Biomarker Qualification Letter of Intent (LOI)

ADMINISTRATIVE INFORMATION

1. **SUBMISSION TITLE:** ***"Machine-Learning based Model for NAFLD Activity Score (NAS) and Fibrosis Scores as a Histologic-based Surrogate Endpoint in Drug Development for NASH Clinical Trials"***

2. **REQUESTING ORGANIZATION:**

PathAI
https://www.pathai.com/

Requestor Organization Address: 120 Brookline Ave., Boston, MA 02215
Phone: 1-617-500-8457
Email: N/A

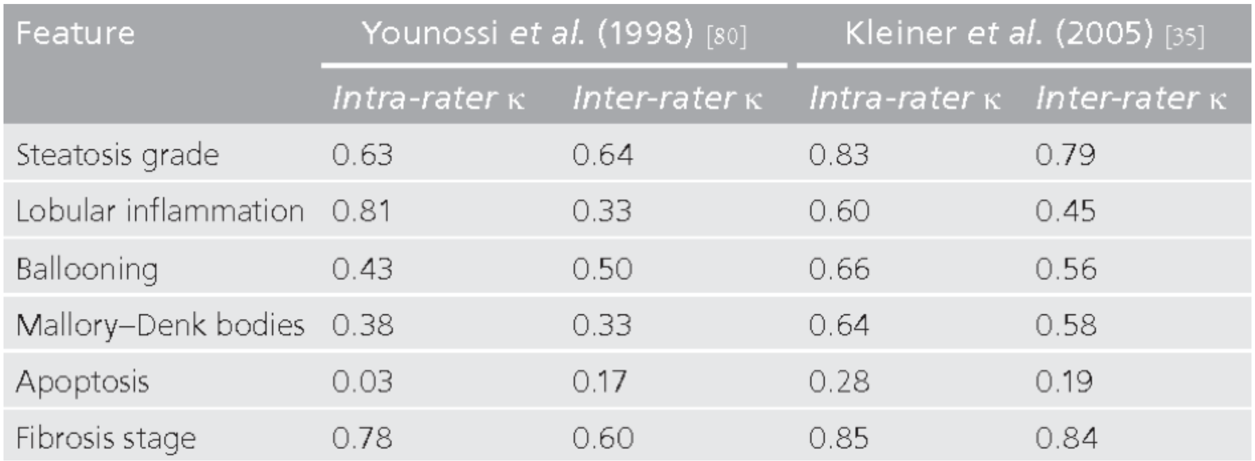
Primary Contact (i.e. POC that receives FDA communications):
Name: Esther Abels
Address: 120 Brookline Ave., Boston, MA 02215
Phone: 857-544-2249
Email: esther.abels@pathai.com

Alternative Contact:
Name: Ginny Kwan
Address: 120 Brookline Ave., Boston, MA 02215
Phone: 617-823-5073
Email: ginny.kwan@pathai.com

3. **SUBMISSION DATE: May 7, 2020**

DRUG DEVELOPMENT NEED STATEMENT

Currently, there are no approved therapies for nonalcoholic steatohepatitis (NASH). Due to its increasing prevalence, NASH is expected to soon become the leading cause of liver transplant [1,2,3]. There is a major and unmet clinical need to accelerate the development of effective therapeutics that result in significant clinical benefit. Pathology plays a critical role in NASH clinical trials with histology being the current reference method to determine change in disease activity, yet manual histological review is complex, subjective and prone to inter- and intra-reader variability and error. Various sets of semi-quantitative pathologic criteria have been proposed for scoring NASH [4,5]. Existing scoring systems show only moderate to fair reproducibility (both intra- and inter-pathologist scoring, Table 1), limiting their utility for clinical research and practice [4,6,7]. In addition to sampling error, the current limitation of manual Pathologist review in histologic endpoint measurement is the intra-rater variability of 15 – 50% (Table 1). This is of particular concern because drug efficacy is likely expected in the same range. In addition, NASH cirrhosis is characterized by heterogeneity in histology, presentation, and prognosis. Thus, there is a critical need for scalable, reproducible and validated tools in quantitative pathology for the assessment of treatment efficacy in NASH for clinical research. Identification and validation of such biomarker tools could significantly accelerate drug development, especially breakthrough therapy [8].



Feature	Younossi *et al.* (1998) [80]		Kleiner *et al.* (2005) [35]	
	Intra-rater κ	*Inter-rater κ*	*Intra-rater κ*	*Inter-rater κ*
Steatosis grade	0.63	0.64	0.83	0.79
Lobular inflammation	0.81	0.33	0.60	0.45
Ballooning	0.43	0.50	0.66	0.56
Mallory–Denk bodies	0.38	0.33	0.64	0.58
Apoptosis	0.03	0.17	0.28	0.19
Fibrosis stage	0.78	0.60	0.85	0.84

κ: Cohen's kappa

Table 1: Intra- and Inter-pathologist agreement for NASH relevant scores

Over the past five years, machine learning (ML), including artificial intelligence (AI), based tools have shown enormous progress for image-based applications offering tremendous promise in medical specialties including ophthalmology, radiology and pathology. FDA has recognized this in various disease areas, having cleared over 30 510(k) premarket applications for computer assisted/ML-based scoring approaches for immunohistochemistry-based breast markers in pathology (using product codes, OEO, NOT and NQN to search FDA 510K database as of Mar 31, 2020 [9]). PathAI has recently shown that a ML approach could be utilized to train models to accurately and efficiently interpret NASH histology, illustrate the heterogeneity of NASH fibrosis and cirrhosis, and has potential for risk stratification [10]. In contrast to the existing histological systems, this approach is quantitative and highly reproducible, providing value for accelerated and traditional approval pathways and for validation of surrogate histological endpoints for NASH.

BIOMARKER INFORMATION AND INTERPRETATION

1. BIOMARKER NAME:

Biomarker Name:	AI-based measurement of histologic endpoints in NASH (AIMHEN)
Type of Biomarker:	Histologic based, Imaging modality, measurement based on machine learning
BEST Classification:	Surrogate endpoint

2. ANALYTICAL METHODS:

There are currently no validated surrogate endpoints for NASH. Given the unmet medical need, it is important to identify and gather evidence for candidate surrogate endpoints that will help to accelerate drug development for both non- cirrhotic and cirrhotic NASH patients. Accurate, precise, reproducible and easy to implement histologic-based serial measurements used during trials will help to evaluate whether patients are responding to and will likely benefit from a therapy when clinical outcome is determined at later time points.

NASH disease activity is assessed histologically in the clinical trial setting by the non-alcoholic fatty liver disease (NAFLD) Activity Score (NAS) and the presence of steatohepatitis [4,5,7,11,12]. In this scoring method (outlined in Table 2), the presence and extent of steatosis (0-3), hepatocellular ballooning (0-2), and inflammation (0-3) are determined, traditionally by a pathologist reviewing first histological hematoxylin & eosin (H&E), and the components summed to give the overall NAS score. Fibrosis stage, using the CRN scoring method (Table 3) is then assessed on trichrome stained tissue slides using a microscope.

Item	**Definition**	**Score**
Steatosis	< 5%	0
	5%-33%	1
	> 33%-66%	2
	> 66%	3
Lobular inflammation	No foci	0
	< 2 foci per 200 x field	1
	2-4 foci per 200 x field	2
	> 4 foci per 200 x field	3
Ballooning	None	0
	Few balloon cells	1
	Many cells / prominent ballooning	2

Table 2: NAFLD Activity Score Components Developed by NASH Clinical Research Network (CRN). Kleiner DE et al. *Hepatology* 41(6):1313-21, 2005

Fibrosis Stage (Evaluated separately from NAS)	
0	None
1A	Mild, zone 3 perisinusoidal
1B	Moderate, zone 3 perisinusoidal
1C	Periportal sinusoidal fibrosis without accompanying zone 3 fibrosis
2	Zone 3 perisinusoidal and portal/periportal
3	Bridging fibrosis
4	Cirrhosis

Table 3. CRN-developed Fibrosis Scoring System. Adapted from Kleiner DE et al. *Hepatology* 41(6):1313-21, 2005

In NASH clinical trials, a follow-up biopsy is collected, often at 52 weeks, and the above NAS and fibrosis scoring systems are again assessed with the objective of capturing change in disease state to assess whether or not endpoints have been met. For these trials, FDA has suggested to include the same or similar patient populations in early and late phase 2 drug development programs as those planned for phase 3 development programs. The FDA suggests an inclusion criteria for NASH trials in phase 3, which includes a NAS greater than or equal to 4 with at least 1 point each in inflammation and ballooning along with a NASH Clinical Research Network (CRN) fibrosis score greater than stage 1 fibrosis but less than stage 4 fibrosis for trials for non-cirrhotic NASH patients, or stage 4 fibrosis for NASH trials in patients with compensated cirrhosis, as defined in the DRAFT FDA Guidance for Industry for Noncirrhotic Nonalcoholic Steatohepatitis With Liver Fibrosis: Developing Drugs for Treatment [13].

While the progression of NASH to various stages of fibrosis is not fully understood, patients with stage 2-3 fibrosis are at higher risk for progression to cirrhosis within 10 years, as well as having an increased mortality risk [11,14]. Given the prognostic value of fibrosis, histologic endpoints are especially recommended in the above referenced FDA DRAFT Guidance [13]. The features contributing to the NAS and fibrosis scores will all be evaluated specifically, in order to determine whether any of the following endpoints from the draft FDA guidance have been met:

Resolution of steatohepatitis (absence or isolated, simple steatosis without steatohepatitis, NAS score 0-1 for inflammation, 0 for ballooning) histologically with no worsening of fibrosis (using NAFLD CRN fibrosis score)

-OR-

Improvement in liver fibrosis greater than or equal to one stage (decreased NASH CRN fibrosis score) and no worsening of NAS score (defined as no increase in NAS for ballooning, inflammation, or steatosis);

-OR-

Both resolution of steatohepatitis and improvement in fibrosis (as defined above).

Given the multiple components that go into the NAS and fibrosis scores and the evidence of both intra- and inter-pathologist variability in scoring, there is a need for a precise, efficient, reproducible method to measure patients' baseline and follow-up biopsy scores. We propose that an image-derived, machine-learning (ML) model determination of NAS and Fibrosis histologic change in scores will serve this need in a consistent and efficient manner, as a pathology imaging-based biomarker to measure histologic-based surrogate endpoints for patients in NASH clinical trials.

3. MEASUREMENTS UNITS AND LIMITS(S) of DETECTION:

A broad range of NAS and fibrosis component features will be included in the development and analytical/clinical testing of the model. In whole slide image (WSI) analyses, using the same (repeats) and different (representing each score within each component) WSIs, the ability to accurately and precisely detect change in NAS and fibrosis score (i.e., using precise, continuous fibrosis scoring at both baseline and follow-up time points) will be determined in an automated fashion, and any limits of detection identified.

4. BIOMARKER INTERPRETATION & UTILITY:

PathAI algorithm training, verification and validation processes require interaction with board-certified pathologists in the form of annotation collection and model evaluation, which is supported through the PathAI research platform. The PathAI research platform contains a slide viewer with data collection capabilities. Board certified pathologists provide annotations used in training through this platform. For example, for the NASH model development thus far, baseline and follow-up (48 week) biopsies from patients considered for STELLAR-3 (https://clinicaltrials.gov/ct2/show/NCT03053050) and STELLAR-4 (https://clinicaltrials.gov/ct2/show/NCT03053063) clinical trials were formalin-fixed, paraffin embedded, sectioned and stained (both H&E and Trichrome). Slides were scanned on a WSI acquisition device, i.e. a scanner. The WSIs used in the entire development pipeline are stored in a cloud storage bucket hosted on Amazon Web Services (AWS), whereas the pathologist annotations and ground truth evaluations are stored in a database. The slides used in model training are sequestered from the slides used in model verification and validation. WSI's in the cloud-based viewer are quality control (QC) checked by in-house pathologists before being sent out for annotation by a pool of qualified pathologists. Pathologists provide annotations to outline and score the various NASH and fibrosis features (central veins, portal triads to identify zones, as well as areas of inflammation, steatosis, ballooning using H&E slides, and fibrosis using Trichrome stained slides; more details in the 'Analytical Considerations' section below). The PathAI research platform has the capability to retrieve the annotations for training algorithms. The model output from training iterations are in the form of overlays and slide level scores, which are exported to cloud storage on AWS. The model outputs are verified against pathologist evaluations using statistical tools for standalone verification.

When interoperability is established between the research platform and the algorithm, integrated verification tests are executed ensuring the algorithm, its components as well as the components within the platform, are performing safely and effectively as intended. The NASH algorithm consists of a result sub system which will provide scores for inflammation, steatosis, ballooning as well as overall scores, based on the CRN- derived NAS scoring system (Figure 1), and can indicate both CRN-based and Ishak fibrosis scores as continuous values on the slide level. If a patient has biopsies from baseline and follow-up timepoints, the model will calculate the absolute change in activity score components and overall NAS and fibrosis scores. In addition, it will provide the qualitative outcome whether a patient has met one or more of the histologic endpoints (see ‘Analytical Methods’ section).

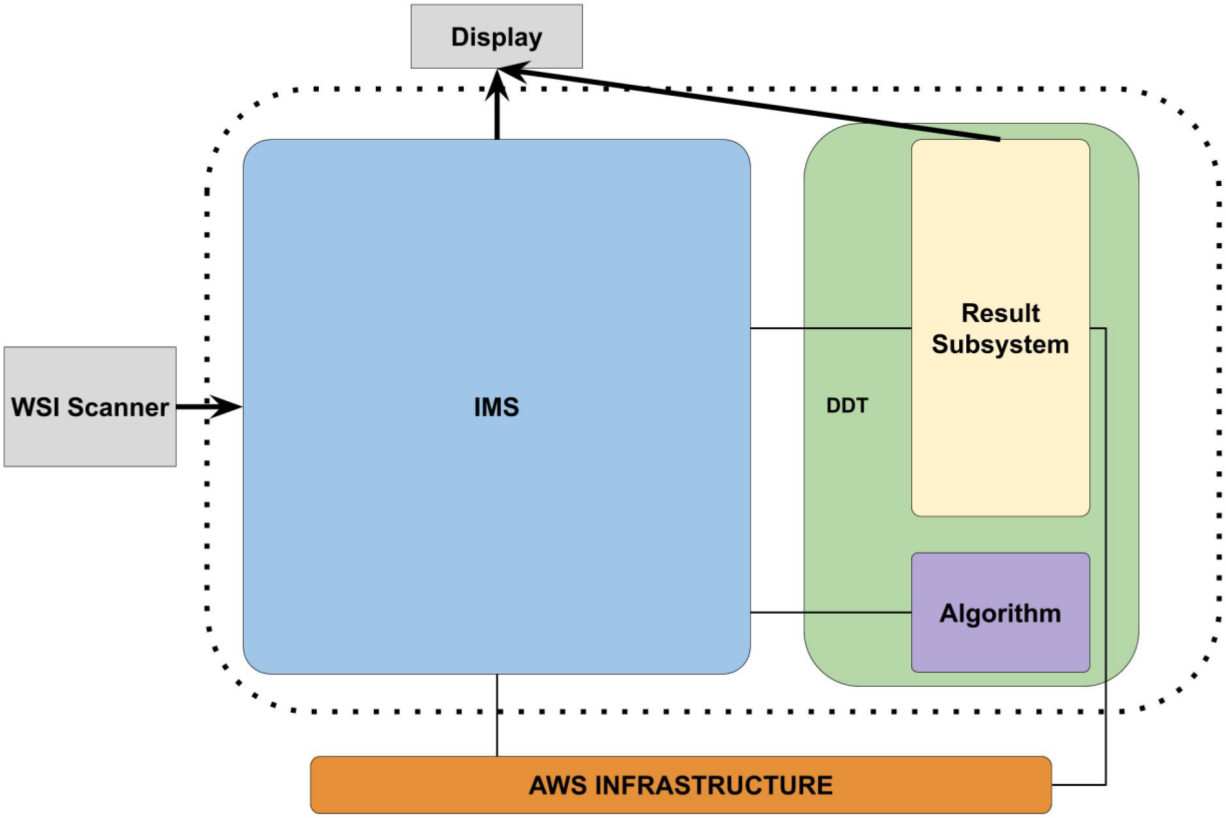


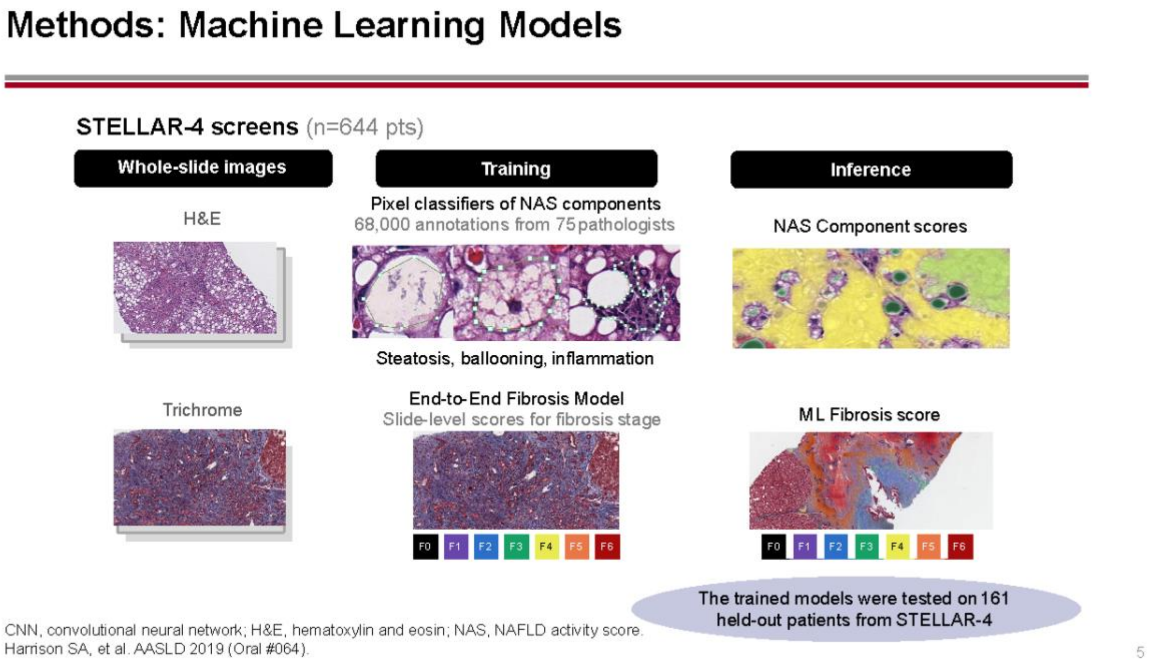
Figure 1. Schematic overview of the DDT, Image Management System (IMS) and Amazon Web Services (AWS) infrastructure

CONTEXT OF USE STATEMENT

A surrogate endpoint biomarker, based on Artificial Intelligence (AI), to measure treatment response based on histological change in Non-Alcoholic Fatty Liver Disease Activity Score (NAS) components (i.e., steatosis, ballooning, inflammation) and fibrosis scores in liver biopsies from baseline to follow-up in patients in clinical trials for treatment of non-alcoholic steatohepatitis (NASH).

ANALYTICAL CONSIDERATIONS

This model was previously trained on liver biopsies from NASH patients who underwent biopsies as part of randomized phase 3 trials of selonsertib (STELLAR- 3, STELLAR-4). Over 68,000 feature annotations (Fig 2) collected from 75 experienced liver pathologists on 642 hematoxylin and eosin (H&E) and 638 trichrome images were obtained and used to train, verify and validate (with a separate subset of cases) according to the predefined framework in Figure 3. Currently the model offers rigorous quantification and reveals heterogeneity for all major NASH pathological features (inflammation, steatosis, fibrosis, ballooning), in patients with and without compensated cirrhosis. (see ‘Supporting Information’ section).



Methods: Machine Learning Models

STELLAR-4 screens (n=644 pts)

Whole-slide images
Training
Inference
H&E
Pixel classifiers of NAS components
68,000 annotations from 75 pathologists
NAS Component scores
Steatosis, ballooning, inflammation
Trichrome
End-to-End Fibrosis Model
Slide-level scores for fibrosis stage
ML Fibrosis score
F0
F1
F2
F3
F4
F5
F6
F0
F1
F2
F3
F4
F5
F6
5

The trained models were tested on 161 held-out patients from STELLAR-4

CNN, convolutional neural network; H&E, hematoxylin and eosin; NAS, NAFLD activity score.
Harrison SA, et al. AASLD 2019 (Oral #064).

Figure 2: Initial model, developed and trained using subsets of STELLAR study patient slides; PathAI Oral Presentation, Pokkalla et. Al, AASLD 2019.

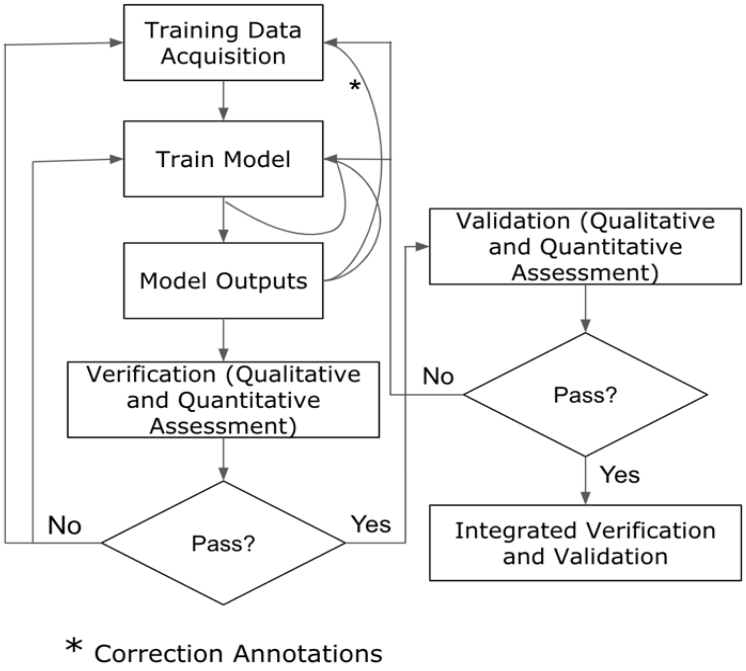

Training Data Acquisition
*
Train Model
Model Outputs
Validation (Qualitative and Quantitative Assessment)
Verification (Qualitative and Quantitative Assessment)
No
Pass?
Yes
No
Pass?
Yes
Integrated Verification and Validation
* Correction Annotations

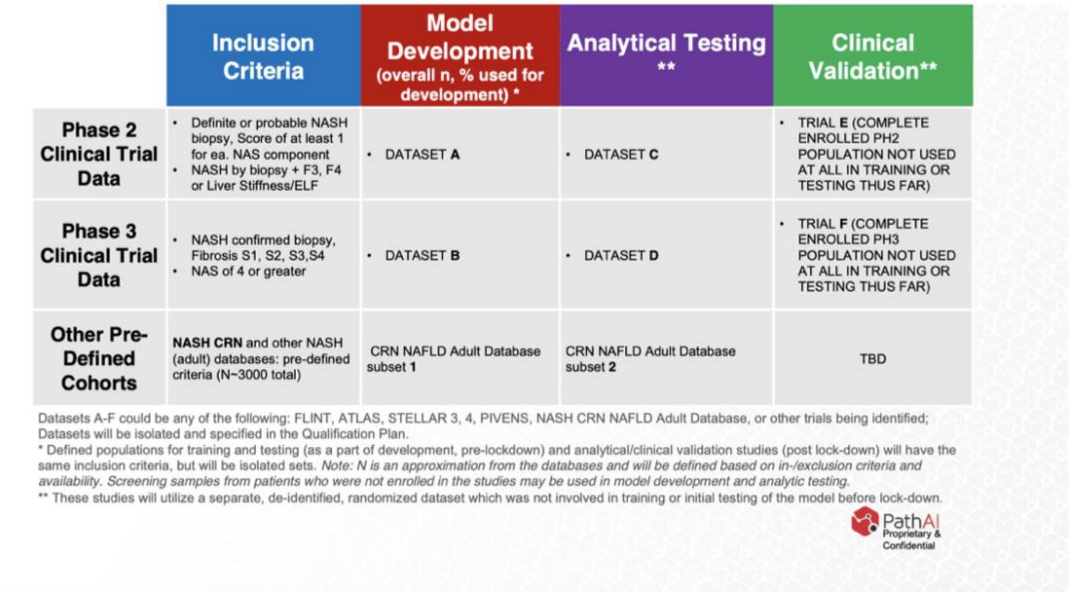
Figure 3: Iterative model training process

For qualification of this proposed DDT, the NASH model will be optimized by training and testing with separate subsets of cases from a large population taken from the NASH CRN database (including the PIVENS and FLINT studies, as well as patients from the adult NAFLD databases) and will include both placebo and treatment arms from studies with various treatment regiments (different targets, as well as combinations).

The additional optimization of the model, based on annotations from a collection of experienced liver pathologists, will be performed to further fine-tune the model using a pre-defined subset of WSIs. The images are acquired on a range of different scanners with a 40x magnification to address possible variances and develop a robust algorithm.

The overall case population for development and testing will reflect a broad spectrum of pre-analytics, such as different labs ranging from community labs to large reference labs with each using their own staining and scanning protocol, as well as the full range of NAS and fibrosis scores in order to include the entire range of disease presentation and histology methodology, as well as to increase the generalizability of the model. The accuracy of the model feature identification and output scores will be verified and validated using a separate held out dataset (not used in training) by expert pathologists before the device will be locked down. The optimized model will then undergo analytical and clinical validation studies using separate sets of samples from phase 2 and 3 trials with entire, held out clinical trial populations (not used in training or testing thus far) for clinical validation. Concordance of the locked-down model will be determined with the NAS, NASH CRN and Ishak fibrosis staging systems as interpreted by the liver pathologists (consensus score) of the studies. The model will then be tested for repeatability, reproducibility and precision, using WSIs that represent a broad range of scores for each of the NAS and fibrosis histologic features. See Table 4 below for an impression of the inclusion criteria, as well as division of data sets across model training, verification and validation. The Statistical Analysis Plan, including the statistical method and analysis as well as sample size calculations, for these analytical and clinical validation studies will be developed and finalized prior to the start of these studies and will be described in the Qualification Plan.

Table 4. Inclusion Criteria and Training, Testing Datasets



	Inclusion Criteria	Model Development (overall n, % used for development) *	Analytical Testing **	Clinical Validation**
Phase 2 Clinical Trial Data	• Definite or probable NASH biopsy, Score of at least 1 for ea. NAS component • NASH by biopsy + F3, F4 or Liver Stiffness/ELF	• DATASET **A**	• DATASET **C**	• TRIAL **E** (COMPLETE ENROLLED PH2 POPULATION NOT USED AT ALL IN TRAINING OR TESTING THUS FAR)
Phase 3 Clinical Trial Data	• NASH confirmed biopsy, Fibrosis S1, S2, S3,S4 • NAS of 4 or greater	• DATASET **B**	• DATASET **D**	• TRIAL **F** (COMPLETE ENROLLED PH3 POPULATION NOT USED AT ALL IN TRAINING OR TESTING THUS FAR)
Other Pre-Defined Cohorts	**NASH CRN** and other NASH (adult) databases: pre-defined criteria (N~3000 total)	CRN NAFLD Adult Database subset **1**	CRN NAFLD Adult Database subset **2**	TBD

Datasets A-F could be any of the following: FLINT, ATLAS, STELLAR 3, 4, PIVENS, NASH CRN NAFLD Adult Database, or other trials being identified; Datasets will be isolated and specified in the Qualification Plan.

* Defined populations for training and testing (as a part of development, pre-lockdown) and analytical/clinical validation studies (post lock-down) will have the same inclusion criteria, but will be isolated sets. *Note: N is an approximation from the databases and will be defined based on in-/exclusion criteria and availability. Screening samples from patients who were not enrolled in the studies may be used in model development and analytic testing.*

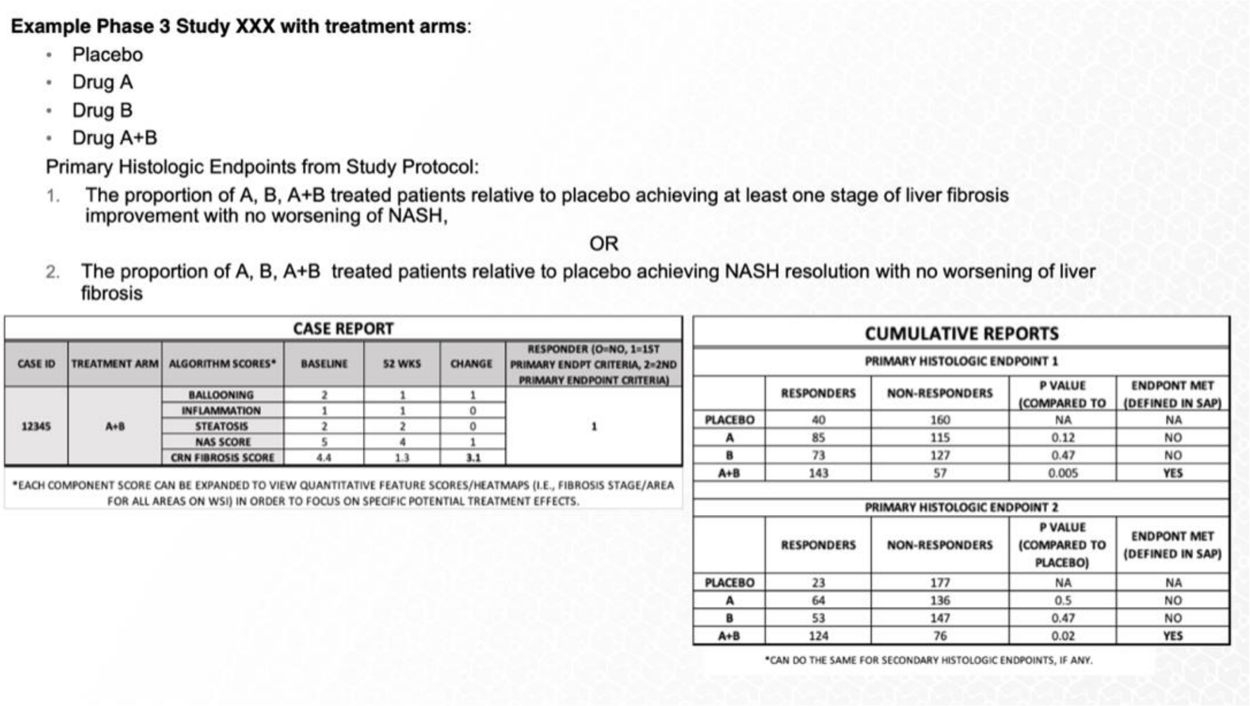
** These studies will utilize a separate, de-identified, randomized dataset which was not involved in training or initial testing of the model before lock-down.

PathAI Proprietary & Confidential

CLINICAL CONSIDERATIONS

Use Statement

As defined in the COU section of this LOI, the proposed AI-based model for precise histologic scoring of NASH biopsy slides can be used as a surrogate endpoint for patients in clinical trials for NASH, based on DRAFT FDA guidelines for accepted histologic endpoints. Both baseline and follow-up biopsy slides must be measured using the model, in order to consistently and accurately capture changes in NAS component scores and fibrosis stage to evaluate the effectiveness of the drug (see Figure 4 for example reports and description for details).



Example Phase 3 Study XXX with treatment arms:

- Placebo
- Drug A
- Drug B
- Drug A+B

Primary Histologic Endpoints from Study Protocol:

1. The proportion of A, B, A+B treated patients relative to placebo achieving at least one stage of liver fibrosis improvement with no worsening of NASH,

OR

2. The proportion of A, B, A+B treated patients relative to placebo achieving NASH resolution with no worsening of liver fibrosis

CASE REPORT

CASE ID	TREATMENT ARM	ALGORITHM SCORES*	BASELINE	52 WKS	CHANGE	RESPONDER (O=NO, 1=1ST PRIMARY ENDPT CRITERIA, 2=2ND PRIMARY ENDPOINT CRITERIA)
12345	A+B	BALLOONING	2	1	1	1
		INFLAMMATION	1	1	0	
		STEATOSIS	2	2	0	
		NAS SCORE	5	4	1	
		CRN FIBROSIS SCORE	4.4	1.3	**3.1**	

*EACH COMPONENT SCORE CAN BE EXPANDED TO VIEW QUANTITATIVE FEATURE SCORES/HEATMAPS (I.E., FIBROSIS STAGE/AREA FOR ALL AREAS ON WSI) IN ORDER TO FOCUS ON SPECIFIC POTENTIAL TREATMENT EFFECTS.

CUMULATIVE REPORTS

PRIMARY HISTOLOGIC ENDPOINT 1

	RESPONDERS	NON-RESPONDERS	P VALUE (COMPARED TO	ENDPONT MET (DEFINED IN SAP)
PLACEBO	40	160	NA	NA
A	85	115	0.12	NO
B	73	127	0.47	NO
A+B	143	57	0.005	YES

PRIMARY HISTOLOGIC ENDPOINT 2

	RESPONDERS	NON-RESPONDERS	P VALUE (COMPARED TO PLACEBO)	ENDPONT MET (DEFINED IN SAP)
PLACEBO	23	177	NA	NA
A	64	136	0.5	NO
B	53	147	0.47	NO
A+B	124	76	0.02	YES

*CAN DO THE SAME FOR SECONDARY HISTOLOGIC ENDPOINTS, IF ANY.

Figure 4: Example Case Level and Cumulative Study Reports

Proposed Conditions of Qualified Use

Population for use:

- Adults, 18 yrs and older.
- Patients enrolled in a NASH clinical trial, with a NASH confirmed biopsy or biochemical criteria and/or imaging evidence of steatosis/steatohepatitis/fibrosis in addition to known risk factors for NASH.

Whole Slide Image Considerations for Clinical Trial Use:

- Formalin-fixed, paraffin-embedded (FFPE) liver biopsy tissue should be stained with H&E and Trichrome according to the package insert.
- H&E and Trichrome-stained slides should be scanned by CRO on validated scanner(s) and the WSIs should be quality checked according to their instructions for use.

General Clinical Validation Plan

As mentioned at the end of the 'Analytical Considerations' section, the optimized model will be validated on the WSI level by comparing model output scores (NAS, Fibrosis) to consensus scores generated by a panel of expert board certified, liver pathologists in concordance studies using Cohen's kappa (on the level of component and overall NAS scores). The repeatability, reproducibility and precision (ability to differentiate between each component score level) will be evaluated on the whole slide image level as well, including a broad range of scores and change in scores.

One of the possible benefits of using this machine-learning derived NASH tool is to be able to more accurately and consistently capture and quantify histologic change in treatment vs. placebo groups in NASH clinical trials. To support and evaluate this benefit, we plan to use retrospective data from the histologic endpoints in ATLAS and possibly PIVENS and FLINT studies (CRN database), as well as an additional completed Phase 3 dataset that was not used in any training or precision studies (see Table 4). In these studies, NAS and fibrosis scores will be generated using manual pathologist reads and using ML models, and the change in NAS and fibrosis scores from baseline to follow-up biopsies will be computed for the manual scores and for the ML-derived scores. Statistical analyses of endpoint treatment group comparisons generated with manual scoring vs. ML model scoring will be performed to determine whether model-based endpoint agreement is equal to or greater than an individual pathologist and pathologist consensus, and therefore offers more statistical power to capture true change through being a more precise and reproducible method.

Overview of Risks and Benefits:

Machine learning algorithms can be understood by defining their internal structure and having clear knowledge about the framework of inputs and the relation to their outputs. The risks of possible over- and underfitting of algorithms need to be acknowledged and balanced with the advantages that well-designed algorithms can provide. PathAI mitigates this risk by using a training, validation, and testing framework and outcome measures (including clinical outcome data).

From a clinical study standpoint, potential benefits of the PathAI device compared to manual scoring are 1) more precise and accurate scoring 2) due to use of an objective and consistent approach the improved ability to identify true change as a result of a therapeutic intervention and 3) fully automated quantitative analysis without human factor errors involved.

Risks of the device are associated with failure of the device to perform as expected, leading to incorrect test results. PathAI will minimize the risk from incorrect results by performing tests, to be defined in the Qualification Plan (QP), to optimize and validate the device.

SUPPORTING INFORMATION

Preliminary Study Results:

The following summaries outline model development and analyses that have been performed thus far and presented at the 2019 AASLD Meeting, which show proof-of-concept and value of use of the model as a qualified DDT in NASH clinical trials compared to the current standard:

Abstract #187: Oral presentation at the 2019 AASLD meeting

Title: MACHINE LEARNING MODELS ACCURATELY INTERPRET LIVER HISTOLOGY IN PATIENTS WITH NONALCOHOLIC STEATOHEPATITIS (NASH)

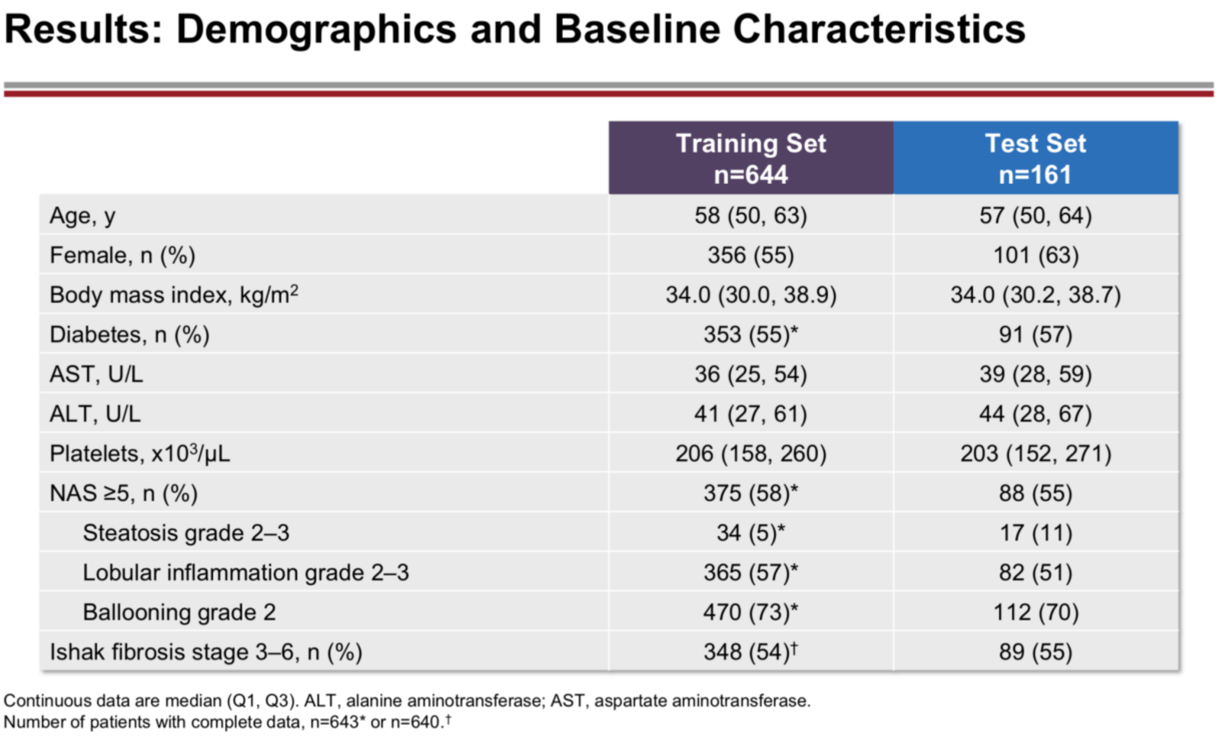
Authors: Harsha Pokkalla, Kishalve Pethia, Benjamin Glass, Jennifer K Kerner, Yevgeniy Gindin, Ling Han, Ryan Huss, Chuhan Chung, Stephen Djedjos, Mani Subramanian, Robert P Myers, Murray Resnick3, Stephen A Harrison4, Zachary D Goodman, Aditya Khosla, Andrew Beck, Ilan Wapinski, and Zobair M. Younossi

Background: Variability in pathological assessment of liver histology in patients with NASH may hinder advances in understanding NASH pathogenesis and confound the results of clinical trials. We hypothesized that a machine learning approach (PathAI) could be utilized to train models to accurately interpret NASH histology.

Methods: Images from 834 liver biopsies from subjects screened for a phase 3 trial of selonsertib (STELLAR-4) were scored by a central pathologist (CP) according to the NAFLD Activity Score (NAS) and NASH CRN and Ishak fibrosis staging systems. The PathAI research platform (PathAI; Boston, MA) was applied to train a convolutional neural network (CNN) with over 20 layers and 8 million parameters using over 68,000 annotations collected from 75 board-certified pathologists on 642 hematoxylin and eosin (H&E) and 638 trichrome images. Annotations included normal liver, steatosis, hepatocellular ballooning, lobular inflammation, portal inflammation, and bile ducts The models were then applied to a separate test set of images (165 H&E, 165 trichrome) to evaluate correlations (Spearman, rs) with consensus mean readings from two independent liver pathologists not included in model training (P1 and P2). For the staging of fibrosis, CNN models were trained using slide-level pathologist scores to recognize unique patterns associated with each fibrosis stage within fibrotic regions of trichrome images. These region-based scores were summarized for each slide and evaluated on nonannotated trichrome images.

Results: There was moderate to strong correlation between the CP and two independent liver pathologists for NAS features including steatosis (CP vs P1, rs=0.70; CP vs P2, rs=0.68; P1 vs P2, rs=0.72), ballooning (CP vs P1, rs=0.73; CP vs P2, rs=0.64; P1 vs P2, rs=0.74), and lobular inflammation (CP vs P1, rs=0.60; CP vs P2, rs=0.58; P1 vs P2, rs=0.56). Correlations between the machine learning model and the consensus mean of readings from P1 and P2 were similar for ballooning (rs=0.68) and lobular inflammation (rs=0.56), but the model demonstrated superior correlation for steatosis (rs=0.86). For the staging of fibrosis, the model predictions in the test set were highly correlated with readings of the CP for both the NASH CRN (rs=0.83) and Ishak staging systems (rs=0.86).

Conclusion: Machine learning models demonstrated high concordance with pathologist interpretations of the histological features of NASH These data highlight the potential of machine learning models for interpretation of NASH histology in clinical trials, and suggest the benefits of generating automated and quantitative readouts for staging and characterizing liver disease.

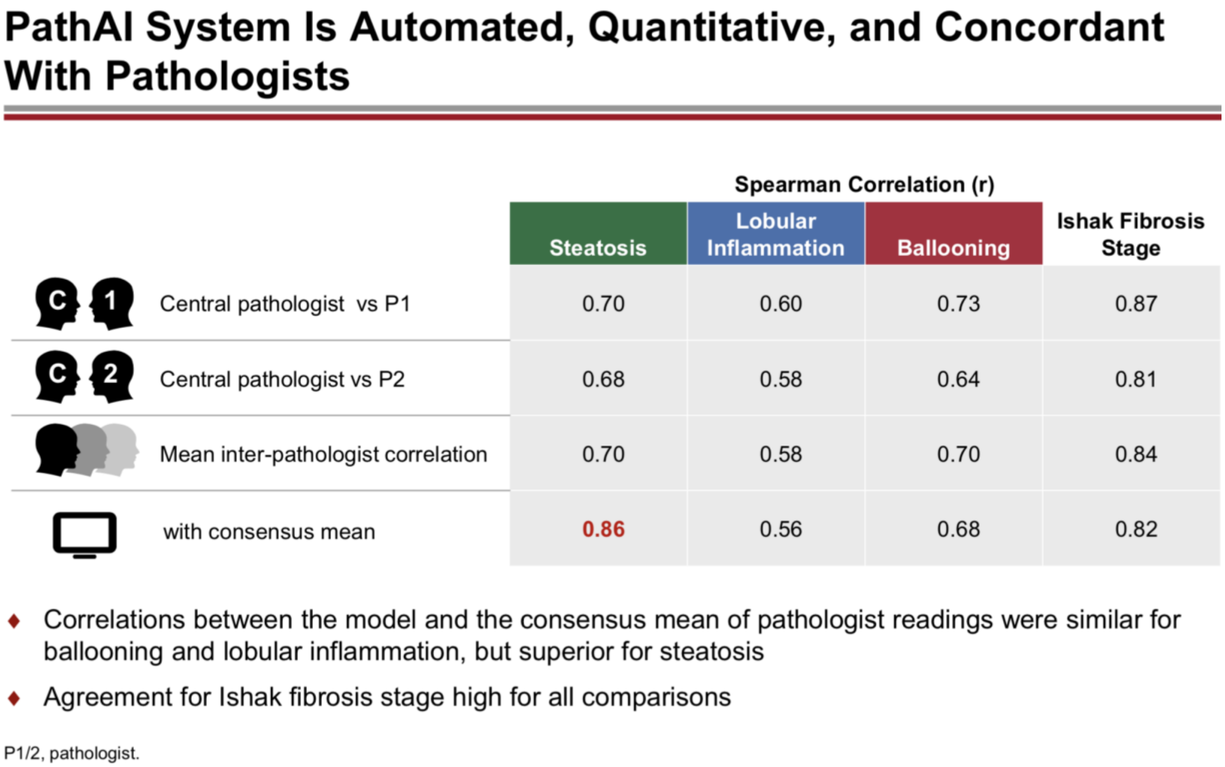


Results: Demographics and Baseline Characteristics

	Training Set n=644	Test Set n=161
Age, y	58 (50, 63)	57 (50, 64)
Female, n (%)	356 (55)	101 (63)
Body mass index, kg/m^2	34.0 (30.0, 38.9)	34.0 (30.2, 38.7)
Diabetes, n (%)	353 (55)*	91 (57)
AST, U/L	36 (25, 54)	39 (28, 59)
ALT, U/L	41 (27, 61)	44 (28, 67)
Platelets, x10^3/µL	206 (158, 260)	203 (152, 271)
NAS ≥5, n (%)	375 (58)*	88 (55)
Steatosis grade 2–3	34 (5)*	17 (11)
Lobular inflammation grade 2–3	365 (57)*	82 (51)
Ballooning grade 2	470 (73)*	112 (70)
Ishak fibrosis stage 3–6, n (%)	348 (54)†	89 (55)

Continuous data are median (Q1, Q3). ALT, alanine aminotransferase; AST, aspartate aminotransferase.
Number of patients with complete data, n=643* or n=640.†

Figure 5: Patient Demographics for Initial Model Development and Testing using STELLAR patient population



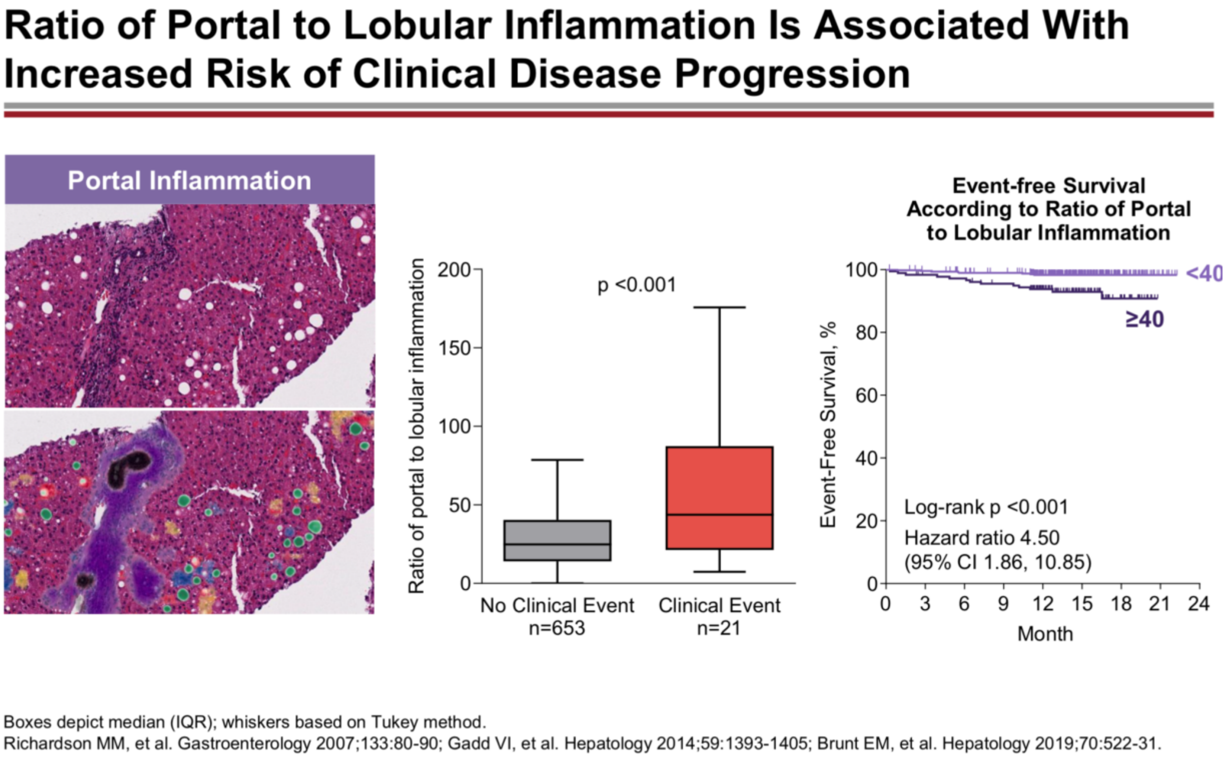
PathAI System Is Automated, Quantitative, and Concordant With Pathologists

	Spearman Correlation (r)			
	Steatosis	Lobular Inflammation	Ballooning	Ishak Fibrosis Stage
C 1 Central pathologist vs P1	0.70	0.60	0.73	0.87
C 2 Central pathologist vs P2	0.68	0.58	0.64	0.81
Mean inter-pathologist correlation	0.70	0.58	0.70	0.84
with consensus mean	**0.86**	0.56	0.68	0.82

- Correlations between the model and the consensus mean of pathologist readings were similar for ballooning and lobular inflammation, but superior for steatosis
- Agreement for Ishak fibrosis stage high for all comparisons

P1/2, pathologist.

Figure 6: Concordance data for Initial model developed with Stellar populations



Ratio of Portal to Lobular Inflammation Is Associated With Increased Risk of Clinical Disease Progression

Boxes depict median (IQR); whiskers based on Tukey method.
Richardson MM, et al. Gastroenterology 2007;133:80-90; Gadd VI, et al. Hepatology 2014;59:1393-1405; Brunt EM, et al. Hepatology 2019;70:522-31.

Figure 7: Portal to Lobular Inflammation Associated With Disease Progression

Abstract #1718: Poster presentation at the 2019 AASLD Meeting

Title: MACHINE LEARNING FIBROSIS MODELS BASED ON LIVER HISTOLOGY IMAGES ACCURATELY CHARACTERIZE THE HETEROGENEITY OF CIRRHOSIS DUE TO NONALCOHOLIC STEATOHEPATITIS (NASH)

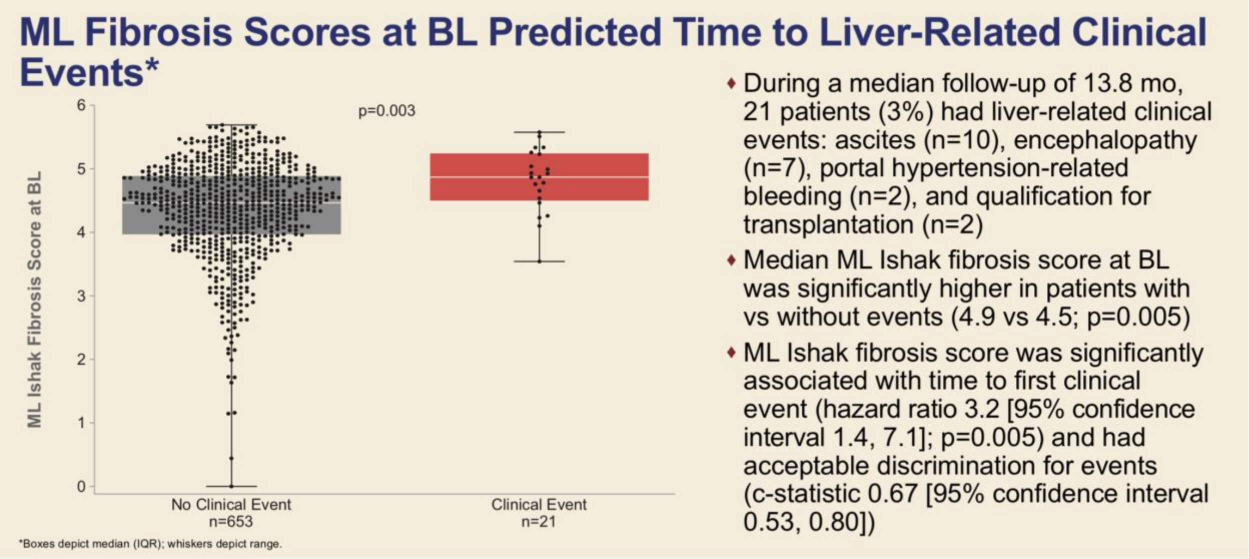
Authors: Zobair M. Younossi, Harsha Pokkalla, Kishalve Pethia, Benjamin Glass, Jennifer Kaplan Kerner, Yevgeniy Gindin, Ling Han, Ryan Huss, Chuhan Chung, Stephen Djedjos, Mani Subramanian, Robert P Myers, Aditya Khosla, Murray Resnick, Stephen A Harrison, Quentin M. Anstee, Vincent Wai-Sun Wong, Ilan Wapinski, Andrew Beck and Zachary D Goodman

Background: NASH cirrhosis is characterized by heterogeneity in histology, presentation, and prognosis. We hypothesized that a machine learning (ML) approach trained on liver histological images would illustrate the heterogeneity of NASH cirrhosis and demonstrate potential for risk stratification.

Methods: Patients with compensated cirrhosis due to NASH were enrolled in a randomized trial of selonsertib (STELLAR-4). Liver fibrosis at baseline (BL) was staged by a central pathologist according to the Ishak and NASH CRN classifications, hepatic collagen and α-SMA were quantified by morphometry, liver stiffness was measured by transient elastography (TE), and ELF and NAFLD Fibrosis Score (NFS) were calculated ML models based on convolutional neural networks with >20 layers and 8 million parameters (PathAI research platform; Boston, MA) were trained to recognize patterns within fibrotic regions associated with each fibrosis stage using slide-level pathologist Ishak fibrosis stages. Once trained, models were applied to another set of trichrome images of BL biopsies, associating each pixel to an Ishak stage Associations were aggregated across the entire image, producing patient-level scores by averaging the scores across all pixels, and associations between these ML Ishak fibrosis scores, clinical parameters, and adjudicated clinical events (e g decompensation, transplantation, death) were determined.

Results: 674 patients with NASH cirrhosis and evaluable images were included. 62% had Ishak F6 fibrosis, and median (IQR) hepatic collagen and ELF were 10.7% (7.6, 14.7%) and 10.62 (10.02, 11.31), respectively. Within this cirrhotic population, the aggregated ML Ishak fibrosis scores were widely distributed (median 4.5; IQR 4.0, 4.9). While a median (IQR) of 20% (7-39%) of total pixels were consistent with Ishak F6 and 34% (24-45%) with F5 based on the ML models, 14% (8-23%), 8% (5-13%), 4% (2-7%), 4% (2-9%), and 0 3% (0 1-1 1%) of pixels were consistent with F4, F3, F2, F1, and F0, respectively. The ML Ishak fibrosis score at BL correlated with hepatic collagen (Spearman ρ=0.36), α-SMA (ρ=0.31), TE (ρ=0.30), ELF (ρ=0.27), NFS (ρ=0.27), and platelets (ρ=-0.25; all p<0.0001). During a median follow-up of 13.8 months, 21 patients (3.1%) had clinical events The median ML Ishak fibrosis score at BL was significantly higher in patients with vs without events (4.9 vs 4.5; HR 3.2 [95% CI 1.4, 7.1], p=0.005) and the score had acceptable discrimination for future events (c-statistic, 0.67; 95% CI 0.53, 0.80)

Conclusion: ML models illustrate the heterogeneity of fibrosis in NASH cirrhosis, correlate with noninvasive markers of fibrosis, and are prognostic. These data highlight the potential of ML models to characterize cirrhotic patients above and beyond conventional histological staging in an automated fashion.



ML Fibrosis Scores at BL Predicted Time to Liver-Related Clinical Events*

*Boxes depict median (IQR); whiskers depict range.

- During a median follow-up of 13.8 mo, 21 patients (3%) had liver-related clinical events: ascites (n=10), encephalopathy (n=7), portal hypertension-related bleeding (n=2), and qualification for transplantation (n=2)
- Median ML Ishak fibrosis score at BL was significantly higher in patients with vs without events (4.9 vs 4.5; p=0.005)
- ML Ishak fibrosis score was significantly associated with time to first clinical event (hazard ratio 3.2 [95% confidence interval 1.4, 7.1]; p=0.005) and had acceptable discrimination for events (c-statistic 0.67 [95% confidence interval 0.53, 0.80])

Figure 8: Fibrosis Scores can predict time to clinical event

PREVIOUS QUALIFICATION INTERACTIONS AND OTHER APPROVALS

This is the first regulatory interaction for this proposed biomarker.

ATTACHMENTS

TABLES AND FIGURES

- **Table 1**: Intra- and Inter-pathologist agreement for NASH relevant scores
- **Table 2**: NAFLD Activity Score Components Developed by NASH Clinical Research Network (CRN). Kleiner DE et al. Hepatology 41(6):1313-21, 200
- **Table 3:** CRN-developed Fibrosis Scoring System. Adapted from Kleiner DE et al. Hepatology 41(6):1313-21, 2005
- **Table 4**: Inclusion Criteria and Training, Testing Datasets
- **Figure 1**: Schematic overview of the DDT, Image Management System (IMS) and Amazon Web Services (AWS) infrastructure
- **Figure 2**: Initial model, developed and trained using subsets of STELLAR study patient slides; PathAI Oral Presentation, Pokkalla et. Al, AASLD 2019.
- **Figure 3**: Iterative model training process
- **Figure 4**: Example Case Level and Cumulative Study Reports
- **Figure 5**: Patient Demographics for Initial Model Development and Testing using STELLAR patient population

- **Figure 6**: Concordance data for Initial model developed with Stellar populations Figure 7: Portal to Lobular Inflammation Associated With Disease Progression Figure 8: Fibrosis Scores can predict time to clinical event

REFERENCES

1. Friedman SL, Neuschwander-Tetri BA, Rinella M, Sanyal. Mechanisms of NAFLD development and therapeutic strategies.Nat Med. 2018 Jul;24(7):908-922. doi: 10.1038/s41591-018-0104-9. Epub 2018 Jul 2.
2. Chalasani N, Younossi Z, Lavine JE, Charlton M, Cusi K, Rinella M, Harrison SA, Brunt EM, Sanyal AJ.The diagnosis and management of nonalcoholic fatty liver disease: Practice guidance from the American Association for the Study of Liver Diseases.Hepatology. 2018 Jan;67(1):328-357.
3. Loomba R and Sanyal AJ. The Global NAFLD Epidemic. Nat Rev Gastroenterol Hepatol 10:686–90, 2013.
4. Claudia Filozof, Shein-Chung Chow, Lara Dimick-Santos, Yeh-Fong Chen, Richard N. Williams, Barry J. Goldstein, and Arun Sanyal. Clinical endpoints and adaptive clinical trials in precirrhotic nonalcoholic steatohepatitis: Facilitating development approaches for an emerging epidemic. Hepatol Commun. 2017 Sep; 1(7): 577–585.
5. Sanyal A, Brunt E, Kleiner D, Kowdley KV, Chalasani N, Lavine JE, et al. Endpoints and clinical trial design for nonalcoholic steatohepatitis. Hepatology 2011;54:344-353.
6. Younossi ZM, Loomba R, Anstee QM, Rinella ME, Bugianesi E, Marchesini G, Neuschwander-Tetri BA, Serfaty L, Negro F, Caldwell SH, Ratziu V, Corey KE, Friedman SL, Abdelmalek MF, Harrison SA, Sanyal AJ, Lavine JE, Mathurin P, Charlton MR, Goodman ZD, Chalasani NP, Kowdley KV, George J, Lindor K. Diagnostic modalities for nonalcoholic fatty liver disease, nonalcoholic steatohepatitis, and associated fibrosis. Hepatology. 2018 Jul;68(1):349-360.
7. Pavlides M, Birks J, Fryer E, Delaney D, Sarania N, Banerjee R, Neubauer S, Barnes E, Fleming KA, Wang LM. Interobserver Variability in Histologic Evaluation of Liver Fibrosis Using Categorical and Quantitative Scores. Am J Clin Pathol. 2017 Apr 1;147(4):364- 369.
8. https://www.fda.gov/regulatory-information/food-and-drug-administration-safety-and-innovation-act-fdasia/frequently-asked-questions-breakthrough-therapies.
9. https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfPMN/pmn.cfm.
10. Harsha Pokkalla, Kishalve Pethia, Benjamin Glass, Jennifer K. Kerner,Yevgeniy Gindin, Ling Han, Ryan Huss, Chuhan Chung, C. Stephen Djedjos, G. Mani Subramanian, Robert P. Myers, Murray Resnick, Stephen A. Harrison, Zachary D. Goodman, Aditya Khosla, Andrew H. Beck, Ilan Wapinski, Zobair M. Younossi. Machine Learning Models. Accurately Interpret Liver Histology in Patients With Nonalcoholic Steatohepatitis. Presented at AASLD, Boston 2019.
11. Sanyal AJ, Friedman SL, McCullough AJ, Dimick-Santos L; American

Association for the Study of Liver Diseases; United States Food and Drug Administration. Challenges and opportunities in drug and biomarker development for nonalcoholic steatohepatitis: findings and recommendations from an American Association for the Study of Liver Diseases-U.S. Food and Drug Administration joint workshop. Hepatology 2015;61:1392- 1405.

12. Kleiner DE, Brunt EM, Van Natta M, Behling C, Contos MJ, Cummings OW, Ferrell LD, Liu YC, Torbenson MS, Unalp-Arida A, Yeh M, McCullough AJ, Sanyal AJ; Nonalcoholic Steatohepatitis Clinical Research Network.Design and validation of a histological scoring system for nonalcoholic fatty liver disease. Hepatology. 2005 Jun;41(6):1313-21.
13. https://www.fda.gov/media/119044/download
14. Angulo P, Kleiner DE, Dam-Larsen S, Adams LA, Bjornsson ES, Charatcharoenwitthaya P, et al. Liver fibrosis, but no other histologic features, is associated with long-term outcomes of patients with nonalcoholic fatty liver disease. Gastroenterology 2015;149:389-397.e10