Identifying Potential Therapeutic Target for Nonalcoholic Steatohepatits (NASH) with Fibrosis

By: Renee Ge, Luyang You, Yun Jin Park

Introduction:

With a global prevalence of 30%, Nonalcoholic Fatty Liver Disease(NAFLD) is the most common liver disease worldwide (Younossi et al., 20123). NAFLD is a condition where fat builds up in your liver. There are two forms of this disease: Nonalcoholic Fatty Liver (NAFL) and Nonalcoholic steatohepatitis (NASH). NAFL is simply fat retention in the liver, which is relatively benign compared to NASH. In NASH, however, one typically observes liver inflammation and damage along with fat in the liver. This inflammation and liver damage can lead to liver fibrosis (scarring). Liver fibrosis is the initial stage of irreversible liver damage known as cirrhosis which may also lead to cancer. (Suppli et al., 2019; U.S. Department of Health and Human Services, 2021)

For prevention and treatment of this prevalent and dangerous disease, better understanding of the disease mechanism and its progression is needed. As such, we'd like to use RNA-seq analysis techniques to examine how cell composition changes through NASH and liver fibrosis progression and identify potential gene targets for future research. In particular, at the cellular level, we are interested in hepatic stellate cells (HSCs) which are the major cell type involved in liver fibrosis. Although their current role is unclear, under healthy conditions we typically see quiescent HSCs (qHSCs). These cells can be triggered to differentiate into activated HSCs which are responsible for the production of collagen which can lead to cirrhosis. Thus, we will focus our analysis on HSCs in general. We replicated some of the results from the paper titled "Single-cell and bulk transcriptomics of the liver reveals potential targets of NASH with fibrosis" by Wang et al (2021).

Data:

Single-cell RNA-seq data of human liver (E-MTAB-10553) was obtained from ArrayExpress. This dataset contained integrated data of 17,810 cells from 6 healthy individuals (Wang et al., 2021) and 157,619 cells from four healthy and three cirrhotic human livers (GSE136103). In our proposal we had planned to analyze a dataset from NCBI Bioproject with accession number PRJNA512027. However, this data was available in the form of 192 SRA files totalling over one terabyte of data and we did not have the computing power to analyze this. As a result we found an alternative dataset similar in construction which also took liver biopsies of individuals at different stages of disease progression of NAFLD. This dataset was Human bulk RNA-seq data from GEO database (GSE126848) containing healthy normal weight (n=14) and obese (n=12) individuals, NAFL (n=15) and NASH (n=16) patients. Mouse bulk RNA-seq data was obtained from the GEO database (GSE116987) of liver fibrosis in vivo and vitro models.

Methods:

Single-cell RNA-seq data:

This dataset was obtained as a Seurat object with cell type assignment contained inside. We made UMAP plots based on this clustering and used Seurat's FindMarkers function to find differentially expressed genes between the aHSC and qHSC cell types. This was sorted into two lists of upregulated and downregulated genes. We also compared the proportion of aHSC cells between cirrhotic and healthy livers.

Human Bulk RNA:

This data was available as a count matrix with genes as rows and samples as columns. We used MuSiC to deconvolute the dataset and obtain the cell type proportions using our scRNA-seq dataset. However, since our scRNA-seq dataset contained 175,429 cells we downsampled this to 50 cells from each of the 15 annotated cell types in order to run the cell type proportion estimation. We then examined the proportion of aHSCs, LSECs, Kuppfer cells, and Cholangiocytes across the four subject levels (healthy, obese, NAFL, NASH) and performed Wilcoxon Rank Sum tests with a Benjamini-Hochberg p-value adjustment. We then used DESeq2 to discover differentially expressed genes in liver cells of 3 different conditions (obese, NAFL, NASH) compared to liver cells in healthy individuals.

Mouse Bulk RNA:

Mouse bulk RNA-seq dataset was obtained as two datasets and was analyzed separately. For the in vivo mouse model, mice were given CCL_4(carbon tetrachloride; a known chemical for inducing liver fibrosis (Dong, 2016)) for either 4, 6, or 8 weeks. Control mice received vehicle for 8 weeks. The dataset was obtained as a count matrix of reads with each sample being the mice at the end of their treatment period (0, 4, 6, or 8 weeks). Based on the experimental information we considered the "0 weeks" (control) mice to be qHSCs and "8 week" time-point as aHSCs(Marcher, 2019). For the in vitro mouse model, qHSCs were seeded on normal plastic cell culture dishes for 12 days and were differentiated into aHSCs. The data we obtained was in the form of a count matrix with samples being a range of time points between 2 hours and 12 days (Marcher, 2019). The two hour sample contained only qHSCs and 12 days contained only aHSCs. We performed differential gene analysis on both models between aHSCs and qHSCs using DEseq-2 and extracted significant up and down-regulated genes and compared them with the differential expression results from the human RNA.

Evaluating Translatability:

Taking all of the genes identified in the differential expression analysis we will identify common upregulated and downregulated genes across all four datasets. We will do this only among the list of 15,252 protein-coding one-to-one orthologs between human and mice provided in the supplementary information of Wang et al. (2021).

Results:

Single-cell RNA-seq Data Analysis:

From the clustering of scRNA-seq data below (Fig. 1) we observe 15 different cell types in this data. We are specifically interested in the aHSCs and qHSCs and after running differential expression analysis between these two cell types we found 777 significantly upregulated genes and 689 significantly downregulated genes. However, the original article for this dataset observed more significantly upregulated and downregulated genes than we did in our analysis. We denoted genes with p-values < 0.05 as significant and the authors could have used a different cutoff, although they did not note explicitly what it was. Among the top 6 upregulated genes (Fig. 2), we note a well-known fibrosis markers (COL1A2), a gene involved in immune responses triggering inflammation (C3), and a gene involved in collagen organization (DCN) (NIH 2009). For all such genes being involved in the fibrotic process is feasible.
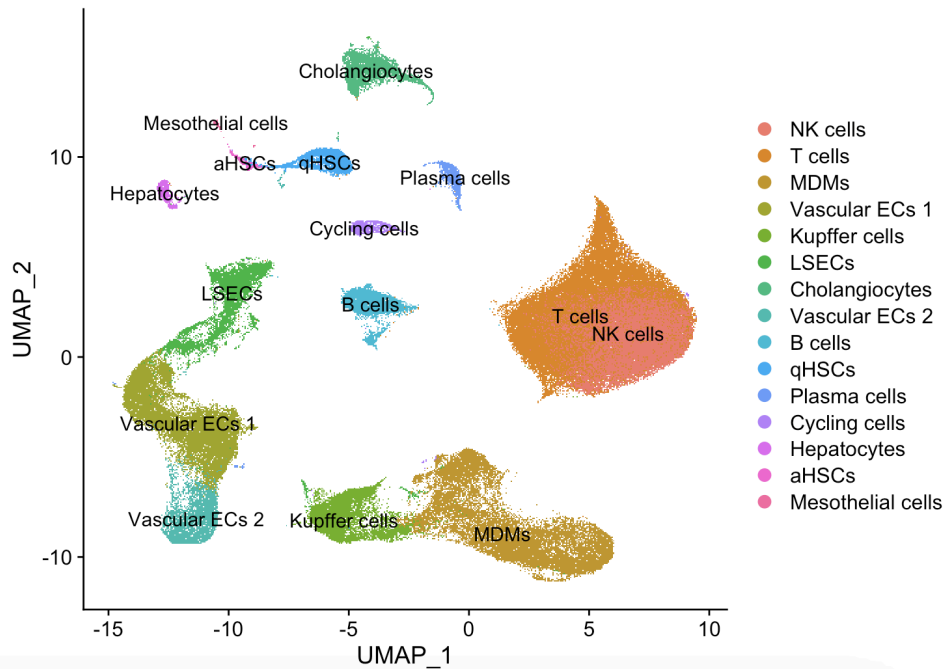
Fig. 1: UMAP Plot of Cell Type clustering in Single Cell RNA-seq dataset obtained from Wang et al.(2021).
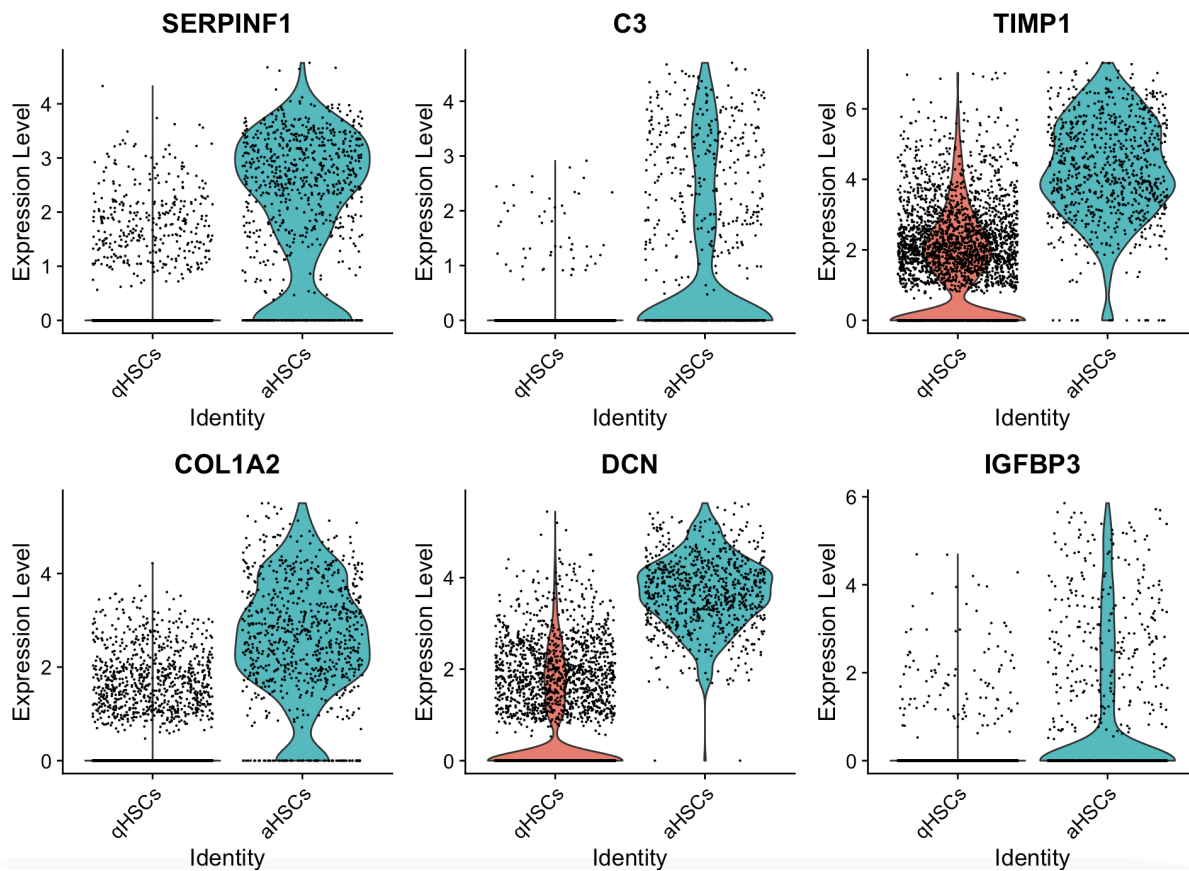

Fig. 2: Violin plots of the top 6 upregulated genes in aHSCs compared to qHSCs in the scRNA-seq data from Wang et al. (2021).
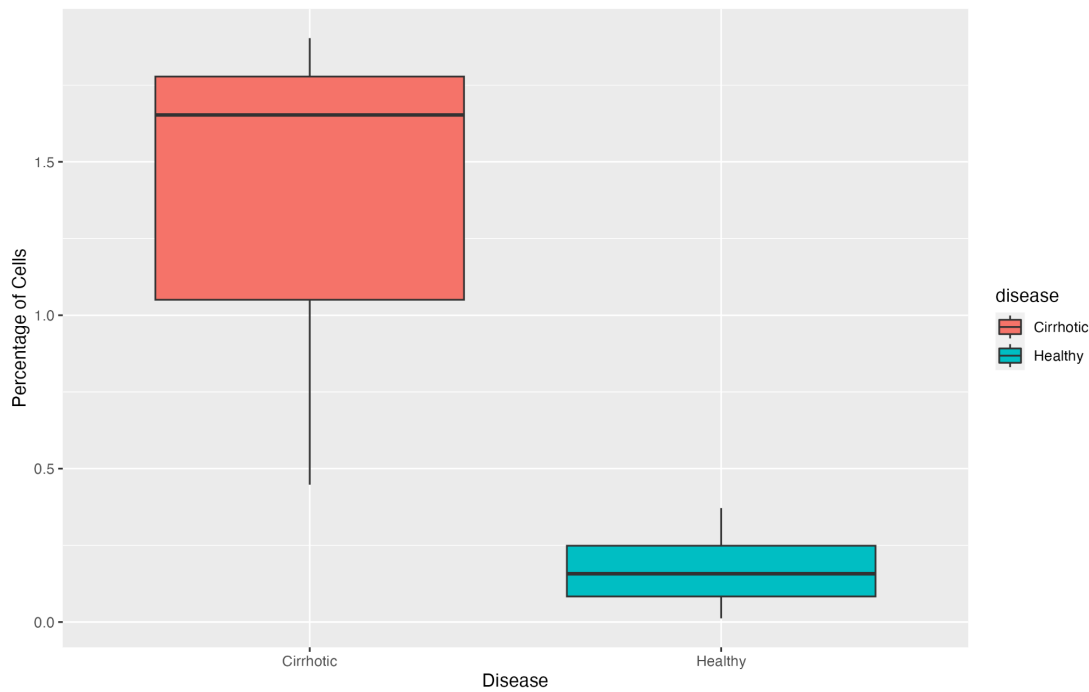
Fig. 3: Boxplot of the percentage of total cells which are aHSCs among the Cirrohitc and Healthy groups from our scRNA-seq data.

From the visualization above of the proportion of aHSCs across individuals (Fig. 3) we do indeed see that cirrhotic livers have a higher proportion of aHSCs compared to healthy individuals. This visualization did not include the healthy participant data generated by Wang et al. (2021) and only included the integrated data taken from Ramachandran et al. (2019). This choice was made for ease of visualization and direct comparison.

Human Bulk RNA Cell Type Composition:
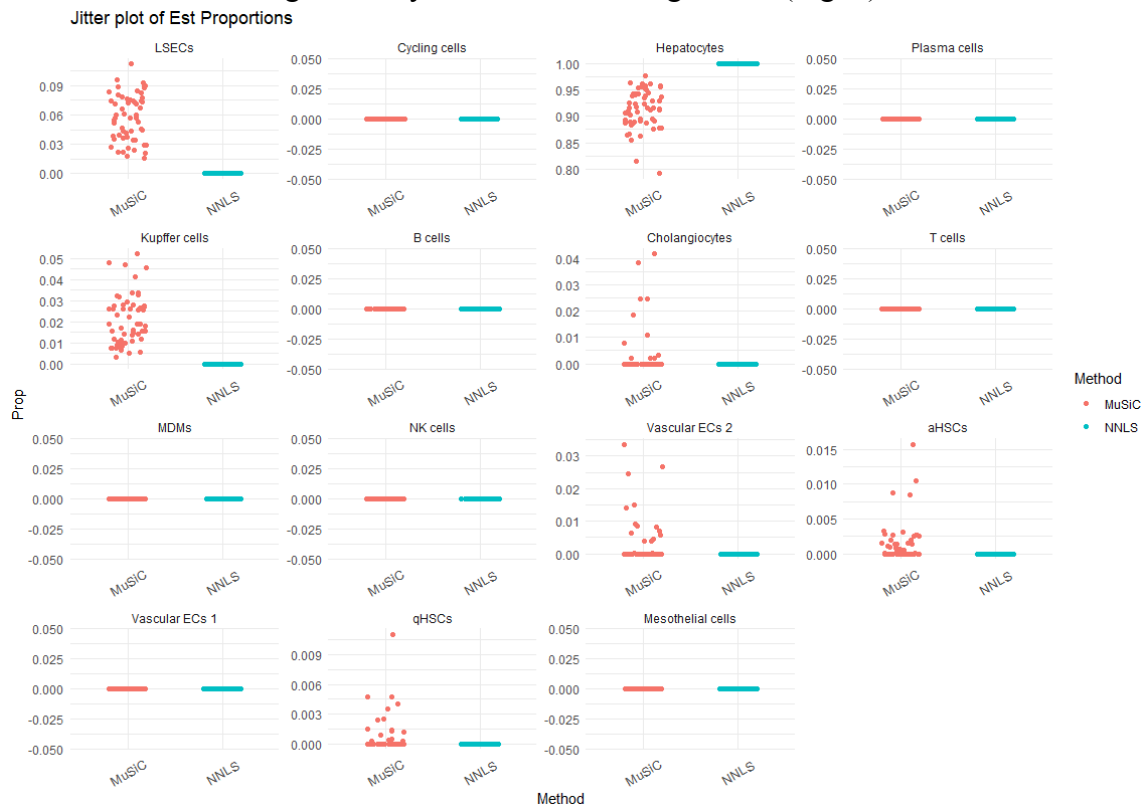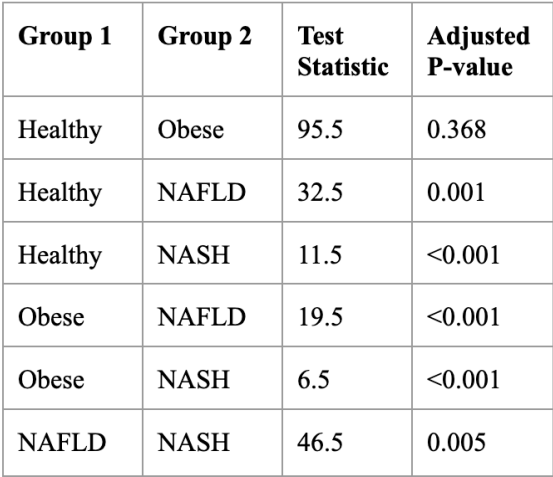Our deconvolution using MuSiC yielded the following results (Fig. 4).

Fig. 4: Results of cell type proportion estimation using MuSiC

We see some differences in the proportions of aHSCs and qHSCs. We also note that the distribution of the proportion of Kupffer cells and LSECs seems very widespread. It also seems that the majority of cells in these liver biopsies are Hepatocytes. MuSiC did not detect the presence of a number of cell types namely B cells, T cells, MDMs, NK cells, Vascular ECs 1, Mesothelial Cells, Cycling cells, and Plasma cells.

For the aHSCs, we filtered these and plotted the proportion in a boxplot based on the disease status of the individual (Fig. 5a). We see that there appears to be an increase in aHSCs as we go from healthy to NAFL to NASH. We conducted pairwise comparisons of these proportions using the Wilcoxon Rank Sum test and found that all differences were significant except between the Healthy and the Obese groups. Test statistics and p-values can be found in the table below (Fig. 5b). We saw that except between the Healthy and Obese groups all other pairwise comparisons had an adjusted p-value below 0.05 and the difference between groups in terms of the proportion of aHSC cells was statistically significant.

Group 1	Group 2	Test Statistic	Adjusted P-value
Healthy	Obese	95.5	0.368
Healthy	NAFLD	32.5	0.001
Healthy	NASH	11.5	<0.001
Obese	NAFLD	19.5	<0.001
Obese	NASH	6.5	<0.001
NAFLD	NASH	46.5	0.005

Fig. 5: a) On the left: This is a boxplot of aHSC cell percentages across different disease status groups in our human bulk RNA-seq dataset. b) Right: This is the result of a Wilcoxon Rank Sum test on aHSC proportions.

Group 1	Group 2	Test Statistic	Adjusted P-value
Healthy	Obese	47	0.072
Healthy	NAFL	176	0.003
Healthy	NASH	173	0.016
Obese	NAFL	19.5	0.002
Obese	NASH	6.5	0.002
NAFLD	NASH	46.5	0.379

Fig. 6: a) On left. This is a boxplot of Kuppfer cell percentages across different disease status groups in our human bulk RNA-seq dataset. b) On right: This is the result of a Wilcoxon Rank Sum test on Kuppfer cell proportions.

We conducted this same test within two other cell types. The first was Kuppfer cells (Fig. 6 above). We found that the proportion of Kuppfer cells was significantly different between the Healthy and NAFL/NASH disease groups and between Obese and the NAFL/NASH disease groups (Fig. 6b). Kuppfer cells are liver macrophages and play a role in maintaining liver function. This could make them a potential target in terms of addressing liver inflammation/fibrosis.

Secondly we examined LSECs. All pairwise comparisons except the NAFL and NASH comparison were found to be statistically significant (Fig. 6b). LSECs (liver sinusoidal endothelial cells) are a special type of endothelial cell with blood cells on one side and HSCs and hepatocytes on the other (Poisson, 2017).
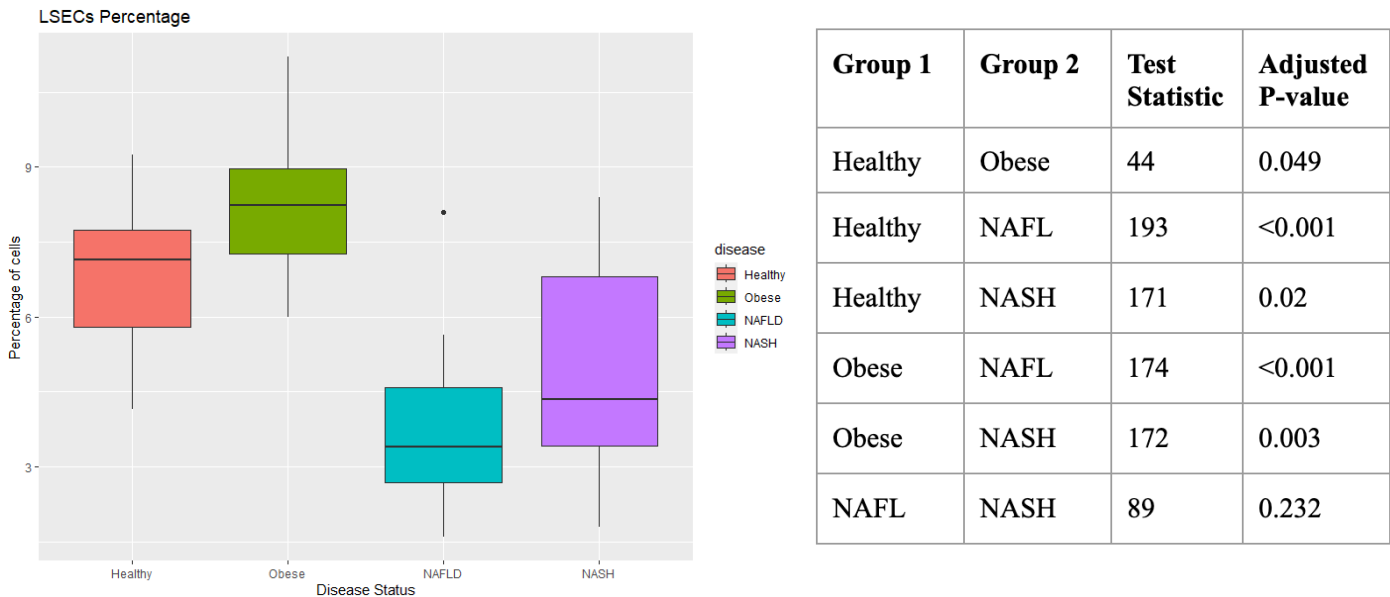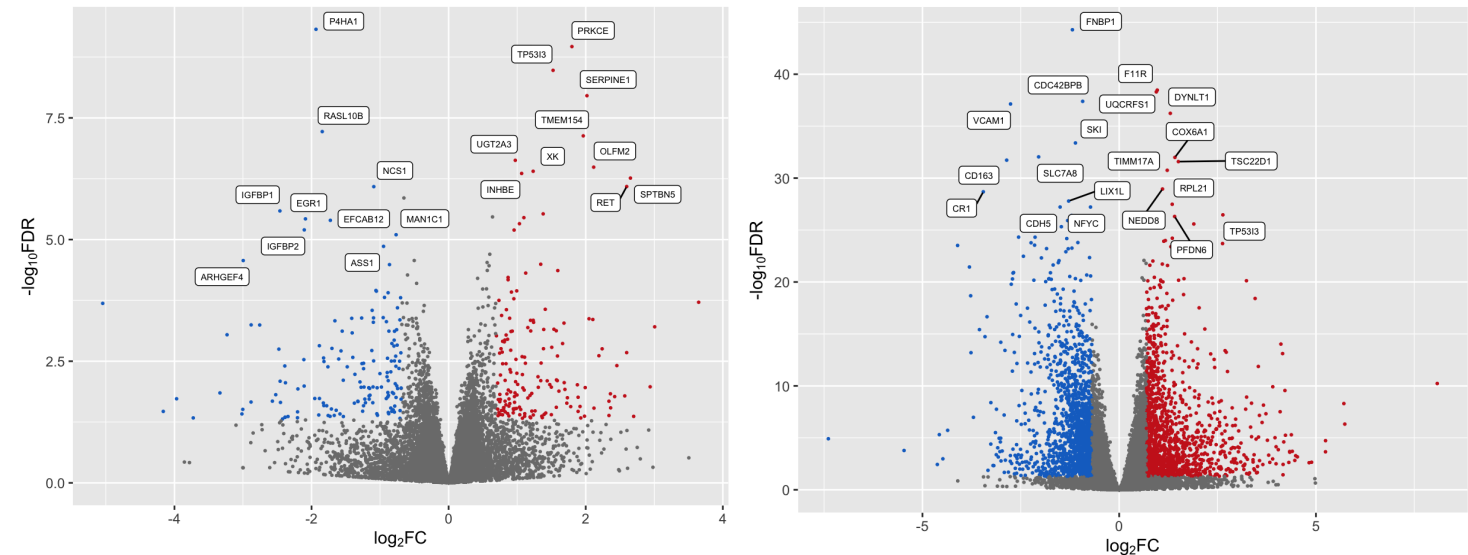

Group 1	Group 2	Test Statistic	Adjusted P-value
Healthy	Obese	44	0.049
Healthy	NAFL	193	<0.001
Healthy	NASH	171	0.02
Obese	NAFL	174	<0.001
Obese	NASH	172	0.003
NAFL	NASH	89	0.232

Fig. 6: a) On left. A boxplot of LSECs percentage across different diseaset status groups in bulk human RNA-seq data. b) On right. Results of a Wilcoxon Rank Sum test on Kuppfer cell proportions.

Differentially Expressed Gene Analysis through Bulk RNAseq

Figure 7 shows the result of differential gene expression analysis via DESeq2. Compared to healthy individuals, obese individuals had 160 upregulated genes and 146 downregulated genes. In NAFL patients, 1395 genes were upregulated and 1457 genes were down regulated. In addition, we found 2195 genes that were significantly upregulated in NASH patients, 1328 genes that were significantly downregulated in the patients. This follows intuition since as we get further from our "healthy" condition in terms of NAFLD we see more differentially expressed genes.
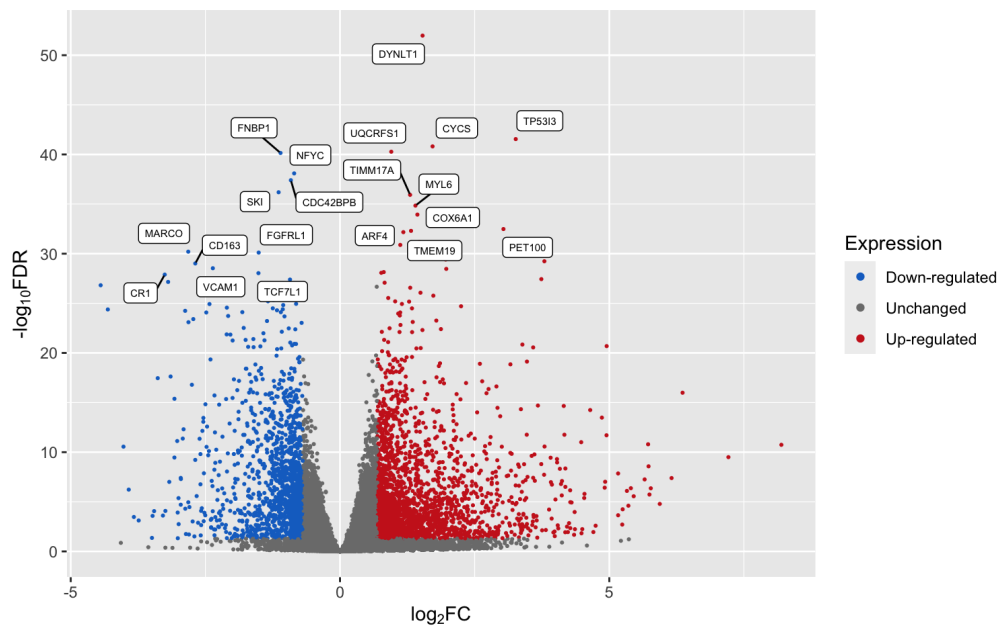
Fig 7. From the top left: differential expressed gene analysis of obese patients, that of NAFL patients, and that of NASH patients.

Then we analyzed differentially expressed genes in the three conditions by comparing genes in three conditions and by conducting gene ontology enrichment analysis. Looking at Fig 8., we can conclude that there are more similarities in differentially expressed genes between NAFL and NASH subjects compared to the similarities with obese subjects.
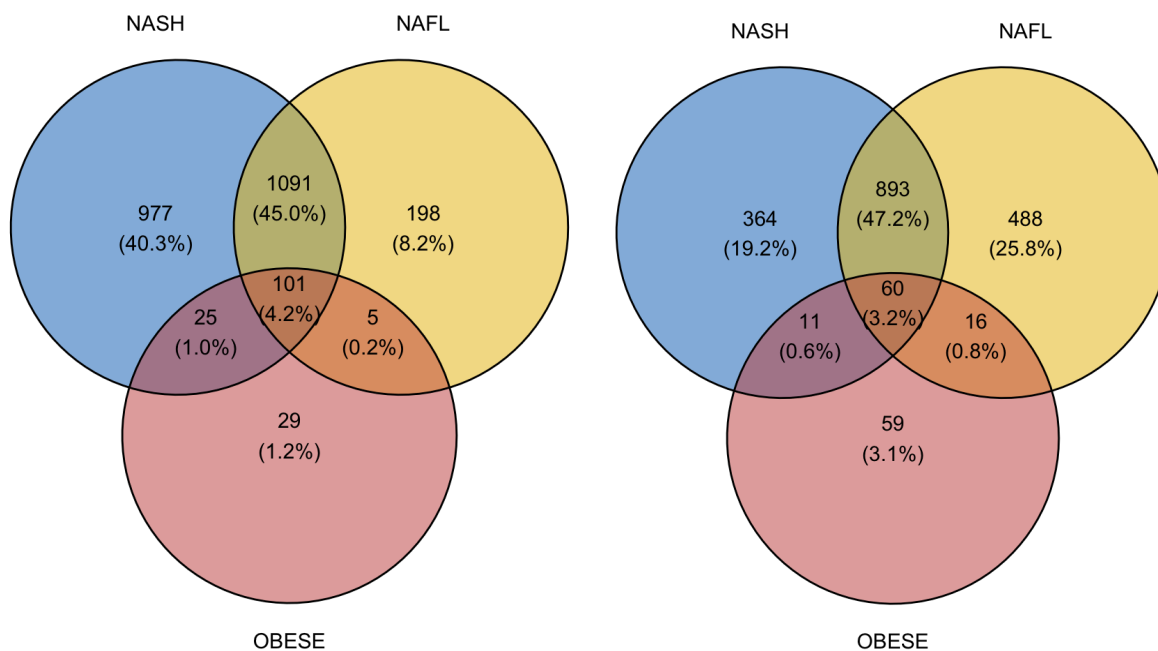

Fig 8. Left is the comparison of upregulated genes and right is the comparison of downregulated genes.

We also conducted Gene Ontology(GO) analysis to see what kind of genes were commonly expressed between NASH and NAFL. From Fig 9., commonly upregulated genes were those related to inflammation.
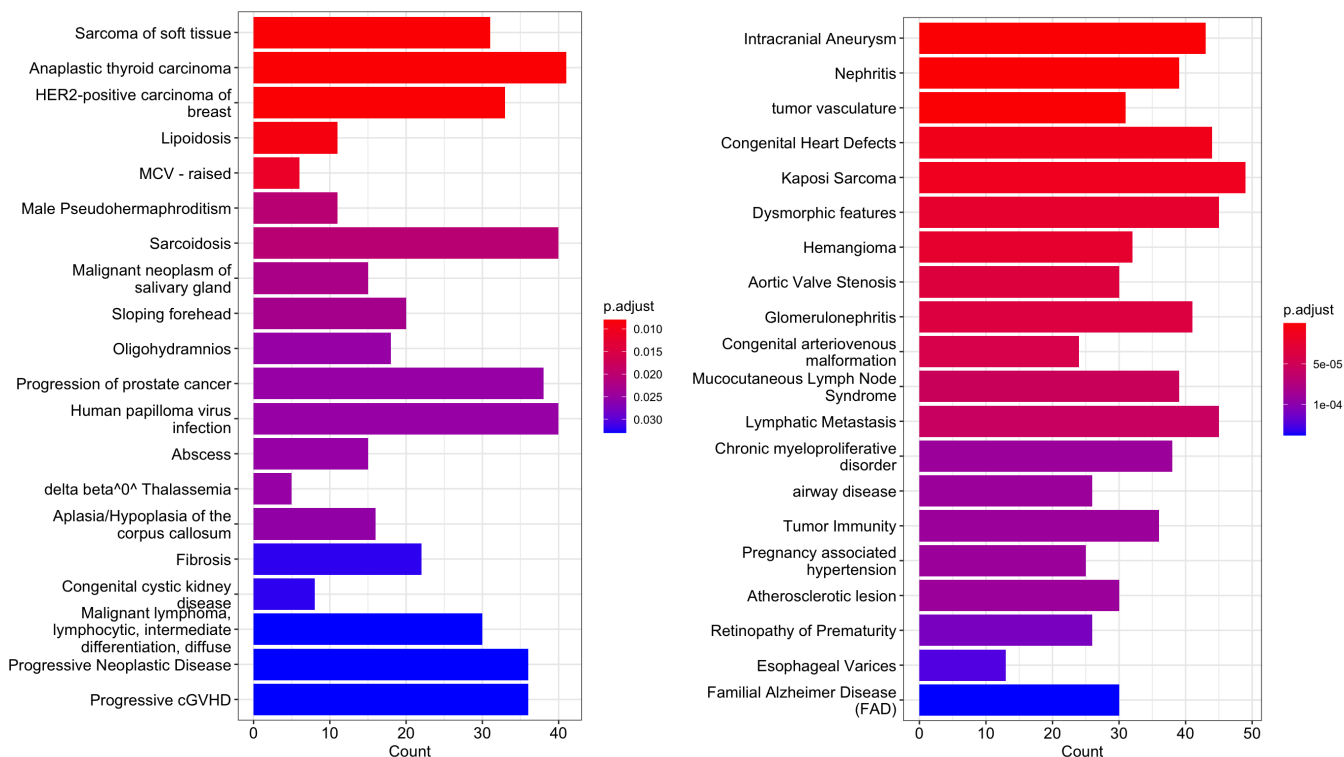
Fig 9. Left is the gene ontology analysis of upregulated genes and left is the gene ontology analysis of downregulated genes.

<u>Translatability of liver fibrosis-associated genes between human and mouse:</u>

For bulk RNA human and mouse data, we conducted an analysis via DESeq2 to find differentially expressed genes related to liver fibrosis. Below is the result for our analysis. In the mouse in vivo model, we identified 2682 up-regulated and 1751 down-regulated genes and in the mouse in vitro model, we identified 3813 up-regulated and 3862 down-regulated genes. Although our identified top 10 up- and down- regulated genes are not identical as the original paper, the pattern of the expression of the genes in both models are still the same. We suspect the reason for the difference in identified specific genes is after alignments between the transcript ID and gene names, there were 632 missing gene names and additional duplications among the aligned gene names which might be caused due to the difference in reference.
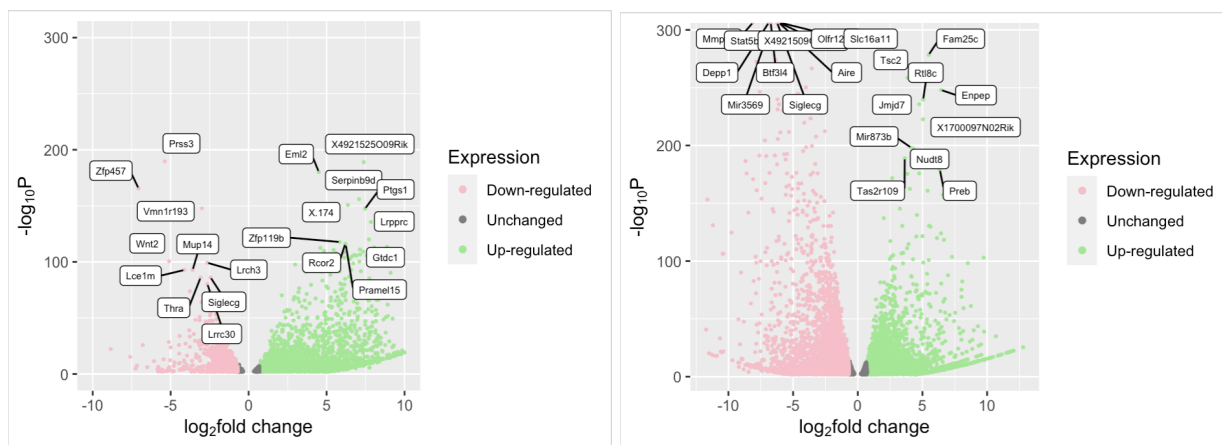

Fig 10. : a) On the left. Differential gene expression analysis between aHSCs and qHSCs (as reference) isolated from healthy and CCl$_4$ mouse livers. Only the top 10 up- or down-regulated genes are labeled with gene symbols. b) On the right. Differential gene expression analysis between aHSCs and qHSCs (as reference)Healthy liver-derived qHSCs were activated on plastic in vitro. Only the top 10 up- or down-regulated genes are labeled with gene symbols

From the results above, we transformed the name of mouse genes into human gene analog. Through the comparison of differentially expressed genes related to liver fibrosis, we got the results in Fig 11. We identified 7 significantly up-regulated and 3 down-regulated genes that behaved similarly in human HSCs in both single cell and bulk compared to mouse HSCs in both in vivo and in vitro models.
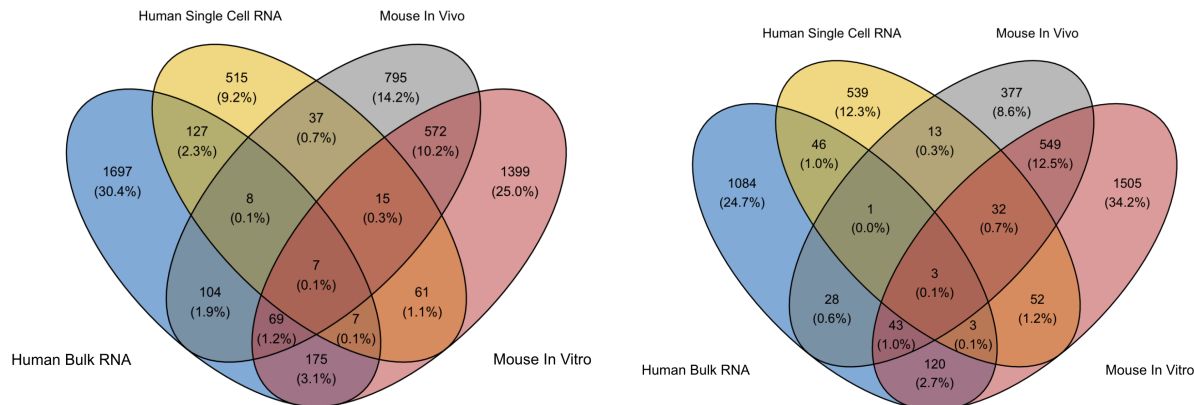

Fig. 11. Venn Diagrams representing the number and proportions of common genes across different datasets. The left venn diagram is the upregulated genes and the right graph indicates downregulated genes.

The seven commonly upregulated genes were *S100A13, PKDCC, PPIC, PLIN2, SPON1, NUCB2, KRT8*. *S100A13* encodes a calcium binding protein, and it is known to cause NASH and drive malignancy in liver cancer. *PKDCC* is a gene that encodes a protein that helps in regeneration of the liver (Delangre, 2022). *PPIC* encodes a protein that is involved in protein folding, and when the gene is knocked out, it shows to reduce the occurrence of liver fibrosis (Yang, 2021). *PLIN2* is a gene encoding perilipin, a protein that constitutes intracellular lipid storage. As it is related to storing lipids in a cell, it plays a major role in the progression of NASH (Najt, 2015). *SPON1*, a gene encoding a protein that constitutes an extracellular matrix is one of the well known genes that is upregulated in NASH(Gao, 2022). *NUCB2* is a gene that encodes a protein that induces the release of tumor necrosis factor(TNF) from vascular endothelial cells. As an inflammatory cytokine, TNF is known to cause liver fibrosis by causing sustained liver inflammation (Yang, 2015). *KRT8* is a gene encoding for protein keratin 8, which is one of the major keratins forming Mallory bodies. Mallory bodies are usually found in liver cells of liver disease patients. Especially since overexpression of keratin 8 over keratin 18 results in the formation of Mallory bodies (Kurtovic, 2022), upregulation of *KRT8* in NASH patients and mouse models is consistent with the biological knowledge. In conclusion, we can see that genes that were commonly upregulated were either genes that were involved in the progression of liver disease, especially NASH and fibrosis, and repairing damaged liver.

On the other hand, the three commonly downregulated genes were EFHD1, MT1X, SOCS2. EFHD1 encodes calcium binding protein, however its relation with liver disease is still unknown. MT1X is a gene that encodes a protein involved in the oxidative stress pathway. This pathway is crucial for cells in that it protects the cell from damage by reactive oxidative species (Cichoż-Lach, 2014).Finally, SOCS2 encodes a protein that inhibits dysregulation of cell function, especially HSC. From the analysis of the downregulated genes, we can conclude that they are significantly downregulated in NASH genes that contribute to homeostasis or repairing damaged cell conditions. These analysis results match with biological knowledge and thus it shows that these genes could be a potential target for diagnosis, prognosis, and treatment of liver disease.

Discussion:

Our analysis of cell type composition change in the bulk human RNA holds potential for future research. We found that Kuppfer cells and LSECs also have statistically significant changes in proportions across disease status. It seems plausible at a glance that these two cell types would play some role in the progression of NASH and liver fibrosis. However, we are not subject matter experts and having such experts take a look at our findings to see if this is clinically/scientifically relevant would be a good next step.

While we conducted a brief overview of our commonly upregulated and downregulated genes, further biological research on how these genes contribute to the progression of NASH seems to be required. Also, consulting with experts in this field would be beneficial in evaluating the true plausibility of these genes as therapeutic targets.

Another direction for future research would be to examine more publicly available NASH or fibrosis related RNA-seq datasets and incorporate it into this comparison or conduct alternate comparisons to create a larger gene pool. We could also loosen some of our significance requirements or use alternate differential expression analysis software and see how the common genes change.

References:

Dong, S., Chen, Q. L., Song, Y. N., Sun, Y., Wei, B., Li, X. Y., Hu, Y. Y., Liu, P., & Su, S. B. (2016). Mechanisms of CCl4-induced liver fibrosis with combined transcriptomic and proteomic analysis. *The Journal of toxicological sciences*, *41*(4), 561–572. https://doi.org/10.2131/jts.41.561

Marcher, AB., Bendixen, S.M., Terkelsen, M.K. et al. Transcriptional regulation of Hepatic Stellate Cell activation in NASH. Sci Rep 9, 2324 (2019). https://doi.org/10.1038/s41598-019-39112-6

Poisson, Johanne et al. "Liver sinusoidal endothelial cells: Physiology and role in liver diseases." Journal of hepatology vol. 66,1 (2017): 212-227. doi:10.1016/j.jhep.2016.07.009

Suppli, M. P., Rigbolt, K. T. G., Veidal, S. S., Heebøll, S., Eriksen, P. L., Demant, M., Bagger, J. I., Nielsen, J. C., Oró, D., Thrane, S. W., Lund, A., Strandberg, C., Kønig, M. J., Vilsbøll, T., Vrang, N., Thomsen, K. L., Grønbæk, H., Jelsing, J., Hansen, H. H., & Knop, F. K. (2019). Hepatic transcriptome signatures in patients with varying degrees of nonalcoholic fatty liver disease compared with healthy normal-weight individuals. *American journal of physiology. Gastrointestinal and liver physiology*, *316*(4), G462–G472. https://doi.org/10.1152/ajpgi.00358.2018

U.S. Department of Health and Human Services. (2021, April). *Definition & Facts of NAFLD & Nash - NIDDK*. National Institute of Diabetes and Digestive and Kidney Diseases. Retrieved April 26, 2023, from https://www.niddk.nih.gov/health-information/liver-disease/nafld-nash/definition-facts#:~:text=Nonalcoholic%20fatty%20liver%20disease%20(NAFLD)%20is%20a%20condition%20in%20which,called%20alcohol%2Dassociated%20liver%20disease

Wang, ZY., Keogh, A., Waldt, A. et al. Single-cell and bulk transcriptomics of the liver reveals potential targets of NASH with fibrosis. Sci Rep 11, 19396 (2021). https://doi.org/10.1038/s41598-021-98806-y

Younossi, Zobair M et al. "The global epidemiology of nonalcoholic fatty liver disease (NAFLD) and nonalcoholic steatohepatitis (NASH): a systematic review." Hepatology (Baltimore, Md.) vol. 77,4 (2023): 1335-1347. doi:10.1097/HEP.0000000000000004

Kurtovic E, Limaiem F. Mallory Bodies. [Updated 2022 Sep 12]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2023 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK545300/

Yang, Y. M., & Seki, E. (2015). TNFα in liver fibrosis. *Current pathobiology reports*, *3*(4), 253–261. https://doi.org/10.1007/s40139-015-0093-z

Gao, R., Wang, J., He, X., Wang, T., Zhou, L., Ren, Z., Yang, J., Xiang, X., Wen, S., Yu, Z., Ai, H., Wang, Y., Liang, H., Li, S., Lu, Y., Zhu, Y., Shi, G., & Chen, Y. (2022). Comprehensive analysis of endoplasmic reticulum-related and secretome gene expression profiles in the progression of non-alcoholic fatty liver disease. *Frontiers in endocrinology*, *13*, 967016. https://doi.org/10.3389/fendo.2022.967016

Delangre, E., Oppliger, E., Berkcan, S., Gjorgjieva, M., Correia de Sousa, M., & Foti, M. (2022). S100 Proteins in Fatty Liver Disease and Hepatocellular Carcinoma. *International journal of molecular sciences*, *23*(19), 11030. https://doi.org/10.3390/ijms231911030

Yang, X., Shu, B., Zhou, Y., Li, Z., & He, C. (2021). Ppic modulates CCl4-induced liver fibrosis and TGF-β-caused mouse hepatic stellate cell activation and regulated by miR-137-3p. *Toxicology letters*, *350*, 52–61. https://doi.org/10.1016/j.toxlet.2021.06.021

Najt, C. P., Senthivinayagam, S., Aljazi, M. B., Fader, K. A., Olenic, S. D., Brock, J. R., Lydic, T. A., Jones, A. D., & Atshaves, B. P. (2016). Liver-specific loss of Perilipin 2 alleviates diet-induced hepatic steatosis, inflammation, and fibrosis. *American journal of physiology. Gastrointestinal and liver physiology*, *310*(9), G726–G738. https://doi.org/10.1152/ajpgi.00436.2015

Cichoż-Lach, H., & Michalak, A. (2014). Oxidative stress as a crucial factor in liver diseases. *World journal of gastroenterology*, *20*(25), 8082–8091. https://doi.org/10.3748/wjg.v20.i25.8082