LIVER FORUM HISTOLOGY SERIES

Session 2: Innovations in Liver Biopsy Assessment- Utilizing Digital Methods and AI Technology
Webinar Summary
October 27, 2020

1608 Rhode Island Avenue NW
Washington, DC 20036
www.forumresearch.org

THE FORUM
For Collaborative Research℠

Berkeley Public Health

INNOVATIONS: UTILIZING DIGITAL METHODS AND NEW TECHNOLOGY

Innovative Tools for Quantitative Analysis of NAFLD Histology

Presenter: Samer Gawrieh, Indiana University

Slides: https://bit.ly/3mCyrf1

Why is innovation needed?
- The accuracy and reproducibility of diagnostic classification of NAFLD phenotypes is a challenge for the field
    - This is in part due to the composite phenotype relying heavily on accurate identification of steatosis, lobular inflammation, and ballooning. Both lobular inflammation and ballooning are challenging to accurately identify and quantify.
    - Another challenging aspect is the lack of continuous scales for biopsy assessment.
- If biopsy based NAFLD phenotyping continues to be used, there must be a clear understanding of the factors that affect sample quality and diagnostic yield. These factors include the length/size, same processing, width/core, number of cores, lobe, plane, and sample analysis.
    - Of these, digital technology can only impact the biopsy sample analysis, by maximizing the data that can be obtained from the biopsy sample provided.

What is the current state of histological analysis?
- Histological analysis is a manual process where pathologists look assess samples through a microscope and provide semi-quantitative grades for each of the NAFLD lesions. Both the NASH CRN system and the FLIP system provide a limited range for the scores (0-4)
    - The limited grades that comprise the scoring system may not capture clinically meaningful or significant change
- Intra- and inter-observer agreement are highest for steatosis and fibrosis; however, intra- and inter-observer agreement is low for inflammation, ballooning, and the diagnostic classification of NAFLD phenotypes
    - These findings have been reported and reproduced by many groups[1,2,3,4,5,6]
    - In an effort to improve the intra- and inter-observer agreement, the implementation of two interventions were tested:
        - Education/training of the study pathologist using teaching slides containing classical examples of diagnostic criteria.

---

[1] Younossi ZM et al. Nonalcoholic fatty liver disease: assessment of variability in pathologic interpretations. Mod Pathol. 1998;11(6):560-5.

[2] Fukusato T et al. Interobserver variation in the histopathological assessment of nonalcoholic steatohepatitis. Hepatology Research. 2005;33(2):122-7.

[3] Kleiner DE et al. Design and validation of a histological scoring system for nonalcoholic fatty liver disease. Hepatology. 2005;41(6):1313-21.

[4] Juluri R et al. Generalizability of the nonalcoholic steatohepatitis Clinical Research Network histologic scoring system for nonalcoholic fatty liver disease. J Clin Gastroenterol. 2011;45(1):55-8.

[5] Gawrieh S et al. Effects of interventions on intra- and interobserver agreement on interpretation of nonalcoholic fatty liver disease histology. Ann Diagn Pathol. 2011;15(1):19-24.

[6] Davison B et al. Suboptimal reliability of liver biopsy evaluation has implications for randomized clinical trials. J Hepatol. 2020; in press.

- A scoring sheet was provided to the pathologist that included the grading system for each component along with simplified written criteria as to what constitutes NASH, borderline NASH, etc.
  - Results of the experiment, only demonstrated a significant improvement in the intra-observer agreement for ballooning.
  - No improvement in inter-observer agreement for lobular inflammation, portal inflammation, ballooning, or the classification of whether NASH is/not present
  - Automation of assessment of NAFLD histologic features has been raised as a solution to improve accuracy and reproducibility of the observer interpretation, as well as provide a continuous scale.

What and how to detect and quantify?
- Cardinal NAFLD lesions: macrosteatosis, lobular inflammation, hepatocyte ballooning, fibrosis
  - Also add portal inflammation which has prognostic value and is the dominate type of inflammation in pediatric patients
  - Consider also the architectural pattern of fibrosis and the collagen proportionate area.
- Machine-learning based
  - Directly identify NAFLD lesions using routinely available stains such as H&E, Mason trichrome, Sirius red, and requiring a digital slide scanner
- Algorithm-based
  - Correlates of NAFLD lesions are quantified using stained or unstained slides, usually requiring second harmonic generation microscopy, or two-photon excitation fluorescence microscopy, and a digital slide scanner
- Other
  - Adobe or color extraction methods are used usually to assess fibrosis and/or steatosis but are less robust and not the focus on the discussion.

What are machine learning and artificial intelligence?
- Machine Learning: algorithms and statistical models that learn from labelled training data, from which they are able to recognize and infer patterns
- Artificial Intelligence: ability of a machine to communicate, reason, and operate independently in both familiar and novel scenarios in a similar manner to a human
  - Commonly interchanged terms

What is the ML approach to NAFLD histology?
- Scientist Team including computer scientists, pathologists, hepatologists → provide digital images of selected NAFLD biopsies for the study → pathologist provides annotations for the biopsies → annotations are used to develop models and internally validate the labeled data, and correlation between continuous measurements and pathologist score → external validation of classifiers and the unlabeled data
- Supervised machine learning:
  - Decide which NAFLD lesion to quantify (i.e., steatosis) and develop a feature vector-based on learning data, a summary of attributes of the lesion that make the classifier most sensitive and accurate for detecting this lesion (size, shape, color, texture, surrounding area, etc)
    - The feature vector is known with this type of machine learning
  - Train the machine learning classifier
  - Test and make predictions on seen and unseen images
- Deep machine learning and neural networks
  - Machine learning models where you provide labeled data

- o Nodes act as mathematical operational centers that maximize the attributes from each area that the node is responsible for handling.
  - o The attributes are not known in deep machine learning algorithms
  - o Could have many layers of nodes, which creates an architectural network that resembles human neurons.
  - o Ultimately the algorithm makes a prediction
    - Convolutional neural networks – popular in video gaming, have recently become of interest in medical fields due to strength at image recognition
- Pathologist annotation software
  - o Web application developed where pathologists can remotely log-in to a system to view digital biopsy images stored on the cloud and provide labels for each NAFLD lesion.
    - Use the lesions to develop feature vectors, applying different thresholds and filters to arrive at the 'best' feature vector for each lesion.
    - Internal testing and validation process:
      - 10-fold cross validation
        - o Divide dataset into many sub-sets and randomly leave one out, while using the remaining data to train the model and test it on the data set that was left out. Repeat process until optimize the model.
- Classification approach
  - o The digital image is split into tiles of equal size, each tile is classified as either containing or NOT containing the feature (i.e., hepatocyte ballooning). The total percent of tissue with the feature is calculated.
    - % ballooning = total area ballooning tiles/ total tissue area
    - Results in a continuous measure
  - o Foundational step is to identify the white regions
    - Microscopic anatomic landmarks: central vein, portal vein, portal vein, bile duct, macrosteatosis, sinusoid
    - Important features to be able to identify the location of the inflammation or fibrosis within the tissue
  - o Initial study has excellent AUC for the classifier built for the white regions
    - 91% precision for identifying the bile duct
    - 82.5% precision for identifying the portal vein
    - 95.7% precision for identifying macrosteatosis
  - o The classifier for macrosteatosis showed high correlation with the average pathologist grade
    - With the automated continuous quantification of macrosteatosis, all the microscopic landmarks are labeled, and the percent of steatosis present in the image is provided (compared with semi-quantitative scoring 0-3).
  - o The classifier for hepatocyte ballooning had modest correlation with the average pathologist grade- due to lack of severe enough or frequent enough lesions to train the model
    - High AUC of 98%, precision of 91%
  - o The classifier for lobular inflammation had weak correlation with the average pathologist grade.
    - High AUC of 94.6&, but low precision of 69.6%
    - Further refined the model for lobular inflammation by adding new labels from more severe NASH biopsy samples- correlation with average pathologist grade improved, AUC remained high at 97.4%, and precision increased to 79.3%
  - o The classifier for portal inflammation has high correlation with the average pathologist grade

- High AUC of 97.9%, good precision of 82.1%
  - o The classifier for fibrosis had good correlation with the two pathologist grades
    - Fibrosis based on collagen proportionate area
    - Developed automated classifier to detect the architectural distribution of liver fibrosis in NAFLD biopsies, had high AUC for detecting bridging patterns and nodules.

Considerations for Development of Automated Methods for NAFLD Histology Analysis
- Development and community agreement on minimum set of acceptable standards for liver biopsy samples that will be used in clinical trial research
- Minimize bias in biopsy selection by including representation of the entire histological spectrum of NAFLD
- Obtain a large quantity of annotations by expert NAFLD pathologists to train the model
- How the biopsies are read, the number of readers providing the labels needs to be discussed and agreed upon.
- Understanding the tradeoffs of setting different thresholds
- External validation ensures the rigor of classifier performance- needs verification by expert pathologists for accuracy of lesion identification on unseen biopsy images, and validation of the classifier performance in completely different cohorts
- Discuss as a community how much weight to put on the strength of correlation between automated continuous measures and semi-quantitative assessments/scores of the regions being quantified.
- Regulatory approval considerations and the lack of 'explanability' for deep learning networks – it is unknown what attributes of a lesion are used or contribute to decision making process in nodes/ networks, this is referred to as the Black Box Factor.
- Machine learning and automated methods may be viewed as complimentary decision aids or guides, but not replacements for pathologists.
- Current State of Histology Analysis
  - o Manual, semi-quantitative, limited scale, issues with intra and inter observer variability, limited poll of experienced NAFLD pathologists, and limited access to experienced NAFLD pathologists.
- Future State of Histology Analysis
  - o Automated, continuous data, large scale, precise, reproducible, available, accessible.
  - o Pathologists are key partners in leading this transformation of the field.

HistoIndex

Presenter: Dean Tai

Slides: https://bit.ly/3bg7oDF

Challenges / Limitations of Liver Biopsy
- Challenge #1: Subjective and Semi-Quantitative Assessment[7,8]
  - o Ishak scoring system for progression and regression is not able to capture what is seen in biopsies- particularly post-treatment biopsies- where features from multiple stages (F5, F1, F3) can be observed within a biopsy sample. The Ishak system cannot explain what to do when there are mixed stages in the sample, which then introduces subjective interpretation.

---

[7] Chang PE, et al. Second harmonic generation microscopy provides accurate automated staging of liver fibrosis in patients with non-alcoholic fatty liver disease. PLOS ONE. 2018;13(6).
[8] Hsiao CY, et al. Improved quantitative assessment of HBV-associated liver fibrosis using second-harmonic generation microscopy with feature selection. Clin Res Hepatol Gastroenterol. 2020;44(1):12-20.

- o qFibrosis quantifies all the features observed in the samples as described in the NASH CRN scoring system simultaneously and provides a continuous assessment.
    - Recent data[9] demonstrating traditional pathologist assessment: 29% treatment arm achieved ≥1 stage reduction in fibrosis, vs 23% placebo (not significant). Second Harmonic Generation (SHG) assessment: 32% treatment arm achieved ≥1 stage reduction in fibrosis, vs 12% in placebo group (p=0.03).
  - o Quantifies automatically defined Zones 1, 2, and 3 – allows visualization of the interplay between steatosis and fibrosis in each zone, and separate quantification of each.
    - Recent data[10] analysis of patients with qSteatosis score reduced by 2 points demonstrates visualization of reduction of steatosis in all zones, and associated reduction in fibrosis in the same zones. The ability to directly co-localize fibrosis with steatosis on the same slide reveals MOA and is vital to assess and quantify drug efficacy.
  - o Distinguishing progression/regression within the same stage can be difficult as the role of septa in bridging fibrosis is a dynamic process. Progressive septa and broken septa can both be classified as F1/2. Both established septa and regressive septa can be classified as F3/4. This dynamic process is difficult to capture with a semi-quantitative assessment.
    - An automated tool to quantify septa dynamics has been developed with data reporting at AASLD 2020.
- Challenge #2: Sampling Error[11,12]
  - o Traditional pathology assessment approach requires minimum 1.5cm biopsy length for stable staging. Quantitative assessment can use 0.5cm biopsy length due to ability to identify the nano-features that are diffuse across biopsy samples.
- Challenge #3: Discrete NAS Scores and Fibrosis Stages Giving Rise to Inter- and Intra-Observer Variability[13,14]
  - o qFibrosis assessment is highly reproducible as a single parameter, as well as multiple combined parameters.
- Challenge #4: Not Linked to Outcome[15,16]

---

[9] Harrison SA, et al. In a Placebo Controlled 36 Week Phase 2 Trial, Treatment with MGL-3196 Compared to Placebo Results in Significant Reductions in Hepatic Fat (MRI-PDFF), Liver Enzymes, Fibrosis Biomarkers, Atherogenic Lipids, and Improvement in NASH on Serial Liver Biopsy. Presented at AASLD 2018. https://www.madrigalpharma.com/wp-content/uploads/2018/11/MGL-3196-Plenary-presentation-Nov-10-NASDAQ.pdf

[10] Harrison SA, et al. Steatosis and Fibrosis Measured as Continuous Variables on Paired, Serial Liver Biopsies in the Resmetirom (MGL-3196) 36-Week Phase 2 NASH Study. Presented at AASLD 2019. https://www.madrigalpharma.com/wp-content/uploads/2019/11/AASLD_Abst-2133_HistoIndex_MGL-3196-poster-FINAL.pdf

[11] Wang B, et al. SHG/TPEF-based image technology improves liver fibrosis assessment of minimally sized needle biopsies. Hepatology International. 2019;13(4):501-509.

[12] Wang Y, et al. Quantifying and monitoring fibrosis in non-alcoholic fatty liver disease using dual-photon microscopy. Gut. 2019;69(6):116-1126.

[13] Goh GBB, et al. Quantification of hepatic steatosis in chronic liver disease using novel automated method of second harmonic generation and two-photon excited fluorescence. Scientific Reports. 2019;9(2975).

[14] Wang Y, et al. Dual-photon microscopy-based quantitation of fibrosis-related parameters (q-FP) to model disease progression in steatohepatitis. Hepatology. 2017;65(6):1891-1903.

[15] Sun Y, et al. Quantitative assessment of liver fibrosis (qFibrosis) reveals precise outcomes in Ishak stable patients on anti-HBV therapy. Scientific Reports. 2018;8(1).

[16] Wang B, et al. Advanced septa size quantitation determines the evaluation of histological fibrosis outcome in chronic hepatitis B patients. Modern Pathology. 2018;31(10):1567-1577.

- o Previous research has linked quantitative assessment of fibrosis to clinical outcomes for HCV and CBV, and recently published data demonstrates a link to clinical outcomes for NASH.

Progress towards validation and regulatory acceptance
- Involved in many NASH phase 2 and 3 clinical trials, as well as pre-clinical studies
- Working towards FDA applications for invitro diagnostic (CDRH) and biomarker qualification program (CDER)
- Involved with multiple NASH consortia: Quantitative Ballooning Consensus, LITMUS, STEATOSITE, EMULSION

PathAI

Presenter: Andrew Beck

Slides: https://bit.ly/3nm4Je2

Focus of PathAI
- Deep-learning based platform focused on improving the accuracy of diagnosis and measurement of therapeutic efficacy for complex diseases, as well as applications in drug development.
- Mission to improve patient outcomes with AI powered pathology, and enable clinical development and approval of effective treatments.
- Grading and staging NASH disease severity
  - o Training a system on thousands of slides and annotations obtained from liver pathologists to then build a system to more accurately quantitate disease activity through the NAS, as well as stage fibrosis with NASH CRN and Ishak scores[17].
- Monitoring treatment response for NASH
  - o In addition to performing assessments of baseline biopsies, also utilizing system for monitoring treatment response in NASH studies. Allows studies to look at quantitative changes across different treatment arms beyond what can be done with manual scoring[18,19].

Identifying novel predictive histologic features
- Initial work completed which correlating AI assessment with that of expert pathologists, which addresses the challenge of reproducibility, accuracy, and precision.
- Going beyond that, recent study reported[20] on use of machine learning models to identify novel histological features that are predictive of clinical disease progression in patients with F3/F4 fibrosis.
  - o Machine learning based histologic features predicted disease progression in patients with bridging fibrosis.
  - o Obtained quantitative readouts of the components of NAS, fibrosis, and others.

---

[17] Pokkalla H, et al. Machine Learning Models Accurately Interpret Liver Histology in Patients With Nonalcoholic Steatohepatitis. Presented at AASLD 2019. https://www.natap.org/2019/AASLD/AASLD_37.htm
[18] Loomba R, et al. Safety and Efficacy of Combination Therapies Including Cilofexor/Firsocostat in Patients With Bridging Fibrosis and Cirrhosis Due to NASH: Results of the Phase 2b ATLAS Trial. Presented at EASL 2020. https://www.natap.org/2020/EASL/EASL_35.htm
[19] Taylor-Weiner A, et al. Validation of a Machine Learning-Based Approach (DELTA Liver Fibrosis Score) for the Assessment of Histologic Response in Patients With Advanced Fibrosis Due to NASH. Presented at AASLD 2020. https://www.natap.org/2020/AASLD/AASLD_121.htm
[20] Pokkalla H, et al. Machine Learning Models Identify Novel Histologic Features Predictive of Clinical Disease Progression in Patients With Advanced Fibrosis Due to Nonalcoholic Steatohepatitis. Presented at EASL 2020. https://www.natap.org/2020/EASL/EASL_65.htm

- o Despite a population of patients at approximately the same stage of disease as assessed by standard pathology (CRN F3), in the machine-learning based analysis there was significant prognostic information for predicting patients at highest risk of progressing to cirrhosis.
  - A 1-point increase in the ML-based NASH CRN fibrosis score (continuous measure, indication of severity of fibrosis) resulted in about 2-fold increase in risk of progression to cirrhosis.
  - The components of the score also provide prognostic information.
  - o Portal inflammation was a strong prognostic factor- large dataset demonstrating statistically significant association of extent of portal inflammation for F3 disease and risk of progression to cirrhosis.
  - o Similar results observed for patients with F4 fibrosis, and quantitative assessment of histological features of fibrosis was a significant predictor of clinical events.

Assessment of histologic response in patients with advanced fibrosis
- The proportion of tissue area at each fibrosis stage are independently associated with risk of progression. Can measure proportion at baseline, monitor post-treatment, and summarize the change as DELTA machine-learning NASH CRN fibrosis stage distribution. This plot shows how the severity of fibrosis has changed.
- Comparing responders and non-responders (based on traditional pathology assessment) for placebo and treatment group, identified that there was no difference in the DELTA score for the patients labeled as responders in the placebo arm.
  - o Compared with difference observed between responders and non-responders in the treatment arm, where the DELTA scores for responders significantly decreased.
  - o Can provide additional data beyond what can be done by manual scoring.

Biomarker and diagnostic development
- Ability to incorporate AI platform into prospective clinical trials and can work with CROs and sponsors to create fit-for-purpose locked algorithms to use within trials in real-time.
- FDA accepted LOI for biomarker qualification program for NASH drug development tool in 2020.

Pharmanest

Presenter: Mathieu Petitjean

Slides: https://bit.ly/3982buO

FibroNest engine and workflow
- Scientific hypothesis is that fibrosis has multiple histology phenotypes, and the mission is to quantity these phenotypes
- Inspired by authors of early scoring methods to focus phenotypic engine on three dimensions that cover the high-level description of the tissue and fibrosis, the morphometric (shape, form of features) description, and the architecture and structure of features
- Describe fibrosis by histogram, based on 37 phenotypic traits, and quantified in 7 statistical dimensions. Each histogram represents 1 biopsy, and as the disease progresses you can observe emergency of traits characteristic of more severe disease. Working to develop and optimize cutoff values to be able to classify the traits.
  - o Generates a method with a signal-to-noise ratio >100 – this strength helps to overcome the noise occurring in the background.
- Fully cloud-based platform, work with collagen-stained biopsy images (Sirius red, Masson's Trichrome, second harmonic generation images) and H&E slides for quantifying disease activity.
  - o After images are uploaded, they go through image normalization to eliminate the variability that comes from the imager (dust, particles), and use algorithms to

normalize the collagen (green) from the tissue (red), and to identify the collagen objects. Collagen is classified into two groups: fine collagen, and assembled collagen.
  - o The output are phenotypic maps which describe specific traits and how they correlate with fibrosis stage, as well as producing continuous scores
  - o Once calibrated, the phenotypic assay is frozen and kept constant for every model.
- ViQi platform uses next generation cloud-based bioimaging and computation infrastructure
  - o Allows web-based viewing of images to facilitate collaboration, visualization of image analysis and quantification, storage of annotations and metadata to use to explore intersection between biopsy data and biomarker data, AI and ML component

Quantifying NASH severity and drug response
- Adults: Results correlate with NASH CRN fibrosis stage[21] – by understanding the traits that change from F2 to F3, have developed an F2/3 score that can enhance the classification power and track the changes that occur between F2 and F3.
- Pediatrics: Results correlate with steatosis scores[22] to more accurately quantitate changes in disease severity. Additionally, have classified NASH type 1 vs NASH type 2 patients based on their different phenotypic traits[23].
- NASH spheroids[24] research exploring efficacy of anti-fibrotic compounds
- NASH fibrosis in rodents exploring dose effect of compounds

Clinical studies
- Noise and controls
  - o Guided by quality assurance programs developed in the field of oncology, as well as regulatory guidance and standards for diagnostic imaging[25,26].
  - o Developed Digital Pathology Imaging Charter, tool to ensure controls are in place to ensure robustness. Working document.
  - o Suggest same-slide workflow: after preparing and staining tissue, recommend digitizing slide and send digital image and slide to pathologist for scoring, while in parallel, upload to platform to generate phenotypic maps. Potential to aid the adjudication team. Result in phenotypic continuous scores to be able to understand what is going on within the categorical buckets.
- Data security

---

[21] Chen L, et al. Evaluation of a novel histology-based fibrosis phenotypic composite score and its correlation with NASH-CRN Fibrosis scores in patients with NASH. Presented at EASL 2020. https://static1.squarespace.com/static/559ae18be4b0dfcd52ac76e9/t/5f3289885127cf236595d82a/15971 47529892/EASL2020+POSTER+-+FibroNest+-+NASH+PSR+-++Final.pdf

[22] Pitkowsky Z, et al. Automated Steatosis Morphometric Scores Benchmark the Pathology-Based Quantification of Steatosis in Pediatric NASH/NAFLD Populations. Presented at AASLD 2019. https://static1.squarespace.com/static/559ae18be4b0dfcd52ac76e9/t/5dc41a109a9f46178c4d400d/15731 32819646/MAFIOSO+steatosis+poster+ASLD+2019+FINAL.pdf

[23] Petitjean M, et al. Automated Morphometric Fibrosis Phenotyping of NAFLD Biopsies Digital Images Helps Classify NASH-Type 1 Vs NASH-Type 2 in Early Fibrosis Pediatric Patients. Presented AASLD 2019. https://static1.squarespace.com/static/559ae18be4b0dfcd52ac76e9/t/5dc41ac29a9f46178c4d4ab9/157313 2995737/AASDL2019-%232299-+NASH1+vs+NASH2.pdf

[24] Petitjean M, et al. Novel phenotypic image analysis of 3D NASH model generate quantitative and continuous scores for the evaluation of fibrosis in vitro. Presented at AASLD 2020. https://static1.squarespace.com/static/559ae18be4b0dfcd52ac76e9/t/5fac3fae47c4690551f78ffd/1605124 016804/AASLD2020+-+FIBRONEST+NASH+SPHEROIDS+AASLD2020+POSTER+10.07.020.pdf

[25] U.S. FDA. Clinical Trial Imaging Endpoint Process Standards Guidance for Industry. https://www.fda.gov/media/81172/download

[26] U.S. FDA. Considerations for Use of Histopathology and Its Associated Methodologies to Support Biomarker Qualification Guidance for Industry. https://www.fda.gov/media/82768/download

- o Platform has had successful audit for attack vulnerability and generate dual-site data back as it is generated. Have a full audit trail and store raw data – developing library of data that can be used later for post-analysis.
- o Tool to aid pathologists in assessment.
- Future considerations
  - o Development of recommendations/ process standards for NASH clinical trial digital pathology endpoints
  - o Define what is an adequate digital liver biopsy (ex. What is the minimum length for a digital biopsy)
  - o Should robotic pathology/ automation participate in the adjudication process?

Panel and Group Discussion

Slides: https://bit.ly/3hK7lRG

Q: How do technologies differentiate between portal and lobular inflammation, especially at interface level?
- Obtain 100,000's of annotations from expert-trained liver pathologists of the different patterns, and deep-learning systems learns from this to sub-classify inflammation into different tissue regions (portal, lobular, interface).

Q: What is the precision and accuracy of identifying each variable (i.e., fibrosis, steatosis, inflammation, and ballooning) when compared with pathologists?
- With AI approaches, reading the same image will consistently produce the same output, compared to traditional pathology when inter and intra observer variability of a slide are a noted issue.
- Accuracy is tied to whatever the 'ground truth' is, or, the ideal expected result.

Q: How is digital pathology and AI technology being used in clinical trials?
- Context for use is really as a validation of what is currently the primary endpoint for clinical trials and together with other surrogates to be able to unravel issues regarding efficacy signals as well as defining mechanisms of action for unique targets. Opportunity to leverage these technologies to be able to really dissect phase 2 and early phase 3 trials.

Q: What are the strengths and limitations of digital pathology?
- Digitizing slides is very useful for clinical trials as it allows pathologists to share and make consensus. Would like opinion of regulatory authorities on this.
- AI is very exciting area though still requires validation. Still need to define the adequate biopsy size. Digital pathology and AI will be able to decrease variability between pathologists, as well as provide information that the pathologist cannot see.
- Identifying the ground truth: steatosis and fibrosis have been the first and easier targets for this technology because they are discreet anatomic structures that everyone can agree on. Inflammation and ballooning are very different.
  - o Balloon cells are not discrete objects with clearly defined definition, and the cells don't all look the same but are rather more like a range. Doing consensus reading to identify the 'best' balloon cells, as well as those less agreed upon to build a scale over a continuum- when the machine has been taught to recognize the continuum, a threshold can be set to identify what should be called ballooning.
  - o Having more labels of all sorts of balloon cells, and the whole spectrum of inflammation would help the model learn to distinguish these features.
  - o There needs to be back and forth between pathologists and the programmers building the tools.

Q: Can CK-18 staining, or other stains, help to identify balloon cells?

- Limited experience with this staining – not feasible in a clinical trial, but rather should be an exploratory approach. Adding a histochemistry on top of everything else adds a layer of complexity that should probably be avoided. Get variable staining through the thickness of the biopsy, affects how you interpret the cells and not sure if provides any advantage.

Q: Human assessment of histology allows for context to be taken into account (age, gender, race, ethnicity, biomarkers, etc), with AI technologies, is there work on advanced modeling paradigms to integrate more nuanced information?

- It is critical to define exactly what you want the system to do and the context of use that you want to validate it for. The tools are developed for a specific use in a specific context (i.e., patient with suspected NASH) and wouldn't necessarily be appropriate for a different use or for a different context (i.e., all liver biopsies) because it would be in a different context and the tool would encounter features it has not been trained to recognize.
- Defining, validating, and using a tool according to its context of use is extremely important. A perfect tool in one context, may end up being dangerous in a different context.
- This would be more relevant in a clinical practice setting, in a clinical trial, the pathologist is blinded to clinical and biological data. In that context you would not want to mix this data into the AI tool, because it would not happen with traditional pathology approach.
    - Pathologist has been trained to understand that clinical and biological context- they may not have the specific clinical history, but would have more context than the AI has to make determinations.
    - Could also be relevant in clinical trials for example in identifying a DILI reaction or an immune mediated drug event.

Q: How do you see this technology being used in clinical practice?

- Clinical trials and clinical practice would have very different contexts – for clinical trials, a very sensitive endpoint is needed to validate the primary outcomes of a trial
- When translating a drug into clinical practice, there are many nuances introduced that are not accounted for in clinical trials. Community pathologist interpretation of presence or absence of NASH is generally not the same as that of a research pathologist in regards to how it's scored and graded. Not sure it is feasible for this technology to be integrated into clinical practice in the near future.

Q: What regulatory guidance exists regarding use of digital pathology and artificial intelligence in clinical trials?

- Two relevant FDA guidance documents:
    - Clinical Trial Imaging Endpoint Process Standards Guidance for Industry: https://www.fda.gov/media/81172/download
    - Considerations for Use of Histopathology and Its Associated Methodologies to Support Biomarker Qualification Guidance for Industry: https://www.fda.gov/media/82768/download

Q: Are there examples from other disease areas to reference?

- There are many examples that can be drawn from the oncology field. Everything starts with the intended use statement and how the AI model is going to be applied.
    - The intended use must be carefully constructed, for example looking at the entry biopsy vs. looking at biopsy after enrollment are two different populations.
    - Regulators examine how good the data is that has been used for the development of the algorithm – how it's been collected, what kind of data (clinical trial, real world), what kind of algorithm is being used and how it works, and perform analytical and clinical validation based on the methodology.

- Analytical validation questions – length of biopsy, staining used, quality of the staining, FDA approved device for digital imaging (Aperio, Philips).
- When examining the data, the cut points selected for the features of NASH are critical.

Q: Why is it possible to reduce the biopsy specimen length with digital assessment?
- Biopsies contain nano features such as hepatocytes and fine collagen which cannot be seen by the pathologist but can be seen in digital platforms.