Gene Expression Analysis of Non-alcoholic Fatty Liver Disease in Diabetes Patients

Lynn Nguyen, Sarah Supp

Denison University, Granville, OH, USA

Abstract

Since non-alcoholic fatty liver disease (NAFLD) has been studied for more than two decades with many different methods and protocols on different patient cohorts, there are gaps in analysis between studies in the early years and more recent stages. This study aims to analyze two NAFLD datasets (GSE89632 (Arendt et al., 2015) and GSE163211 (Subudhi et al., 2022)) to find gene expression patterns of diabetes patients that are involved in stages of NAFLD, steatosis and nonalcoholic steatohepatitis. The identification of differentially expressed genes was performed using the DESeq2 package and molecular function pathways were discovered with GO analysis. More than 150 differentially expressed genes were found, along with molecular pathways associated with them, such as DNA binding, miRNA-binding transcription factors, serine hydrolase activity, and insulin-like growth factors, which some of them were also found in initial studies and some were novel. These pathways also verified to be consistent with independent studies in their roles with disease progression. These data provide insights and biomarkers into the dynamic pathogenesis of NAFLD regarding high-risk individuals with inclusive data and analysis.

Introduction

Non-alcoholic fatty liver disease (NAFLD) is a heavy burden on global health concerns (H. Ye et al., 2023). It is possible to have NAFLD without any specific symptoms, and the diagnosis of NAFLD is typically because of liver biopsy (LaBrecque et al., 2014). After diagnosis, the treatment is usually based on managing lifestyle and weight loss, and to our best understanding, no drug for NAFLD treatment is approved by the American Agency for Food and Drug Administration (FDA). Besides, it is also imperative to have noninvasive treatments that would not disrupt the natural fibrosis progression which helps prevent liver-related mortality risk (Friedman et al., 2018).

Since it is one of the most frequent causes of abnormal liver dysfunction because of excess fat, it was found that gene profiling that is associated with NAFLD stages is different compared to the normal or non-diabetic liver (Takamura et al., 2004). Even though obesity is an independent risk factor for NAFLD (Fan et al., 2018), the mechanism has not been fully understood since not all obese patients develop NAFLD in the early stage (simple steatosis) and only a subset of these develop into more serious stage, Nonalcoholic Steatohepatitis (NASH) (Bellentani et al., 2010).

In the process of understanding the mechanism of NAFLD progression for the development of a better diagnostic and/or treatment for patients, gene expression profiling and phenotypes is an increasingly recognized method for studying the development of NAFLD (Adams et al., 2013; Subudhi et al., 2022). However, many studies have investigated the approach of examining gene expression in liver NAFLD patients before, they have different methods of collecting data and analysis that developed over time and on different patient cohorts, along with different treatments

(Brunt et al., 2015; Lallukka et al., 2013; L. Li et al., 2016). With this, there is a possibility of having gaps in the findings of these studies compared to each other, especially in this era of technological development rapidly over time.

The overarching research question guiding this study is to provide a meta-analysis to identify gene expression profiling and molecular function pathways associated with the progression of nonalcoholic fatty liver disease (NAFLD), along with the clinical characteristics of diabetes patients from two US databases in 2015 and 2022. I hypothesize that with a different analysis approach on these databases, it is potentially that there will be similar and different findings regarding differentially expressed genes and pathways. The ultimate aim of this meta-analysis study is to identify signatures of NAFLD stages associated with diabetes and involved in each stage during the progression, as well as propose target genes for treatments and improve our understanding of this disease among different studies.

Data and Method

In this study, I performed a meta-analysis of microarray experiments and RNA sequencing on liver tissue of NAFLD diabetes patients on two datasets from the US that were held and distributed on Gene Expression Omnibus (GEO) (Barrett et al., 2013), with accession numbers GSE89632 (Arendt et al., 2015) and GSE163211 (Subudhi et al., 2022) on the NCBI GEO website (Barrett et al., 2013). These two datasets have information on gene expression values, phenotype characteristics, and metadata of patients, which makes them sufficient for the meta-analysis. There were 63 patients in the dataset GSE89632 (Arendt et al., 2015), and 318 patients in

GSE163211 (Subudhi et al., 2022). For GSE89632 (Arendt et al., 2015), the study was approved by the local Research Ethics Board, was registered (NCT02148471, www.clinicaltrials.gov), and patients were recruited between March 2007 and November 2011 and the study followed the guidelines of the 1975 Declaration of Helsinki and its revisions. For GSE163211 (Subudhi et al., 2022), patients were enrolled between December 2010 and 2016 under the Massachusetts General Hospital (MGH) NAFLD clinic with written consents from participants, and there was no clear ethics registration indicated in the report.

From these datasets, I analyzed gene expression values (microarray data values), patients' diabetes situation (have diabetes or do not have diabetes), and their NAFLD stages (with the range from less serve to more serve: healthy control, SS (Simple Steatosis) or NASH (Nonalcoholic Steatohepatitis)). Every dataset is analyzed separately to respect their experiment design and patient cohorts. Data is retrieved using the GEOquery package in R (Davis & Meltzer, 2007).

I excluded missing values in these variables and those that did not fail within the listed categories. Gene expression values had been normalized and quality controlled by the authors of these original publications. For GSE89632, they used GeneSpring v12.5 (Agilent Technologies) and for GSE163211, they normalized with NanoString (Bhattacharya et al., 2020) to housekeeping genes while performing a quantitative reverse transcription-polymerase chain reaction. With this, it is required to back-transform the first dataset to approach variance between samples in the analysis to achieve the whole-number raw data for a more detailed analysis. After preprocessing, I spanned these variables into a matrix and used the DESeq2 package

(Love et al., 2014) in R (R version 4.3.0) between gene expression values with NAFLD stages and diabetes status as conditions. These differentially expressed transcripts have a threshold False Discovery Rate of 5% and the distance of Wald test (log_2Fold Change > 2 or < -2), which is the validation of their performance. After that, differentially expressed transcripts were encoded and mapped to gene symbols using illuminaHumanv4ACCNUM (Dunning et al., 2015), and the molecular function pathway was identified using the enrichGO (Yu et al., 2012).

Statistical analysis was performed using DESeq2 (Love et al., 2014) with normalization of raw reads using size factor normalization. The False Discovery Rate of 5% was controlled using the Benjamini-Hochberg test with p-values. Dispersal (variability) of input transcripts were examined using the empirical Bayes approach and the threshold for determining differentially expressed genes was based on the Wald tests for specifying contrasts between design conditions (in this study, conditions (or treatment) would be patients in stages of NAFLD and the control would be healthy patients).

All analyses were performed using R version 4.3.0 (R Core Team, 2023), and data were retrieved using the GEOquery package (Davis & Meltzer, 2007). The analysis and data are on the GitHub repository NAFLD_genes (Nguyen, 2024) and accessibility can be provided per request.

Results

1. Clinical characteristics

Two datasets did not have missing data in gene expression assay; however, since some patients were missing proper classification for this meta-analysis they should have either failed into the category of “have diabetes” or “no diabetes”. With that, I lowered the number of patients included and the input of analysis has 29 patients with 29377 transcripts for GSE89632 and 318 patients with 800 transcripts for GSE163211 which satisfied our requirements for further analysis (Table 1).

		GSE89632 (29377 transcripts)	**GSE163211** (800 transcripts)
Diabetes (number of patients)			
	Yes	4	93
	No	25	225
NAFLD stages (number of patients)			
	Healthy	5	76
	Simple Steatosis	12	88
	Nonalcoholic Steatohepatitis (NASH)	12	154

Table 1: Summarize the number of patients with diabetes status and NAFLD stages in two datasets after the pre-processing step.

2. Differentially expressed genes analysis

There were three sub-analyses based on conditions (stages) for these two datasets: Healthy versus Simple Steatosis, Healthy versus Nonalcoholic steatohepatitis, and Simple Steatosis versus Nonalcoholic steatohepatitis. With each dataset, I examined differences in gene expression of healthy liver and stages of NAFLD of patients included.

A. GSE89632

With the threshold of False Discovery Rate of 5% (p-value < 0.05) and threshold generated based on Wald tests of greater than 2 for upregulated transcripts or less than -2 for downregulated transcripts ($\log_2$FoldChange > 2 or $\log_2$FoldChange < 2), comparisons between treatment and control groups were conducted. Among Healthy and Simple Steatosis patients, there were 155 upregulated transcripts and 10 downregulated transcripts (Fig. 1A). With this, 143 identified genes are differentially expressed compared to other genes in this category (Table A.1 and Fig. 2A). Comparing differentially expressed genes between Healthy and NASH patients, there were 16 upregulated transcripts and 7 downregulated transcripts (Fig. 1B). From this, 21 associated genes were identified (Table A.2 and Fig. 2B). Besides, there were 2 differentially expressed transcripts between Simple Steatosis and NASH (Fig. 1C), with 1 upregulated and 1 downregulated genes. In addition, the comparison between Healthy and stages of NAFLD (Simple Steatosis and NASH) shows that 19 shared significant genes that clearly distinguished healthy patients from patients with NAFLD in any stage.

Figure 1A

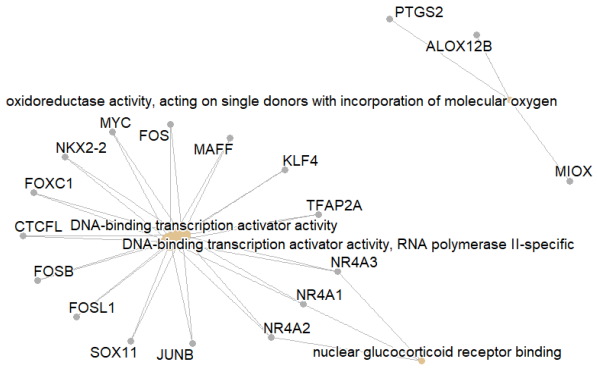

Figure 2A

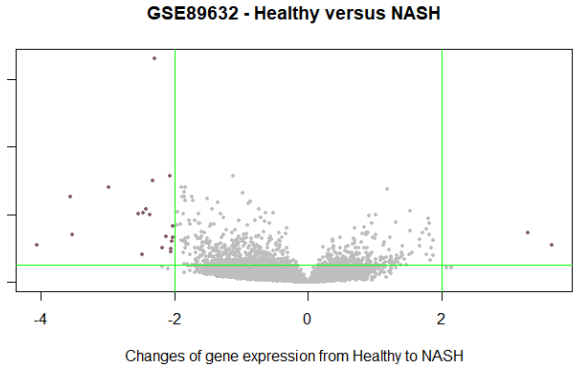

Figure 1B

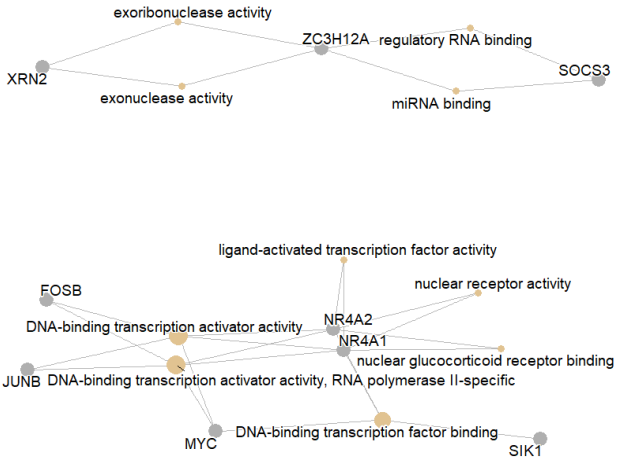

Figure 2B

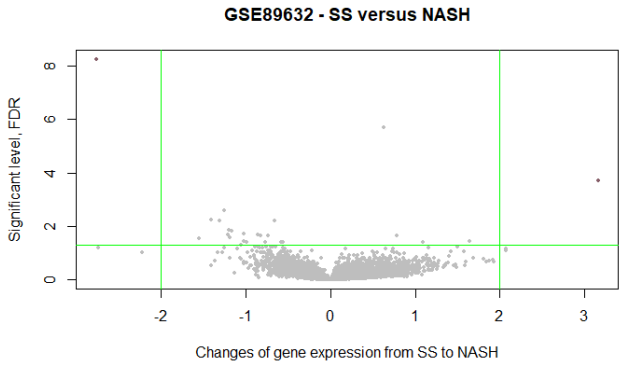

Figure 1C

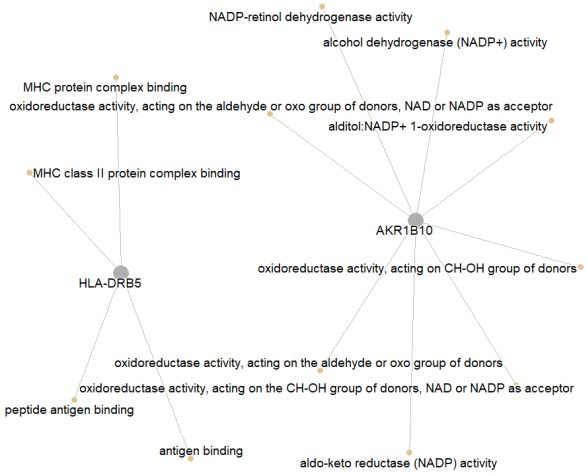

Figure 2C

Figure 1 (1A, 1B, 1C): Log_2FoldChange of transcripts expression and False Discovery Rate (FDR) in GSE89632. We examined Healthy versus Simple Steatosis (1a), Healthy versus Nonalcoholic Steatohepatitis (1b), and Simple Steatosis versus Nonalcoholic Steatohepatitis (1c). Dark purple dots are differentially expressed transcripts with an FDR of 5%.

Figure 2 (2A, 2B, 2C): Network plot shows expressed genes and their clusters of participated molecular pathways in GSE89632. Yellow dots are pathways and gray dots are genes, and the size of gray dots shows the number of genes that contributed to its linked pathway.

DNA-binding related activity pathways were the most dominant molecular functional pathways of shared differentially expressed genes in both Healthy versus Simple Steatosis and Healthy versus Nonalcoholic steatohepatitis (Fig. 3, p-value < 0.05). Besides that, more pathways associated with differentially expressed genes in each comparison were recorded. In Simple Steatosis patients compared to healthy, we observed oxidoreductase activity and nuclear receptor binding (p-value < 0.001, Fig. 2A). Similar to that, dominant molecular pathways found between NASH patients and healthy were RNA-binding related, ligand-activated transcription and exonuclease activity (p-value < 0.05, Fig. 2B). However, with genes shared among SS and NASH, since there were only 2 shared genes which are not found in previous comparisons of stages, each gene associated with many different pathways such as MHC protein complex binding, oxidoreductase, or even NADP-related pathways (p-value < 0.05, Fig. 2C and Table A.3). Overall, these genes and their pathways are generally related to

pathways that stem from transcription regular activities and receptor binding (p-value < 0.05, Fig. 4). Still, there are also variances in pathways and genes (Benjamini-Hochberg test, p-value < 0.01).

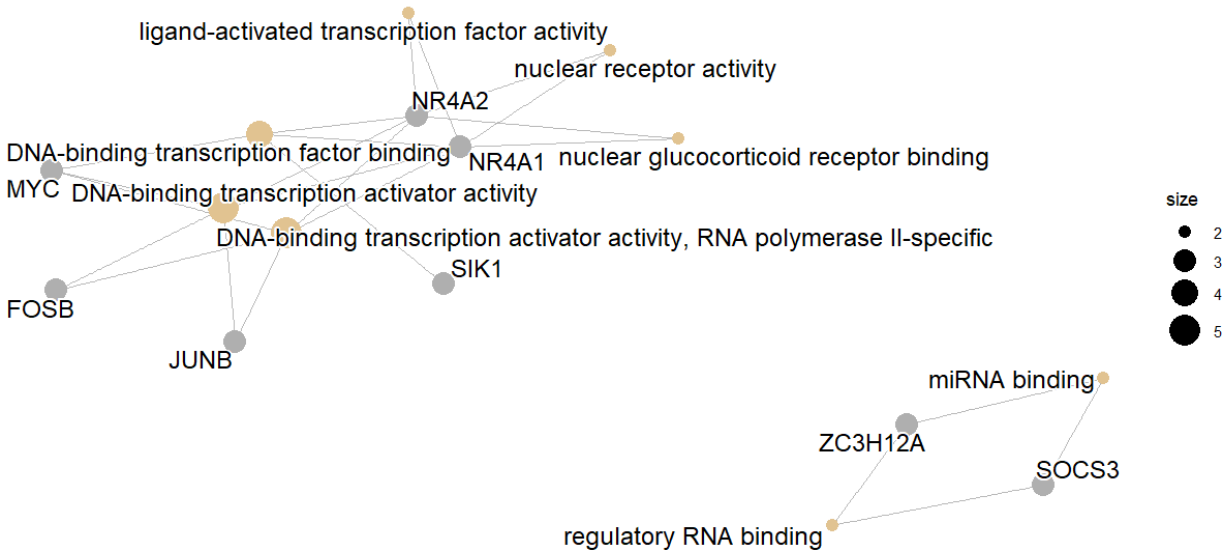

Figure 3: Shared differentially expressed genes and their clusters of participating in the same molecular pathways. Yellow dots are pathways and gray dots are genes, and the size of gray dots shows the number of genes that contributed to its linked pathway.

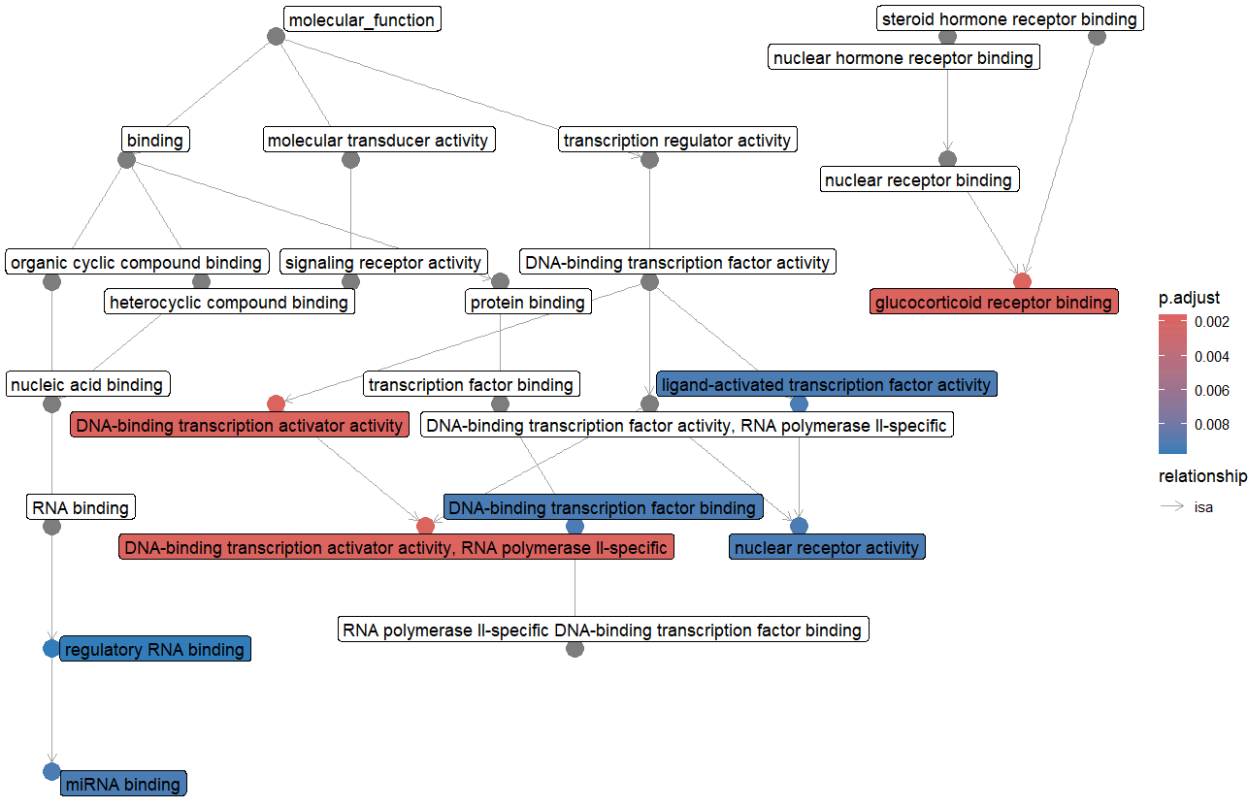

Figure 4: Network of molecular function pathways of genes that can lead to both NAFLD stages (SS and NASH) in patients. Colored pathways are pathways found in our differentially expressed genes with different colors for different p-values, and non-colored are pathways that might lead to these pathways but are not included in the findings.

B. GSE163211

Using a similar threshold category to the previous dataset with the False Discovery Rate of 5%, but I utilize Wald test-based threshold to the range of -0.5 to 0.5 ($\log_2$FoldChange > 0.5 or $\log_2$FoldChange < 0.5) since the GSE163211 consists of genes expressions (not microarray expressions as the previous study), there are similarities and differences between genes found among the two cohorts of patients

(Healthy versus Simple Steatosis and Healthy versus Nonalcoholic Steatohepatitis). Comparing Healthy and Simple Steatosis patients, there were 7 downregulated genes that are differentially expressed and no upregulated genes (p-value <0.05, Fig. 5A). Comparing Healthy and NASH patients, 10 upregulated genes and 1 downregulated gene changed their expression levels when changing conditions (p-value <0.05, Fig. 6A). Regarding the comparisons between SS and NASH, 10 upregulated genes and 1 downregulated genes were observed (p-value <0.05, Fig. 6A).

Figure 5A

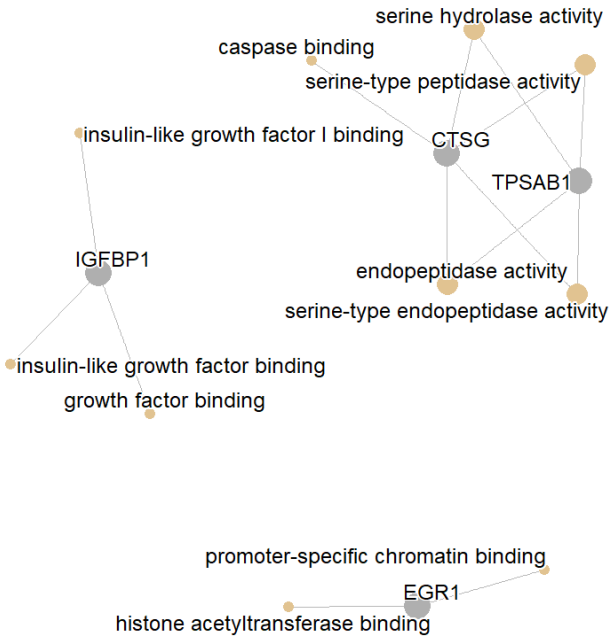

Figure 6A

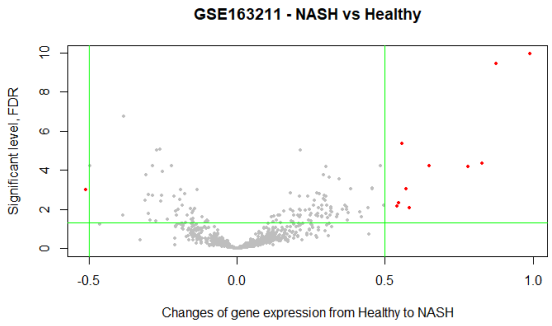

Figure 5B

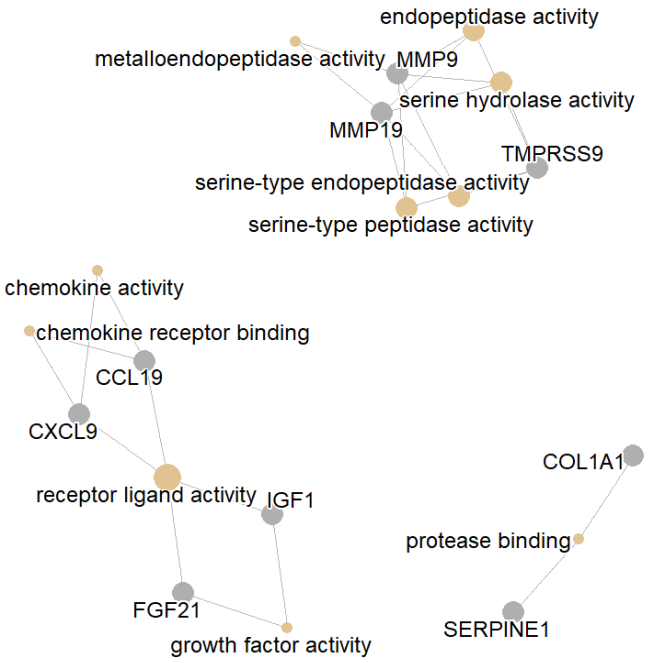

Figure 6B

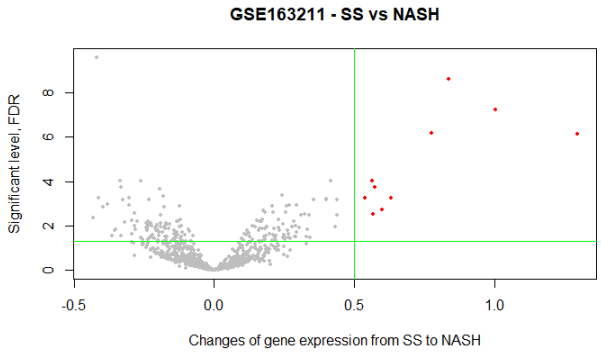

Figure 5C

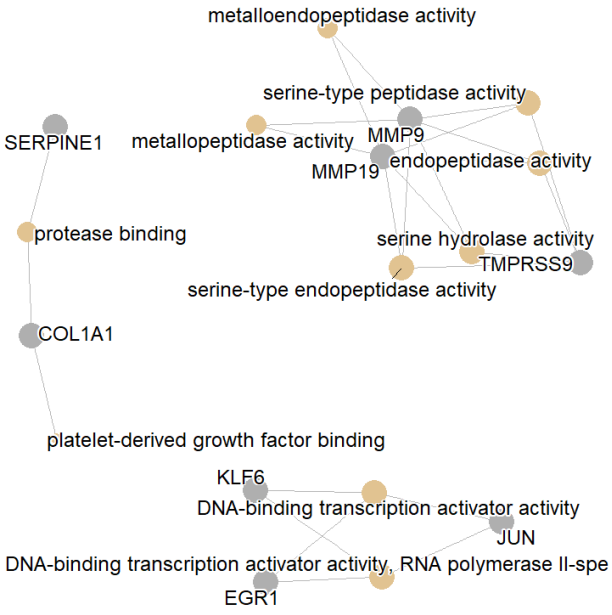

Figure 6C

Figure 5 (5A, 5B, 5C): Log_2FoldChange of transcripts expression and False Discovery Rate (FDR) in GSE163211. We examined Healthy versus Simple Steatosis (5A), Healthy versus Nonalcoholic Steatohepatitis (5B), and Simple Steatosis versus Nonalcoholic Steatohepatitis (5C). Dark purple dots are differentially expressed transcripts with an FDR of 5%.

Figure 6 (6A, 6B, 6C): Network plot shows expressed genes and their clusters of participated molecular pathways in GSE163211. Yellow dots are pathways and gray dots are genes, and the size of gray dots shows the number of genes that contributed to its linked pathway. Similar to previous plots, Healthy versus Simple Steatosis (6A), Healthy versus Nonalcoholic Steatohepatitis (6B), and Simple Steatosis versus Nonalcoholic Steatohepatitis (6C).

In this dataset, there are no shared differentially expressed genes between the comparison of Healthy and Simple Steatosis versus Healthy and Nonalcoholic steatohepatitis patients. However, there are some similar molecular function pathways among these comparisons, such as serine hydrolase activity and insulin-like growth factor, as well as other growth factor binding pathways (Fig. 6p-value < 0.05). Other pathways associated with differentially expressed genes in each comparison were recorded. In Simple Steatosis patients compared to Healthy, promoter-specific chromatin binding and endopeptidase activity (p-value < 0.05, Fig. 6A) were recognized. Some molecular pathways found between NASH patients and Healthy were chemokine activity and protease binding (p-value < 0.05, Fig. 6B). With genes and pathways shared among SS and NASH, a wide range of pathways from protease

binding to DNA-binding activities and endopeptidase activities (p-value < 0.05, Fig. 6C and Table A.3) were recorded. Overall, these genes and their pathways are generally related to pathways that stem from transcription regular activities, protein binding and catalytic activity (p-value < 0.05, Fig. 4). Still, there are also variances in pathways and genes (Benjamini-Hochberg test, p-value < 0.01).

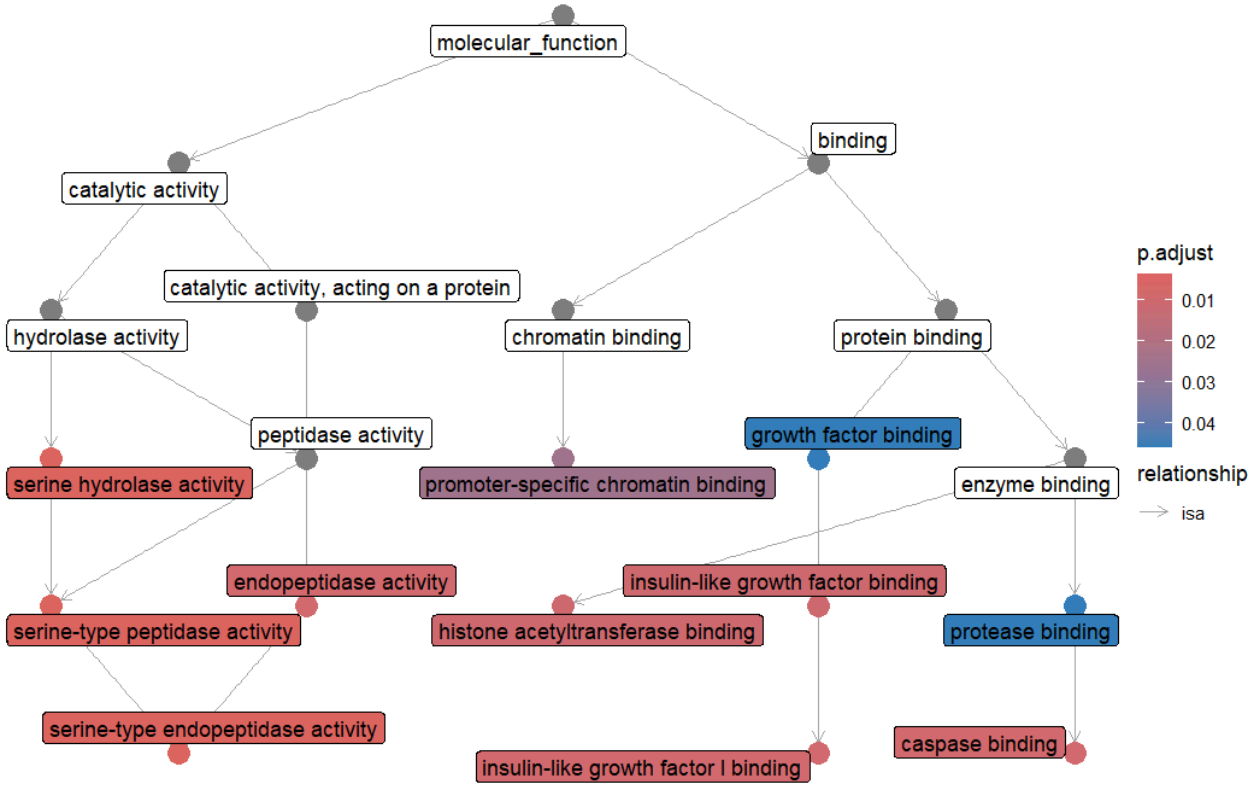

Figure 7: Network of molecular function pathways of genes that lead to Simple Steatosis in GSE163211. Colored pathways are pathways found in our differentially expressed genes with different colors for different p-values, and non-colored are pathways that might lead to these pathways but are not included in the findings.

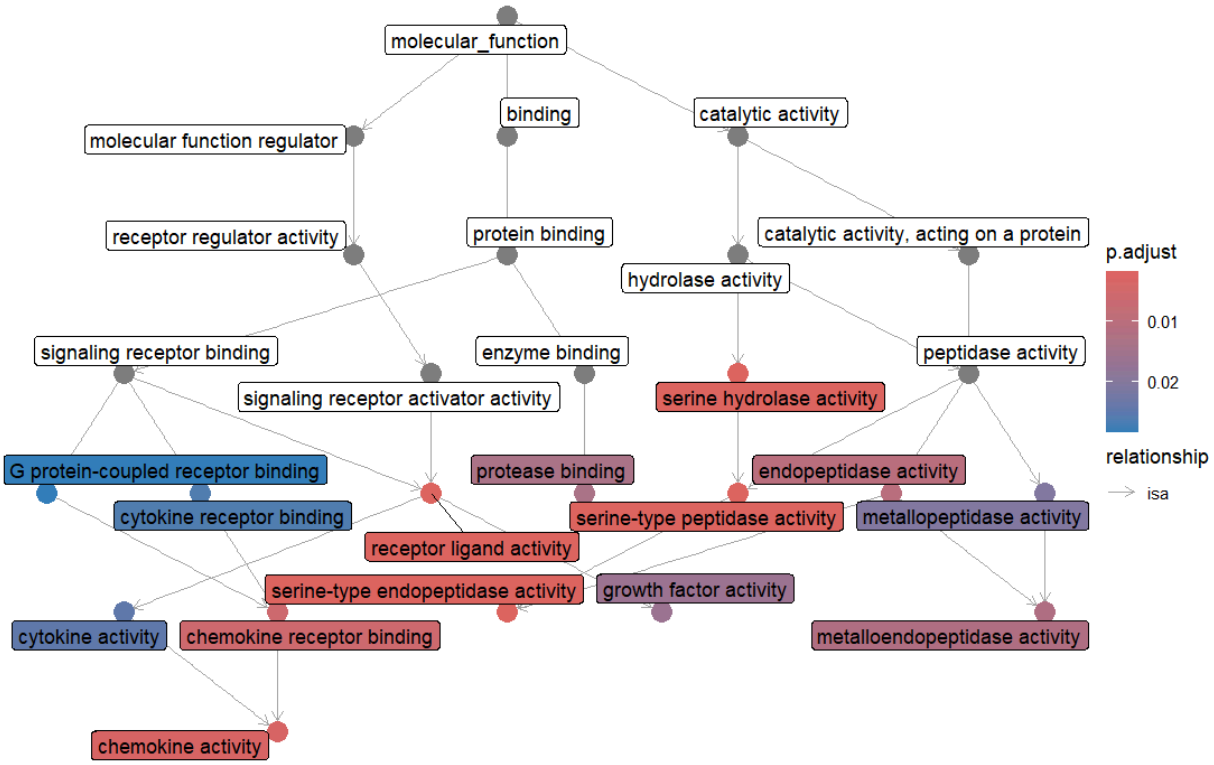

Figure 8: Network of molecular function pathways of genes that lead to Nonalcoholic Steatohepatitis in GSE163211. Colored pathways are pathways found in our differentially expressed genes with different colors for different p-values, and non-colored are pathways that might lead to these pathways but are not included in the findings.

Discussion

This meta-analysis reports hepatic gene-expression profiling of NAFLD through mRNA and identifies genes that were differentially expressed with obesity status from two datasets, GSE89632 and GSE163211. Patients and genes in the analysis were considered with two specific conditions, diabetes or non-diabetes patients, and healthy liver versus Simple Steatosis (SS) and Nonalcoholic Steatohepatitis (NASH). It also

investigates molecular function pathways that these differentially expressed genes get involved in and rise from. The relationship between these genes and discovered some genes were differentially expressed in both stages of NAFLD is also interpreted, as well as the differences in their roles in each stage of the disease. This meta-analysis poses a unique opportunity to evaluate data and findings from different countries and eras regarding obesity, a major risk factor of NAFLD.

As pointed out before, in GSE89632, the DNA-binding factor was the most dominant pathway in Healthy patients to stages of NAFLD (SS and NASH), with the involvement of downregulated genes such as NR4A2, FOSB, JUNB,... (Table A.1.2). mi-RNA and regulatory RNA binding also have downregulated genes involved (Figure 3 and Table A.1.2). This indicates that when these genes and pathways decrease the expression levels, the severity of patients might get worse, which agrees with previous findings (Vachher et al., 2022). In GSE163211, even though I did not find similar differentially expressed genes when comparing Healthy and stages of NAFLD (Healthy versus SS and Healthy versus NASH), I found equivalent and similar pathways of these comparisons such as serine-related activities, insulin growth factor binding, and the presence of DNA-binding factors.

A range of up and down-regulated genes in each pathway involved in NAFLD progressions contributed to how that pathway influences the severity of patients. In GSE89632, from healthy to SS (Table A.1.2), patients experienced the downregulation of genes that contributed to the DNA-binding activities and oxidoreductase activity, indicating the decrease in controlling DNA-binding transcription and oxidoreductase caused the decrease of the ability to self-fixed gene issues in the liver, making they

experienced worse condition (J. Ye et al., 2020, Yoneda et al., 2008). This pattern was also found when examining patients that have NASH and Healthy, with the reduction in DNA-binding activity and exoribonuclease activity because of downregulated genes associated with these pathways (Table A.1.2). In the comparison of gene expression in SS and NASH, there is an upregulated gene involved in oxidoreductase and a downregulated gene related to protein and antigen binding. This created oxidative stress and a lower level of control on binding and surpassed the checkpoint from SS to NASH, created liver abnormal activity, and caused worse progression, which is similar to previous findings (Z. Li et al., 2023; Yoneda et al., 2008). In GSE163211, from Healthy to SS, the chromatin and histone binding are upregulated, the insulin growth factor pathway is downregulated, and both up and down-regulated genes contribute to the serine-related activities (serine hydrolase and serine-type peptidase). This shows that when obese patients get to the SS stage, there are pathways that contribute to the decrease of insulin productivity and protein, but increase the activity of chromatin and histone binding, causing an increase in liver destruction. This is because when the liver exhibits insulin resistance (decrease of insulin productivity pathways), it causes histone modification changes that may lead to the dysregulation of multiple biological processes associated with NAFLD, which agrees with the previous studies (Kitade et al., 2017; Shi & Qi, 2023). This explains the upregulation of genes and pathways in the comparison of healthy and NASH patients (Fig. 5B), such as receptor-ligand activity, protease binding, and serine-related pathways. These increased dysregulated processes create oxidative stress and inflammation, making the condition more harmful. This also explains why examining the comparison between SS and NASH

patients (Fig. 5C), protease binding, serine-related, and DNA-binding pathways were all upregulated. With this, the increase of these processes was proved to cause the progression of NAFLD to get worse, which agrees with previous studies (Shi & Qi, 2023, Z. Li et al., 2023).

Even though the approach (DESeq2, Love et al., 2014) allowed robust quantitation of gene expression of NAFLD, some limitations need to be considered when replicating the results and using data. These datasets (GSE89632 and GSE163211) provided machine-preprocessed data, which means they slightly normalized raw data while sequencing, while the DESeq2 package requires un-normalized and raw gene expression values. Hence, I back-transformed processes based on the described normalization procedure (GeneSpring v12.5 (Agilent Technologies) and NanoString (Bhattacharya et al., 2020)); however, there might be miscalculations and wrong results and interpretations. Moreover, there were not too many differentially expressed genes identified in the comparison of Healthy versus SS in GSE163211, which might be explained because of the sequencing assay, but this needs more care when replicating the study. Limitations of collected data were declared in their initial reports, such as NanoString sequencing may have biases towards SS and NASH, and in GSE89632, they got data from hepatectomy which may influence the expression of healthy control genes. Besides, the interpretations in this study also can not explain related pathways that were not mentioned, hence limiting the generalizability of the study.

Despite these limitations, this gene expression analysis in obese patients with NAFLD is uniquely robust due to the number of patients included and unbiased

whole-genome approach. There is a lot of potential for next steps and expansion of this study. This meta-analysis can be expanded with analyzing more datasets generated using more high throughput assays such as single-cell and whole-genome transcriptomics, which will provide a greater breadth of understanding gene-profiling of NAFLD and obesity. We can also expand the usage of data by developing a model to find the importance levels of shared genes and fully understand the potential of these genes, or maybe including information regarding survival rate from recorded time of stages and recovery rate.

In summary, I report gene-expression profiles of liver tissue from two different cohorts with a large number of patients and transcripts included, providing a robust analysis of gene and pathways found. This study identifies pathways that agree with previous findings and consistent with initial reports of two datasets, as well as highlights genes that can be further investigated and act as biomarkers for drug targets in NAFLD diabetes patients for Simple Steatosis and NASH. A combination of well-interpretation and well-design studies will be sufficient to understand NAFLD pathogenesis and develop molecular biomarkers for this disease progression in the future.

Supplemental material

Please see the supplemental materials on Github of the study (Nguyen, 2024) and in the attached package of this study.

Works Cited

Adams, L. A., White, S. W., Marsh, J. A., Lye, S. J., Connor, K. L., Maganga, R., Ayonrinde, O. T., Olynyk, J. K., Mori, T. A., Beilin, L. J., Palmer, L. J., Hamdorf, J. M., & Pennell, C. E. (2013). Association between liver-specific gene polymorphisms and their expression levels with nonalcoholic fatty liver disease. *Hepatology*, *57*(2), 590–600. https://doi.org/10.1002/hep.26184

Arendt, B. M., Comelli, E. M., Ma, D. W. L., Lou, W., Teterina, A., Kim, T., Fung, S. K., Wong, D. K. H., McGilvray, I., Fischer, S. E., & Allard, J. P. (2015). Altered hepatic gene expression in nonalcoholic fatty liver disease is associated with lower hepatic n-3 and n-6 polyunsaturated fatty acids. *Hepatology*, *61*(5), 1565–1578. https://doi.org/10.1002/hep.27695

Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., Marshall, K. A., Phillippy, K. H., Sherman, P. M., Holko, M., Yefanov, A., Lee, H., Zhang, N., Robertson, C. L., Serova, N., Davis, S., & Soboleva, A. (2013). NCBI GEO: Archive for functional genomics data sets—update. *Nucleic Acids Research*, *41*(D1), D991–D995. https://doi.org/10.1093/nar/gks1193

Bellentani, S., Scaglioni, F., Marino, M., & Bedogni, G. (2010). Epidemiology of Non-Alcoholic Fatty Liver Disease. *Digestive Diseases*, *28*(1), 155–161. https://doi.org/10.1159/000282080

Bhattacharya, A., Hamilton, A. M., Furberg, H., Pietzak, E., Purdue, M. P., Troester, M. A., Hoadley, K. A., & Love, M. I. (2020). An approach for normalization and quality control for NanoString RNA expression data. *Briefings in Bioinformatics*, *22*(3), bbaa163. https://doi.org/10.1093/bib/bbaa163

Brunt, E. M., Wong, V. W.-S., Nobili, V., Day, C. P., Sookoian, S., Maher, J. J., Bugianesi, E., Sirlin, C. B., Neuschwander-Tetri, B. A., & Rinella, M. E. (2015). Nonalcoholic fatty liver disease. *Nature Reviews Disease Primers*, *1*(1), Article 1.

https://doi.org/10.1038/nrdp.2015.80

Davis, S., & Meltzer, P. S. (2007). GEOquery: A bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics*, *23*(14), 1846–1847. https://doi.org/10.1093/bioinformatics/btm254

Dunning, Mark, Andy Lynch, and Matthew Eldridge. "illuminaHumanv4. db: Illumina HumanHT12v4 annotation data (chip illuminaHumanv4)." *R package version* 1.0 (2015). https://doi.org/10.1002/hep4.1789

Fan, R., Wang, J., & Du, J. (2018). Association between body mass index and fatty liver risk: A dose-response analysis. *Scientific Reports*, *8*(1), 15273. https://doi.org/10.1038/s41598-018-33419-6

Friedman, S. L., Neuschwander-Tetri, B. A., Rinella, M., & Sanyal, A. J. (2018). Mechanisms of NAFLD development and therapeutic strategies. *Nature Medicine*, *24*(7), Article 7. https://doi.org/10.1038/s41591-018-0104-9

Kitade, H., Chen, G., Ni, Y., & Ota, T. (2017). Nonalcoholic Fatty Liver Disease and Insulin Resistance: New Insights and Potential New Treatments. *Nutrients*, *9*(4), 387. https://doi.org/10.3390/nu9040387

LaBrecque, D. R., Abbas, Z., Anania, F., Ferenci, P., Khan, A. G., Goh, K.-L., Hamid, S. S., Isakov, V., Lizarzabal, M., Peñaranda, M. M., Ramos, J. F. R., Sarin, S., Stimac, D., Thomson, A. B. R., Umar, M., Krabshuis, J., LeMair, A., & Team, R. (2014). World Gastroenterology Organisation Global Guidelines: Nonalcoholic Fatty Liver Disease and Nonalcoholic Steatohepatitis. *Journal of Clinical Gastroenterology*, *48*(6), 467. https://doi.org/10.1097/MCG.0000000000000116

Lallukka, S., Sevastianova, K., Perttilä, J., Hakkarainen, A., Orho-Melander, M., Lundbom, N., Olkkonen, V. M., & Yki-Järvinen, H. (2013). Adipose tissue is inflamed in NAFLD due to obesity but not in NAFLD due to genetic variation in PNPLA3. *Diabetologia*, *56*(4), 886–892. https://doi.org/10.1007/s00125-013-2829-9

Li, L., Liu, D.-W., Yan, H.-Y., Wang, Z.-Y., Zhao, S.-H., & Wang, B. (2016). Obesity is an independent risk factor for non-alcoholic fatty liver disease: Evidence from a meta-analysis of 21 cohort studies. *Obesity Reviews*, *17*(6), 510–519. https://doi.org/10.1111/obr.12407

Li, Z., Zhang, H., Li, Q., Feng, W., Jia, X., Zhou, R., Huang, Y., Li, Y., Hu, Z., Hu, X., Zhu, X., & Huang, S. (2023). GepLiver: An integrative liver expression atlas spanning developmental stages and liver disease phases. *Scientific Data*, *10*(1), 376. https://doi.org/10.1038/s41597-023-02257-1

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550. https://doi.org/10.1186/s13059-014-0550-8

Nguyen, L. (2024). *Lynnnguyen1/NAFLD_gene* [Computer software]. https://github.com/lynnnguyen1/NAFLD_gene (Original work published 2024)

Shi, Y., & Qi, W. (2023). Histone Modifications in NAFLD: Mechanisms and Potential Therapy. *International Journal of Molecular Sciences*, *24*(19), 14653. https://doi.org/10.3390/ijms241914653

Subudhi, S., Drescher, H. K., Dichtel, L. E., Bartsch, L. M., Chung, R. T., Hutter, M. M., Gee, D. W., Meireles, O. R., Witkowski, E. R., Gelrud, L., Masia, R., Osganian, S. A., Gustafson, J. L., Rwema, S., Bredella, M. A., Bhatia, S. N., Warren, A., Miller, K. K., Lauer, G. M., & Corey, K. E. (2022). Distinct Hepatic Gene-Expression Patterns of NAFLD in Patients With Obesity. *Hepatology Communications*, *6*(1), 77–89. https://doi.org/10.1002/hep4.1789

R Core Team (2023). _R: A Language and Environment for Statistical Computing_. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>

Takamura, T., Sakurai, M., Ota, T., Ando, H., Honda, M., & Kaneko, S. (2004). Genes for systemic vascular complications are differentially expressed in the livers of Type 2

diabetic patients. *Diabetologia*, *47*(4), 638–647. https://doi.org/10.1007/s00125-004-1366-y

Vachher, M., Bansal, S., Kumar, B., Yadav, S., & Burman, A. (2022). Deciphering the role of aberrant DNA methylation in NAFLD and NASH. *Heliyon*, *8*(10), e11119. https://doi.org/10.1016/j.heliyon.2022.e11119

Ye, H., Sun, M., Su, M., Chen, D., Liu, H., Ma, Y., Luo, W., Li, H., & Xu, F. (2023). Identification of disease-related genes and construction of a gene co-expression database in non-alcoholic fatty liver disease. *Frontiers in Genetics*, *14*. https://www.frontiersin.org/articles/10.3389/fgene.2023.1070605

Ye, J., Lin, Y., Wang, Q., Li, Y., Zhao, Y., Chen, L., Wu, Q., Xu, C., Zhou, C., Sun, Y., Ye, W., Bai, F., & Zhou, T. (2020). Integrated Multichip Analysis Identifies Potential Key Genes in the Pathogenesis of Nonalcoholic Steatohepatitis. *Frontiers in Endocrinology*, *11*. https://doi.org/10.3389/fendo.2020.601745

Yoneda, M., Endo, H., Mawatari, H., Nozaki, Y., Fujita, K., Akiyama, T., Higurashi, T., Uchiyama, T., Yoneda, K., Takahashi, H., Kirikoshi, H., Inamori, M., Abe, Y., Kubota, K., Saito, S., Kobayashi, N., Yamaguchi, N., Maeyama, S., Yamamoto, S., … Nakajima, A. (2008). Gene expression profiling of non-alcoholic steatohepatitis using gene set enrichment analysis. *Hepatology Research*, *38*(12), 1204–1212. https://doi.org/10.1111/j.1872-034X.2008.00399.x

Yu, G., Wang, L.-G., Han, Y., & He, Q.-Y. (2012). clusterProfiler: An R Package for Comparing Biological Themes Among Gene Clusters. *OMICS: A Journal of Integrative Biology*, *16*(5), 284–287. https://doi.org/10.1089/omi.2011.0118