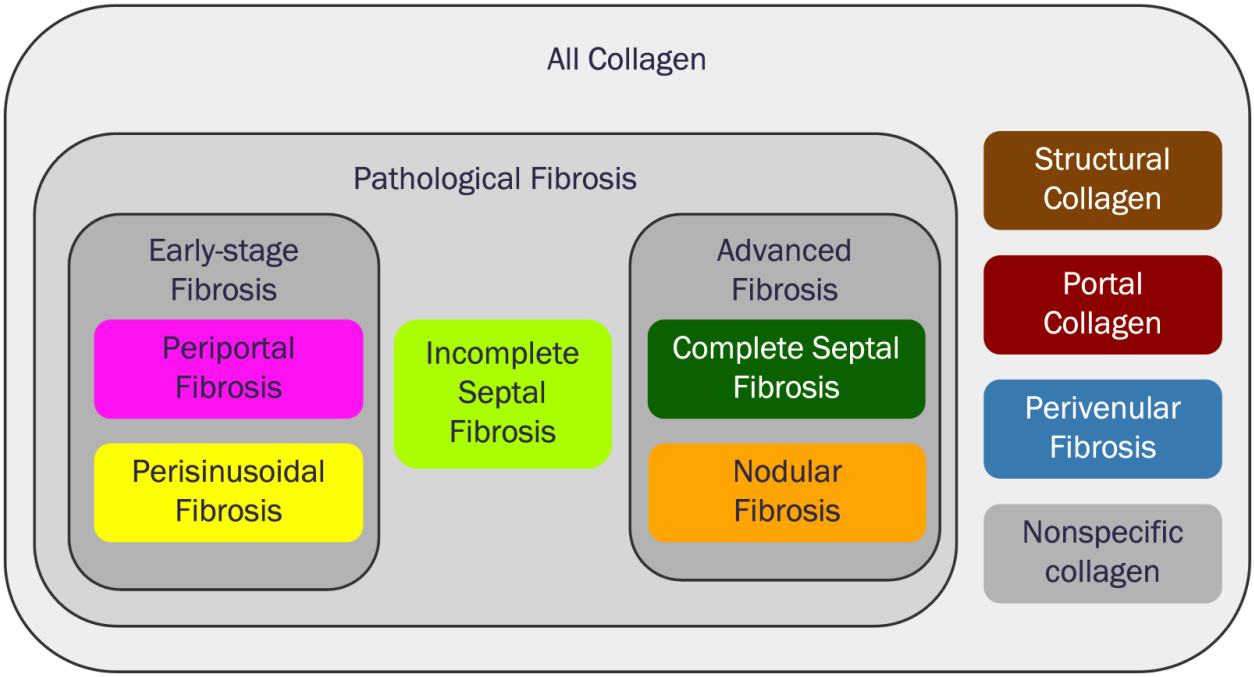
Supplementary Figure S2. Liver Explore Tissue Model Ontology



Supplementary Figure S3. Liver Explore Cell Model Ontology



Supplementary Figure S4. Liver Explore Zonal Model Ontology



Supplementary Figure S5. Liver Explore Fibrosis Model Ontology