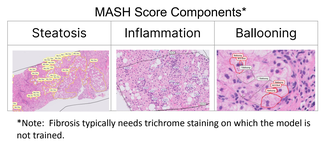
# Prediction of MASH features from liver biopsy images using a pre-trained self-supervised learning model

Yang Wang*, Saurabh Vyawahare*, Carson McNeil*, Jessica Loo*, Marc Robbins*, Roman Goldenberg*
*Verily Life Sciences, 249 E Grand Ave., South San Francisco, CA 94080, USA

**Contact**:Roman Goldenberg
rgoldenberg@verily.com

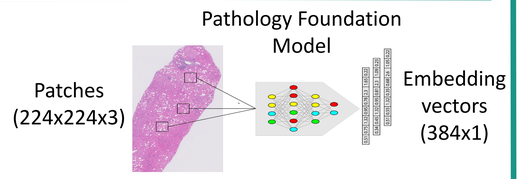
## Introduction

- Neural networks that identify features in histology slides have been widely adopted. These are trained using pathologist annotations. However, obtaining annotations is very expensive due to the large image sizes, large number of features, and high labor costs.
- Annotation requirements can be significantly reduced by using a self-supervised pathology foundation model as a module of the final network architecture. Such models are trained on existing corpuses of *unlabeled* training data, such as the Cancer Genome Atlas (TCGA).
- We apply a pathology foundation model to predict features relevant to metabolic dysfunction-associated steatohepatitis (MASH), including inflammation, steatosis, and ballooning.

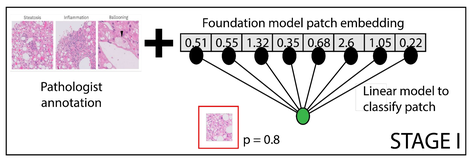



## Method

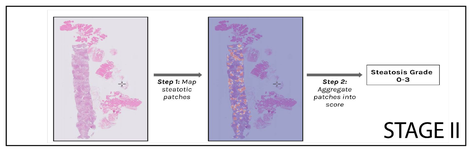
- Our pathology self-supervised learning (SSL) model is based on vision transformers[1], trained on hematoxylin and eosin (H&E) slides from TCGA. This model produces embeddings from patches.




- Patch embeddings are used to predict features in a two-stage process: I) Produce patch classifier using embeddings, II) Aggregate patch classifications over slides and use Bayesian classification for slide-level prediction.




- The results from SSL-based model are compared against fully supervised model[2] to demonstrate any benefits on generalizability and required sample size for training.




## Dataset

The MASH-labeled data is from the CENTAUR study[3] which we previously analyzed using fully supervised models.[2] Using liver tissue biopsy H&E images we created a dataset consisting of 110, 77, and 179 training, evaluation, and test images respectively.

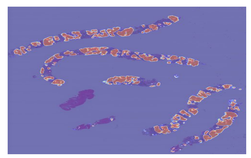
## Results

- Pathology foundation model produced patch-level results (Stage I) similar to fully supervised models we reported previously. However, at the slide level (Stage II) superior results were obtained for grading of ballooning and lobular Inflammation, while steatosis performance remained the same.
- We believe this reflects better generalizability of the model and resilience to messy patch-level labels.
- Less data is required to patch-wise segment features of MASH to a given level of accuracy, when the foundation model is finetuned compared to finetuning a (pretrained) ResNet model, suggesting better data efficiency.

| **AUC scores** | Fully Supervised[2] | Foundational Model |
|---|---|---|
| Inflammation | 0.76 | **0.85** |
| Ballooning | 0.77 | **0.88** |
| Steatosis | **0.91** | **0.91** |




Example of segmentation map for steatosis



## Conclusions

Self-supervised pathology foundation models can improve machine learning performance in digital pathology for liver diseases like MASH. These models can lead to better downstream generalization and enable training using smaller labeled datasets.

### References

1) Lai, Jeremy, et al. "Domain-specific optimization and diverse evaluation of self-supervised models for histopathology." *arXiv preprint arXiv:2310.13259* (2023).
2) McNeil, Carson, et al. "An End-to-End Platform for Digital Pathology Using Hyperspectral Autofluorescence Microscopy and Deep Learning Based Virtual Histology." *Modern Pathology* 37.2 (2024): 100377.
3) Ratziu, Vlad, et al. "Cenicriviroc treatment for adults with nonalcoholic steatohepatitis and fibrosis: final analysis of the phase 2b CENTAUR study." *Hepatology* 72.3 (2020): 892-905.

### Model Availability

Pathology foundational model is available here:
https://github.com/Google-Health/imaging-research/tree/master/path-foundation

### Disclosures

All authors are current employees with equity interests at Verily Life Sciences LLC.

### Acknowledgements

We thank Allergan/AbbVie for their collaboration and sharing the CENTAUR clinical trial data, and Google Health team for helpful discussions.