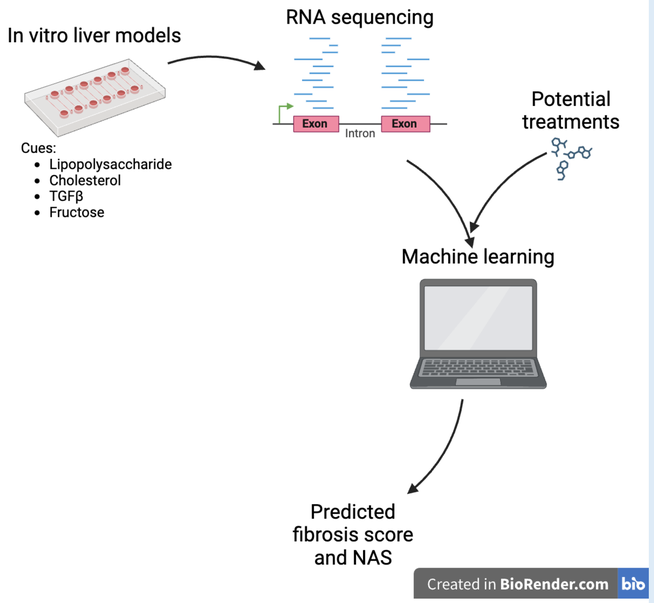
Transcriptomics-based histological scoring for metabolic-associated steatohepatitis using machine learning

Jeannie She[1], Nikolaos Meimetis[1], Douglas Lauffenburger[1]
[1]MIT Department of Biological Engineering

Metabolic-associated steatohepatitis (MASH) lacks clinical therapies that target liver inflammation

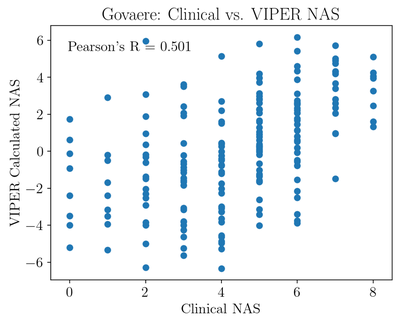
- MASH, the advanced stage of metabolic-associated steatosis liver disease, is characterized by **severe accumulation of fat in the liver**.
- Patient disease severity is determined by **histological scoring** of liver biopsy, producing two clinical scores:
  - ➢ **Fibrosis stage score** (ranges from 0 to 4)
  - ➢ **NAS** (NASH Activity Score) (ranges from 0 to 8)

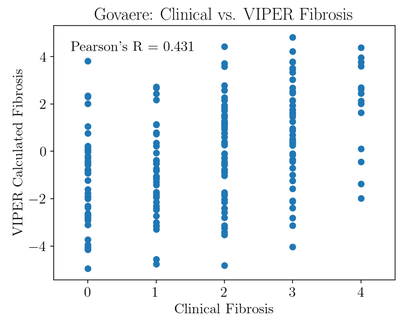


- The effort towards **developing accurate *in vitro* liver models** is crucial for better understanding disease progression and therapy development.
- **Goal:** create a **machine learning model** to **histologically score** transcriptomic data with **fibrosis stage score** and **NAS**, bringing **clinical meaning** to the *in vitro* liver model conditions.

Inferred histological score using known markers of the disease correlates well with clinical scores

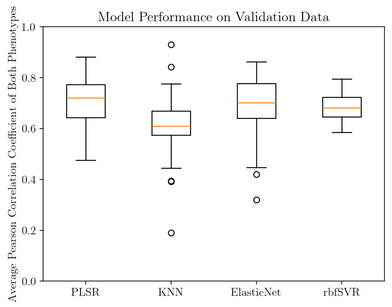
- VIPER algorithm[2] evaluates the differential enrichment of a protein's transcriptional targets (its "regulon") across conditions.

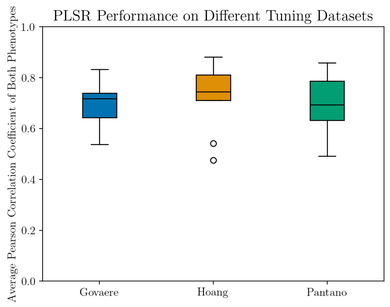




Machine learning predicts histological scores from transcriptomic data

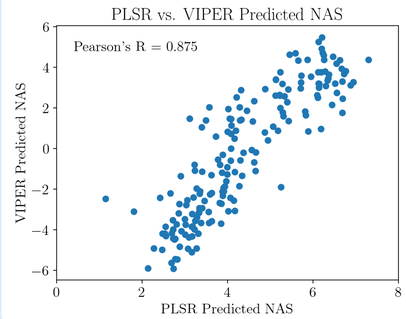
- Trained different ML models using 10-fold cross validation to predict NAS and fibrosis.
- Validated in external clinical datasets.
- Identified the **Govaere et al.**[3] **dataset** as the ideal dataset to use to train a **PLSR** model due to greatest breadth of data and high median Pearson's correlation coefficient.

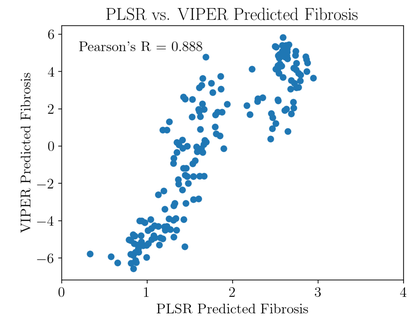




Scoring the *in vitro* model

- Both PLSR or known markers seem to report similar information about the model; suggests using a consensus of both in the future
- Few models were scored the **extreme** Fibrosis stage score (score of 4) or NAS (score of 8).
  - ➢ The number of datapoints in the training dataset with severe scores were far fewer than the less severe scores (n=14 for Fibrosis score of 4, n=51 for Fibrosis score of 2).
  - ➢ The *in vitro* models[1] may not have necessarily captured extreme disease states.





Significance and next steps

Given the transcriptomic profile of an *in vitro* liver model, we can now predict how a clinician may histologically score its MASH severity with a ML model that produces a Fibrosis stage score and NAS.

We hope to utilize this model in a pipeline for *in silico* drug repurposing, where we input drugs' **chemical structures** into a ML model to predict the **histological scoring** of a liver model. This puts us further on the path towards finding a more targeted MASH therapy.

Thank you to Nikolaos Meimetis and the Lauffenburger Lab for the support on this project.

[1]Kostrzewski, T., Snow, S., Battle, A.L., Peel, S., Ahmad, Z., Basak, J., ... Hughes, D. (2021). Modelling human liver fibrosis in the context of non-alcoholic steatohepatitis using a microphysiological system. *Commun Bio, 4*(1080). https://doi.org/10.1038/s42003-021-02616-x.

[2]Alvarez, M.J., Shen, Y., Giorgi, F.M., Lachmann, A., Ding, B.B., Ye, B.H., Califano, A. (2016). Functional characterization of somatic mutations in cancer using network-based inference of protein activity. *Nature Genetics*. *48*(8):838-47. https://doi.org/10.1038/ng.3593.

[3]Govaere, O., Cockell, S., Tiniakos, D., Queen, R., Younes, R., Vacca, M., … Anstee, Q.M. (2020). Transcriptomic profiling across the nonalcoholic fatty liver disease spectrum reveals gene signatures for steatohepatitis and fibrosis. *Sci Transl Med*. *12*(572). https://doi.org/10.1126/scitranslmed.aba4448.