Jeannie She

20.260

Professors Lauffenburger and White

22 Mar 2024

20.260 Midterm Project

Introduction to the Biological Problem

Non-alcoholic fatty liver disease (NAFLD) affects almost 25% of the population globally and has become the leading cause of liver disease worldwide. Non-alcoholic steatohepatitis (NASH) is a subtype of NAFLD, which is characterized by "lipid accumulation, infiltration of immune cells, hepatocellular ballooning, collagen deposition and liver fibrosis" (Kostrzewski et al., 2021). In their paper, Kostrzewski et al. expressed the high unmet need of an accurate fibrotic disease model on which new drugs and disease conditions can be tested. Therefore, they developed an *in vitro* microphysiological system (MPS) and proved its phenotype to be highly reminiscent of an advanced disease state.

Insight Gained from the Original Study

Kostrzewski et al. began by performing transcriptomic profiling of their NASH MPS model and a control, "lean" microtissue model with RNAseq from both samples. They found that their NASH model identified a correlation with human fatty liver disease as characterized by the DISEASES database. Then, they found that up to 45% of the differentially expressed genes that represent "key changes in human NAFLD/NASH," according to the 2016 study by Teufel et al, matched in the same analysis of their disease model. Surprisingly, only "0.01% to 10% of the DEGs in NASH patients were also differentially expressed in the mouse pre-clinical models"

(Kostrzewski et al., 2021). The MPS evidently proves to be a much more useful disease model than the commonly used mouse model for NASH.

Next, the authors tested different media conditions on their MPS, ranging from lean to fat media, and added various permutations of cues (sugar (fructose), fat (cholesterol), TGFβ, and lipopolysaccharide (LPS)), all of which were selected for "clinical association with NASH" (Kostrzewski et al., 2021). They then performed gene ontology (GO) term enrichment and cluster analysis to visualize which gene sets were differentially expressed in each cue and media condition. Finally, Obeticholic acid and Elafibranor, two compounds in late stage clinical development with promising anti-NASH effects, were added at gradient concentrations to the MPS to evaluate the model's efficacy at pre-clinical drug testing.

My Approach to Bioinformatics Analysis

First, the authors reported the DEGs between their NASH MPS model and a control lean microtissue as a validation of their model. Later on, they had the MPS cultured in CN Bio Innovations' HEP-Fat and HEP-Lean media as part of their cue condition experiments. Their pathway enrichment analysis was produced by comparing each cue condition with the HEP-Lean, low non-parenchymal cells (NPC) sample. However, the differences between the HEP-Lean and HEP-Fat media types were not observed in great detail. Therefore, I decided to perform differential gene expression analysis and gene set enrichment analysis contrasting these two samples specifically. Not only would this give insight to how the media impacted the cells, but it might also suggest differences between the lean control in the cue experiment versus the lean control in the initial MPS disease model validation.

I used R's "voom" package to assist in normalizing the data: first with counts per million (CPM) normalization, and later log2-normalization, since gene expression data tends to follow a

log-normal distribution. The “limma” package in R produced t-statistic values for each gene that represented the level of differential expression after contrasting the HEP-Fat condition versus the HEP-Lean condition. Afterwards, gene ontology (GO) term analysis was performed to identify the top gene sets enriched between these two conditions. In Figure 1, the normalized enrichment score (NES) is plotted for each of the top 35 differentially expressed pathways.

Second, while the authors performed GSEA on their samples with different cue conditions, they did not perform any sort of clustering analysis. Therefore, I felt it apt to see if the conditions would cluster together in hierarchical clustering. First, I filtered for the most highly expressed genes from the raw RNAseq counts by only keeping those that had an expression of 50,000 or greater, which resulted in 97 genes total. Importantly, the data was log-transformed ahead of clustering to ensure that all of the data could be analyzed on the same scale. I used Python’s “seaborn” package to assist with performing and displaying the hierarchical clustering. In Figure 2, the dendrograms resulting from hierarchical clustering of the samples and genes can be viewed on the x- and y-axis respectively.

Discussion of Biological Significance from Analysis

In Figure 1, the pathways with a positive NES score indicate an upregulation of that gene set in the MPS with HEP-Fat media, and those with a negative NES score indicate a downregulation. Most of the very highly upregulated pathways are related to an uptick in mitochondrial activity and cytoplasmic protein production, including the ones mentioning ATP, nucleoside triphosphates, and ribosomal biogenesis. Interestingly, the gene set relevant for regulation of fatty acid oxidation also demonstrated upregulated activity. It has been characterized that the inhibition of fatty acid oxidation results in the fatty acid accumulation in the liver, directly contributing to the disease phenotype seen in NAFLD and NASH patients

(Reddy & Sambasiva, 2006). Therefore, I hypothesize that the cells subjected to the HEP-Fat media are experiencing unusual levels of inhibition of this specific pathway; therefore, the pathway is upregulated with an effort to overcome this perturbation.

In terms of the downregulated gene sets, cerebellum morphogenesis and hindbrain morphogenesis stood out to me in particular. Upon first glance, it feels particularly peculiar since the MPS consists of primary human liver cell culture, namely hepatocytes. However, it turns out that the DLC-1 (deleted in liver cancer 1) gene is part of the hindbrain morphogenesis gene set (Blake et al., 2021). DLC-1 was initially identified to be frequently deleted in hepatocellular carcinoma (a cancer that can follow from severe NASH). It was identified in 2008 that DLC-1 is a tumor suppressor gene that is commonly methylated or otherwise deleted in numerous other cancer types as well (Lian & Lo, 2008).

In Figure 2, we can see that the samples that received similar cues tended to cluster together more than the samples with the same media. In the bottom right, those samples that received TGFβ regardless of their non-parenchymal cell (NPC) amount or their growth media appear to cluster together well. Significantly, they all had a downregulated expression of genes like A2M and SELENOP (which have been related to aggressive tumors and cancer-associated fibroblasts in lung adenocarcinoma) and TM4SF4 (which is related to cell development and proliferation and emerging as a possible anticancer target) (Vasiukov, 2023; Rahim, 2023). TGFβ was specifically chosen as a cue by Kostrzewski et al. for its regulation of collagen production and therefore important role in liver fibrosis.

In the middle of the plot, most of the samples with reduced levels of NPC ended up clustering together as well. The researchers had intended for low NPC to correlate with a non-physiological model, as a contrast with their high NPC, physiological models. The

clustering is a confirmation of the fact that the amount of non-parenchymal cells significantly affects the accuracy of a liver model. Ultimately, from the hierarchical clustering I hypothesize that the effect of the cues outweighs the effect of the media, but still depends heavily on the types of cells present in the MPS.

Possible Experimental Next Steps

To evaluate the role of different media conditions on the MPS, looking into the media composition could give light to how the disease model is being curated. To examine and identify additional NASH related cues in the development of the disease, more cue permutations could be introduced to the MPS, as they seem to have perturbed the MPS significantly with a lot of potential biological conclusions to be made. Though there likely exists a lot of known cues related to NASH and liver fibrosis, it would be elucidating to also introduce cues that researchers wouldn't expect to be related to the fatty liver phenotype. These cues can be generated again with bioinformatics analysis by looking at the differentially expressed pathways and identifying small molecules whose concentrations upstream of a pathway significantly impact the downstream effects. Possibly a unique combination of unrelated cues, added to the media of MPSes, may generate a notable phenotype that can again be analyzed by achieving RNAseq data and performing genomics analysis.

Additionally, the authors mentioned that most of the MPS disease models had represented a milder phenotype, whereas the introduction of the cues in this paper were focused on developing a severe NASH disease phenotype. It would be extremely useful for pre-clinical drug experiments if a media and cue formulation for the different stages of NAFLD, NASH, hepatocellular carcinoma, or cirrhosis diseases could be developed for the liver MPS. This could be developed by expanding the search for more media types and cue concentrations that may

provoke different levels of disease severity. It would be interesting to observe how the differentially expressed genes differ from stage to stage so we can have an even clearer understanding of the evolution of these diseases.

Conclusions

Ultimately, Kostrzewski et al. have established a convincing argument for the advantages of a microphysiological system as a pre-clinical disease model for non-alcoholic steatohepatitis. Performing different methods of bioinformatics research (hierarchical clustering, differentially expressed gene analysis, and gene set enrichment analysis) on the same data as the authors has given rise to intriguing conclusions that signal a clearer path forward for more useful experimental analysis. Amidst the successful development of other viable microphysiological systems and the continued application of genomics analysis, the future of pre-clinical models is on its way to change.

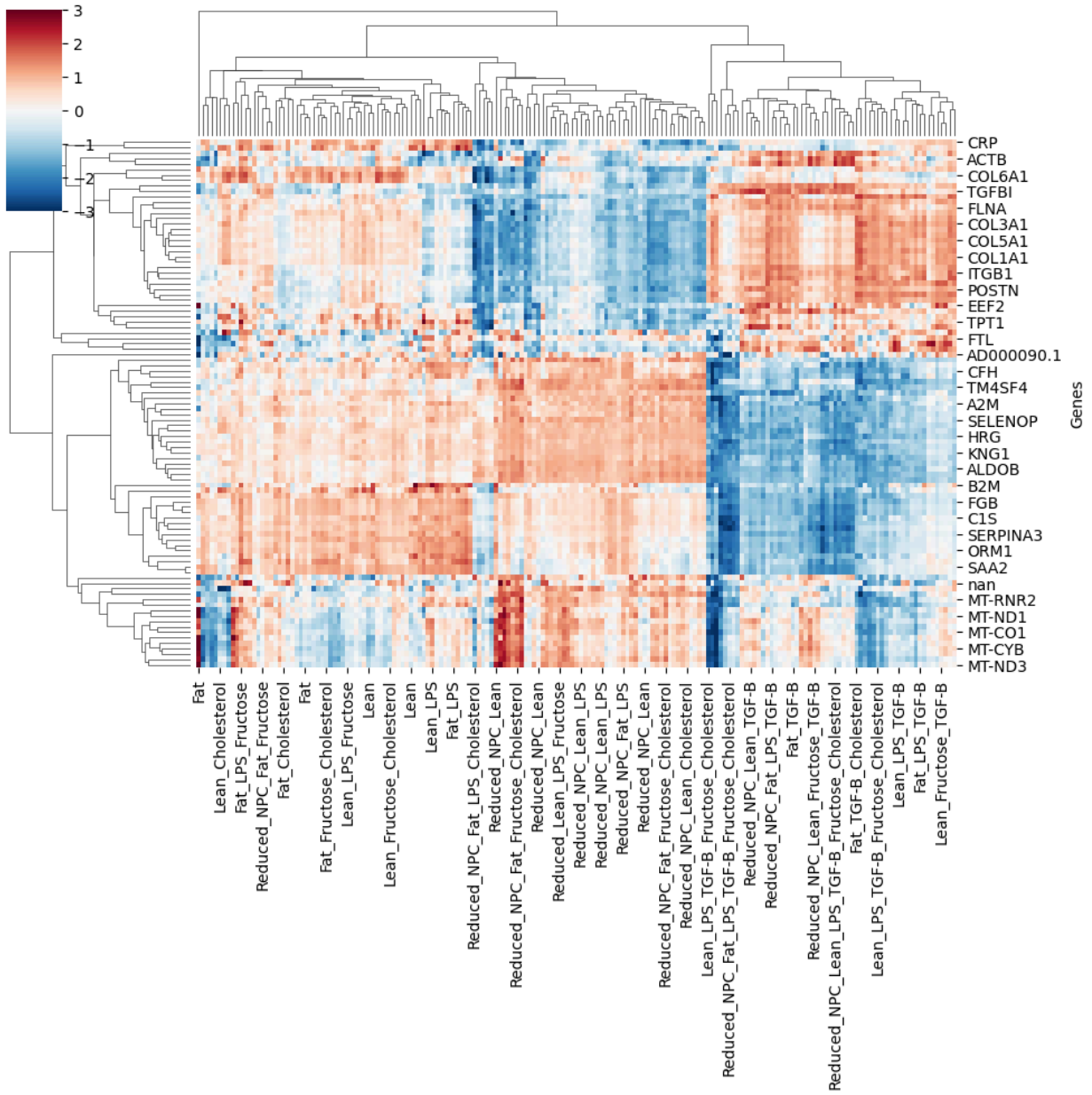

Figure 1. Hierarchical clustering of the top 97 highly expressed genes by the liver microphysiological system treated with Lean media, Fat media, sugar (fructose), fat (cholesterol), TGFβ, and lipopolysaccharide (LPS). Some samples had reduced non-parenchymal cells (NPC), and the others had physiological levels of NPC.

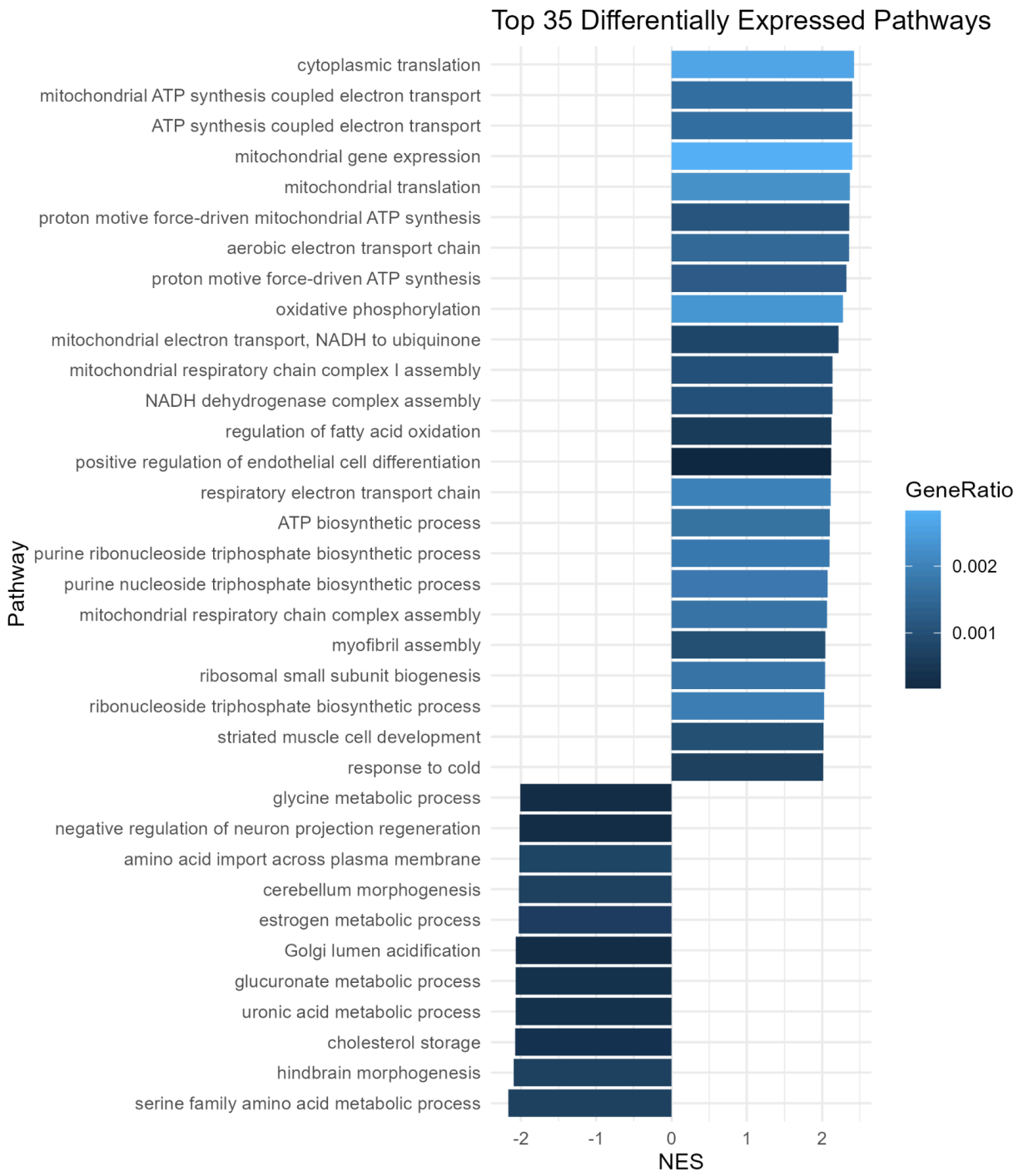

Figure 2. Top 35 differentially expressed pathways of a liver microphysiological system grown in HEP-Fat media compared to a system in HEP-Lean media, determined by GO term enrichment analysis. All of the pathways had an adjusted p-value smaller than 0.05, indicating significance.

Works Cited

Blake JA, Baldarelli R, Kadin JA, Richardson JE, Smith CL, Bult CJ; Mouse Genome Database Group. 2021. Mouse Genome Database (MGD): Knowledgebase for mouse-human comparative biology. Nucleic Acids Res. 2021 Jan 8;49(D1):D981-D987.

Kostrzewski, T., Snow, S., Battle, A.L. *et al.* Modelling human liver fibrosis in the context of non-alcoholic steatohepatitis using a microphysiological system. *Commun Biol* 4, 1080 (2021). https://doi.org/10.1038/s42003-021-02616-x.

Rahim NS, Wu YS, Sim MS, Velaga A, Bonam SR, Gopinath SCB, Subramaniyan V, Choy KW, Teow SY, Fareez IM, Samudi C, Sekaran SD, Sekar M, Guad RM. Three Members of Transmembrane-4-Superfamily, TM4SF1, TM4SF4, and TM4SF5, as Emerging Anticancer Molecular Targets against Cancer Phenotypes and Chemoresistance. Pharmaceuticals (Basel). 2023 Jan 11;16(1):110. doi: 10.3390/ph16010110. PMID: 36678607; PMCID: PMC9867095.

Reddy, J. K., & Sambasiva Rao, M. (2006). Lipid metabolism and liver inflammation. II. fatty liver disease and fatty acid oxidation. *American Journal of Physiology-Gastrointestinal and Liver Physiology*, *290*(5). https://doi.org/10.1152/ajpgi.00521.2005

Teufel, A., Itzel, T., Erhart, W., Brosch, M., Wang, X. Y., Kim, Y. O., von Schönfels, W., Herrmann, A., Brückner, S., Stickel, F., Dufour, J. F., Chavakis, T., Hellerbrand, C., Spang, R., Maass, T., Becker, T., Schreiber, S., Schafmayer, C., Schuppan, D., & Hampe, J. (2016). Comparison of Gene Expression Patterns Between Mouse Models of Nonalcoholic Fatty Liver Disease and Liver Tissues From Patients. *Gastroenterology*, *151*(3), 513–525.e0. https://doi.org/10.1053/j.gastro.2016.05.051

Vasiukov G, Zou Y, Senosain MF, Rahman JSM, Antic S, Young KM, Grogan EL, Kammer MN,

Maldonado F, Reinhart-King CA, Massion PP. Cancer-associated fibroblasts in early-stage lung adenocarcinoma correlate with tumor aggressiveness. Sci Rep. 2023 Oct 17;13(1):17604. doi: 10.1038/s41598-023-43296-3. PMID: 37848457; PMCID: PMC10582049.

Yi-Chun Liao, Su Hao Lo, Deleted in liver cancer-1 (DLC-1): A tumor suppressor not just for liver, The International Journal of Biochemistry & Cell Biology, Volume 40, Issue 5, 2008, Pages 843-847, ISSN 1357-2725, https://doi.org/10.1016/j.biocel.2007.04.008.