Exploring key genes in NAFLD based on glutamine metabolism: a comprehensive analysis combining multi-omics, machine learning and SHAP

doi: 10.6133/apjcn.202512/PP.0005
Published online: December 2025

Running title: Glutamine Multi-omics in NAFLD

Changan Chen BM[1], Wenfeng Liu BM[2], Yongtao Lan BM[2], Fuxiong Li MMed[3], Xiaoman Li BM[4]

[1]Department of Gastroenterology, Affiliated Hospital of Guangdong Medical University, Zhanjiang, China
[2]The First School of Clinical Medicine, Guangdong Medical University, Zhanjiang, China
[3]School of Basic Medicine, Guangdong Medical University, Zhanjiang, China
[4]Department of Endocrinology, Affiliated Hospital of Guangdong Medical University, Zhanjiang, China

Authors' email addresses and contributions:
CC: chenchangan0319@gdmu.edu.cn
Contribution: data analysis, visualization and paper writing.

LW: lww1236@gdmu.edu.cn
Contribution: methodology

LY: 412543965@gdmu.edu.cn
Contribution: data collection

LF: lxiong621@gdmu.edu.cn
Contribution: data analysis.

LX: l_man@gdmu.edu.cn
Contribution: paper editing.

Corresponding Author: Dr Xiaoman Li, Department of Endocrinology, Affiliated Hospital of Guangdong Medical University, Zhanjiang, 524001, China. Affiliated Hospital of Guangdong Medical University, No. 57, Renmin Avenue South, Xia Shan District, Zhanjiang, Guangdong, 524001, China. Tel:. Email: l_man@gdmu.edu.cn

ABSTRACT

Background and Objectives: Non-alcoholic fatty liver disease (NAFLD) is a prevalent liver condition globally, with an escalating incidence and a strong association with various metabolic disorders, thus presenting a significant public health challenge. Currently, there is a scarcity of effective preventive or therapeutic methods for NAFLD. This study used multi-omics, machine learning (ML), and SHAP comprehensive analysis to explore NAFLD-related metabolites and genes, hoping to provide new insights. **Methods and Study Design:** We initially conducted MR analysis on 1,400 serum metabolites and two NAFLD datasets, identifying glutamine as causally linked to NAFLD. In single-cell RNA sequencing, hepatocytes were categorized into high-synthesis and low-synthesis glutamine groups for cell communication analysis. We extracted differentially expressed genes from these two groups and performed GO and KEGG enrichment analysis. Further screening of these genes was followed by the application of least absolute shrinkage and selection operator (LASSO) regression to identify hub genes for ML. We constructed the ML model using Catboost, NGboost, and XGboost algorithms. Finally, we employed the SHAP method to interpret the model, identifying key genes with significant model contributions. **Results:** MR analysis demonstrated that the glutamine-to-alanine ratio and levels of 1-linoleoyl-2-arachidonoyl-GPC (18:2/20:4n6) were associated with a reduced incidence of NAFLD. We identified 19 hub genes for ML, with validation set AUCs of 0.83 for Catboost, 0.82 for NGboost, and 0.86 for XGboost. The SHAP analysis highlighted ASL, LGALS1, and GLUL as genes with the contributed significantly to the models. **Conclusions:** Our MR findings suggest that specific metabolites may lower the risk of NAFLD. A comprehensive analysis underscores the significant role of glutamine metabolism and related genes in NAFLD pathogenesis, offering new potential targets for NAFLD diagnosis and treatment.

Key Words: multi-Omics, machine Learning, SHAP, non-alcoholic fatty liver disease, glutamine metabolism

INTRODUCTION

Non-alcoholic fatty liver disease (NAFLD) is widespread, affecting approximately 32.4% of the global population.[1] It is defined by the accumulation of lipid droplets in hepatocytes in the absence of excessive alcohol intake, with at least 5% hepatic steatosis.[2] The condition poses a significant public health issue due to its increasing prevalence and strong linkage to various metabolic disorders such as obesity, type 2 diabetes, and dyslipidemia.[3-4] Annually, NAFLD-

related medical costs surpass €35 billion across four major European nations (the United Kingdom, France, Germany, and Italy) and exceed $100 billion in the United States, imposing a substantial clinical burden.[5] In 2020, an international consensus expert panel recommended the term metabolic dysfunction-associated fatty liver disease (MAFLD) as a replacement for NAFLD, considering MAFLD's broader and independent scope.[6-7] Despite extensive research efforts, no effective prevention or treatment methods for NAFLD are currently available.

Recent genomic and metabolomic studies have offered new insights into the pathogenesis of NAFLD.[8] Variants in genes such as patatin-like phospholipase domain-containing 3, transmembrane 6 superfamily member 2, and 17-beta hydroxysteroid dehydrogenase 13 are linked to NAFLD susceptibility.[9] Metabolic processes involving lipids, amino acids, and bile acids are thought to contribute to NAFLD's pathogenesis.[10-11] For instance, changes in circulating amino acids, such as elevated levels of branched-chain and aromatic amino acids, along with reduced levels of amino acids related to glutathione synthesis, are common in NAFLD patients.[12-14] NAFLD patients also show significantly increased plasma total primary bile acids and decreased secondary bile acids.[15] Nevertheless, causal relationships between these metabolites and NAFLD remain to be clarified.

Mendelian randomization (MR) offers a method of causal inference, utilizing germline genetic variations as instrumental variables (IVs) to minimize bias from residual confounding or reverse causation.[16-17] Machine learning (ML) is adept at analyzing large datasets, revealing valuable patterns and explanations.[18] However, the "black box" nature of ML limits the use of more sophisticated ML methods in medical decision support. Shapley Additive exPlanations (SHAP), rooted in game theory, provides a transparent explanation of ML models.

Our study integrated MR and multi-omics analysis with ML to examine serum metabolites and genes linked to NAFLD pathogenesis, using the SHAP algorithm to elucidate the constructed ML model. This aims to offer new directions and insights for NAFLD diagnosis and treatment.

MATERIALS AND METHODS

Mendelian randomization design

Selection and description of data sources

The Genome-Wide Association Studies (GWAS) data utilized in this study were sourced from IEU OpenGWAS (https://gwas.mrcieu.ac.uk/), the GWAS Catalog (https://www.ebi.ac.uk/gwas/), and FINNGEN (https://www.finngen.fi/en).[19-21] All these data are publicly accessible. For exposure data, we analyzed 1,400 plasma metabolites with

accession numbers GCST90199621-90201020 from the GWAS Catalog.[22] The outcome data concerning NAFLD were sourced from IEU, dataset: ebi-a-GCST90091033, and FINNGEN, dataset: finngen_R10_NAFLD, primarily involving European participants.

Instrumental variables selection

(a) We filtered 1,400 plasma metabolic single nucleotide polymorphisms (SNPs) using a significance threshold of $p < 1 \times 10^{-5}$. (b) We ensured a physical distance of more than 10,000 k between each pair of SNPs, and the r^2 threshold for linkage disequilibrium was set at < 0.001. Palindromic SNPs with inconsistent effect allele frequencies across the exposure and outcome datasets were excluded. (c) We included only SNPs with an F-statistic > 10.[23]

MR Analysis

The "TwoSample MR" package in R version 4.3.1 was employed for MR analysis. Primarily, we utilized the inverse variance weighting (IVW) method to estimate the causal relationship between exposure and outcome, applying a fixed-effect meta-analysis model that assumes all SNPs exert a single true effect size.[24] We conducted MR analysis across 1,400 plasma metabolites and the two NAFLD datasets (ebi-a-GCST90091033 and finngen_R10_NAFLD). A p-value < 0.05 via the IVW method indicated significant associations. To confirm robustness, we applied additional MR methods (MR-Egger, Weighted Median, Weighted Mode, and Simple Mode). We identified compounds with causal relationships by intersecting results from both datasets.

Sensitivity analysis

Cochran's Q test p-value assessed heterogeneity, with $p < 0.05$ signaling significant heterogeneity.[25] We used the MR-Egger intercept test to detect horizontal pleiotropy.[26] MR-PRESSO was employed to identify outlier instrumental variables and provide adjusted causal estimates after their removal.[27] Additionally, the leave-one-out method was used to evaluate result stability by sequentially removing each SNP.

Single-cell RNA sequencing data processing

The single-cell RNA sequencing (scRNA-seq) dataset for NAFLD was obtained from the National Center for Biotechnology Information Gene Expression Omnibus (GEO) (https://www.ncbi.nlm.nih.gov/geo/) GSE202379.[28] The "Seurat" package (version 4.3.0) was used for single-cell analysis. Samples with less than 10% mitochondrial content, fewer than

4000 genes, and more than 200 genes were retained after quality control. High-quality samples were integrated and then normalized using the NormalizeData function. Principal component analysis was conducted via the RunPCA function for clustering and dimensionality reduction, and cell clusters were annotated using the SingleR package. The intersection of the MR results of the two data sets gave the glutamine to alanine ratio and 1-linoleoyl-2-arachidonoyl-GPC (18:2/20:4n6) levels, which can reduce the risk of NAFLD. Among them, glutamine is the most abundant and widely used amino acid in the human body and is related to immune regulation in the human body.[29] Moreover, in ebi-a-GCST90091033, glutamine can reduce the risk of NAFLD (IVW $p = 0.01$). In the liver, alanine can react with α-ketoglutarate to generate pyruvate and glutamate, and glutamate can further synthesize glutamine in hepatocytes. Therefore, we determined that the synthesis of glutamine would be the next research direction. We downloaded the gmt file of glutamine synthesis: GOBP_GLUTAMINE_FAMILY_AMINO_ACID_BIOSYNTHETIC_PROCESS from MSigDB (https://www.gsea-msigdb.org/gsea/msigdb). In the processed single-cell dataset, the score of glutamine synthesis was added by read.gmt and AddModuleScore functions. It was found that hepatocytes had the highest score (Figure 4B). Hepatocytes were divided into two groups (synthesis_high and synthesis_low) according to the median score of glutamine synthesis, and a ratio graph was drawn.

Transcription factor activity prediction, cell communication, and enrichment analysis

The dorothea package predicted transcription factor activity differences between synthesis_high and synthesis_low groups within NAFLD patients, selecting the top 20 transcription factors for visualization. The CellChat package examined cell communication, particularly among synthesis_high, synthesis_low, and other cells. Differentially expressed genes in synthesis_high and synthesis_low groups were identified using the FindMarkers function. KEGG and GO enrichment analysis were carried out, significant at an adjusted p-value < 0.05.

Construction of training and validation sets and screening of hub genes

GSE61260 was chosen as the training set.[30] GSE48452 served as the test set.[31] Gene symbols were matched to probes based on the GPL11532 platform, and common genes between the sets were extracted. Batch effect correction was applied, and subsequent analysis and modeling utilized these corrected sets. Further filtering of differentially expressed genes involved choosing those with a p-value < 0.05 and log2FC > 0.25, excluding mitochondrial,

ribosomal, and erythrocyte genes. LASSO regression was applied to identify hub genes. We utilized the glmnet package in R to fit a binomial LASSO regression model. To ensure the robustness of the model and to determine the optimal value of the penalty parameter (lambda), we performed 10-fold cross-validation using the cv.glmnet function. The lambda value that minimizes the binomial deviance was selected as the optimal parameter. A PPI network for the hub gene was constructed using the STRING database (version 12.0).

Machine learning and SHAP algorithm analysis

For hub genes selection, CatBoost, XGBoost, and NGBoost were utilized as ML models due to their flexibility, scalability, and high usability, making them prevalent in various research domains.[32-34] SHAP, based on game theory, was employed to elucidate the importance of each model feature.[35] Post-ML model construction, SHAP interpretation provided insights into the most influential genes impacting the model. Figure 1 illustrates the study design flowchart.

RESULTS

Causal association of glutamine-to-alanine ratio with reduced NAFLD risk

Results with IVW *p* value less than 0.05 in the MR analysis of 1,400 metabolites and NAFLD are detailed in Supplementary Table 1–2. By MR analysis, we obtained the ratio of glutamine to alanine ratio and 1-linoleoyl-2-arachidonoyl-GPC (18:2/20:4n6) levels in ebi-a-GCST90091033 and finngen_R10_NAFLD, with p values less than 0.05 in the IVW method (Figure 2A). The glutamine to alanine ratio was associated with a decreased risk of NAFLD (Figure 2B-C, Table 1), and glutamine was shown to lower NAFLD risk within the ebi-a-GCST90091033 dataset (Figure 2D, Table 1). Their scatter plot and leaveoneout plot are shown in Figure 3. Sensitivity analysis confirmed no pleiotropy regarding glutamine to alanine and glutamine to NAFLD associations (Table 2). MR-PRESSO and leave-one-out analysis detected no abnormal instrumental variables.

Single-cell RNA sequencing reveals hepatocyte glutamine synthesis heterogeneity and endothelial cell interactions

The annotated cell plan view is shown in Figure 4A. Incorporating the glutamine synthesis score into scRNA-seq revealed hepatocytes as having the highest synthesis score (Figure 4B). Based on this median score, hepatocytes were categorized into high and low glutamine synthesis groups (synthesis_high and synthesis_low). The distribution chart indicates a higher proportion of synthesis_high in healthy individuals compared to NAFLD patients (Figure 4C),

aligning with MR analysis findings. In cell communication analysis, synthesis_high demonstrated more extensive interactions with endothelial cells than synthesis_low (Figure 4D-E). Previous studies have shown that endothelial cells exert anti-inflammatory and anti-fibrotic effects by inhibiting Kupffer cell and hepatic stellate cell activation and regulating intrahepatic vascular resistance and portal vein pressure.36 Based on cell communication results, synthesis_high hepatocytes may exert their anti-inflammatory effects by activating endothelial cells to regulate the hepatic vascular microenvironment. Transcription factor analysis showed opposing predictions between synthesis_high and synthesis_low (Figure 4F). The differential genes of synthesis_high and synthesis_low were extracted, and a total of 4785 differential genes were obtained for enrichment analysis. KEGG enrichment analysis showed that the differential genes were not only enriched in NAFLD but also in various neurodegenerative diseases (Figure 4G). The homeostasis of glutamine in the body is also related to neurodegeneration.[37-38] GO enrichment analysis highlighted gene involvement in RNA splicing, ribosome, mitochondrial inner membrane, and cadherin binding (Figure 4H).

Machine learning models and SHAP uncover key regulators (ASL, LGALS1, GLUL) with diagnostic utility in NAFLD

Further screening of 4,785 differential genes with a p-value < 0.05 and log2FC > 0.25, excluding mitochondrial, ribosomal, and erythrocyte genes, resulted in 30 candidate genes, with LASSO regression narrowing this to 19 hub genes (Figures 5A-B). PPI analysis of these hub genes utilized the STRING database (Figure 5C). We developed ML models using Catboost, NGboost, and XGboost for these hub genes. In the validation set, the Areas Under the Curve (AUC) were 0.83 for Catboost, 0.82 for BGboost, and 0.86 for XGboost (Figure 5D), affirming high diagnostic accuracy (AUC > 0.7). SHAP analysis provided model insights through Summary Plots, Beeswarm Plots, and Heatmap Plots (Figure 5E-M). ASL contributed most significantly to Catboost, LGALS1 to NGboost, and LGALS1 also to XGboost. Notably, ASL, LGALS1, and GLUL consistently ranked among the top five contributors in all models. LGALS1 correlated positively with NAFLD risk, whereas ASL and GLUL correlated negatively. This suggests that ASL, LGALS1, and GLUL may play important roles in NAFLD.

DISCUSSION

In this study, we employed MR, multi-omics analysis, ML, and SHAP methods to explore metabolites and key genes implicated in the pathogenesis of NAFLD. Our findings showed

that the glutamine to alanine ratio and 1-linoleoyl-2-arachidonoyl-GPC (18:2/20:4n6) levels can reduce NAFLD risk. Hepatocytes with high glutamine synthesis (synthesis_high) serve as protective factors against NAFLD. Utilizing ML and SHAP algorithms, we constructed diagnostic models that identified ASL, LGALS1, and GLUL as significant genes. These metabolites and genes may offer novel therapeutic directions for NAFLD.

Over the past few decades, the intricate link between metabolism and the immune system has been increasingly recognized. Glutamine emerged as a protective factor in our analysis. It is well known that glutathione, a tripeptide comprising glutamate, cysteine, and glycine, protects tissues from oxidative damage by detoxifying reactive species and/or repairing cellular damage. Glutamine can reduce pro-inflammatory gene and protein levels in adipocytes, and its increased breakdown is observed in NAFLD-affected livers.[39-40] Targeted glutaminase therapy has shown promise in improving nonalcoholic steatohepatitis by facilitating very low-density lipoprotein triglyceride assembly.[41] Decreased serum glutamine levels correlate with liver fibrosis progression and depressive symptoms in NAFLD patients.[42-43] Hypoxia-inducible factor 2a activation can lead to glutaminolysis by inhibiting YAP phosphorylation and increasing YAP nuclear translocation, thereby enhancing NAFLD fibrosis and progression.[44] Glutamine supplementation has been shown to ameliorate intestinal flora dysbiosis and NAFLD induced by high-fat diets in mice.[45] Our MR analysis also identified that 1-linoleoyl-2-arachidonoyl-GPC (18:2/20:4n6) could lower NAFLD risk. NAFLD patients often show reduced levels of sphingolipids and phosphocholine in both liver and plasma, compared to healthy individuals.[46] Phosphatidylcholine, in particular, is effective in reducing endotoxin-stimulated tumor necrosis factor-α release in the liver.[47] Additional metabolites like vitamin A and deoxycholate, shown to reduce NAFLD risk, were also identified. Notably, vitamin A intake inversely correlates with NAFLD risk, and tauroursodeoxycholic acid has been shown to moderate intestinal inflammation and NAFLD progression in mice.[48-49] Further studies are needed to elucidate the mechanisms through which these metabolites affect NAFLD.

Through the transcription factor prediction analysis of synthesis_high and synthesis_low, we found that the predictions of transcription factors for these two cells were completely opposite. Synthesis_high correlates positively with transcription factors such as CREB1, ATF4, PPARA, HNF1B, and TCF7L2. CREB1 modulates stearoyl-CoA desaturase 1, a contributor to hepatic steatosis.[50] ATF4's involvement in endoplasmic reticulum stress and autophagy modulation can mitigate hepatic steatosis.[51] PPARA regulates hepatic lipid metabolism.[52] HNF1B impacts liver fat content in high-fat diet mice, and TCF7L2 maintains

bile acid and lipid homeostasis via gut-liver communication.[53-54] Hence, synthesis_high potentially inhibits NAFLD progression through these pathways.

Using SHAP to analyze our ML models revealed GLUL, ASL, and LGALS1 as key gene contributors. ASL and GLUL were inversely related to NAFLD risk, while LGALS1 showed a positive correlation. GLUL encodes glutamine synthetase and is vital in detoxifying ammonia, glutamate signaling, and various cellular processes. Decreased GLUL expression was observed in a 2,3,7,8-tetrachlorodibenzo-p-dioxin-induced mouse NAFLD model.[55] GLUL expression in the subcutaneous adipose tissue of obese patients was significantly lower than that in lean women.[56] Butyrate can increase GLUL expression to reduce metabolic disorders induced by a high-fat diet.[57] There are currently no research reports on the mechanisms of ASL and LGALS1 in NAFLD. ASL encodes a member of the lyase 1 family, which mainly catalyzes the reversible hydrolytic cleavage of argininosuccinate into arginine and fumarate, which is an important step in the liver's detoxification of ammonia through the urea cycle. In other inflammatory diseases, ASL upregulation can induce endogenous nitric oxide production in intestinal epithelial cells, thereby improving epithelial integrity and alleviating colitis and inflammation-related colon cancer.[58] In patients with psoriasis, ASL reduces the inflammatory response by increasing arginine.[59] LGALS1 is a protein-coding gene, and this gene product may act as an autocrine negative growth factor that regulates cell proliferation. LGASL1 is involved in the apoptosis of liver cancer cells.[60] In summary, these genes are related to the regulation of inflammation or apoptosis, suggesting that they may serve as new targets for the diagnosis and treatment of NAFLD in the future.

Our study's strength lies in leveraging MR analysis to estimate causal relationships using genetic variants, minimizing confounding biases. Combining multi-omics, ML, and SHAP methods, we identified critical NAFLD-related genes, improving the robustness of our results. However, there are some limitations to our analysis. First, the GWAS data used in our study mainly involved European populations, which may limit the generalizability of our findings to other populations. Second, the validation set we chose has a relatively small sample size. Validating our results on a larger cohort will be the direction of our future research. Finally, this study has not yet verified the function and mechanism of the results. For example, how hepatocytes regulate the anti-inflammatory effect of endothelial cells by synthesizing glutamine is also a focus of our future research.

Conclusion

In summary, through the comprehensive analysis of MR, multi-omics, ML, and SHAP, we found that some metabolites such as glutamine and 1-linoleoyl-2-arachidonoyl-GPC (18:2/20:4n6) levels can reduce the risk of NAFLD. Moreover, we also found three key genes related to NAFLD: GLUL, ASL, and LGALS1. This suggests that these metabolites and genes may provide new therapeutic targets for NAFLD. Future research efforts could be aimed at validating these hypotheses and uncovering implications for clinical practice, potentially leading to innovative treatments that address the immune and metabolic components of NAFLD.

SUPPLEMENTARY MATERIALS

All supplementary tables and figures are available upon request from the editorial office, and are also accessible on the journal's webpage (apjcn.qdu.edu.cn).

ACKNOWLEDGEMENTS

We want to acknowledge the participants and investigators of the FinnGen study and all individuals and research teams who provided raw data, as well as individuals who participated in the research.

CONFLICT OF INTEREST AND FUNDING DISCLOSURE

The authors declare no conflict of interest.

This research was not funded.

REFERENCES

1. Riazi K, Azhari H, Charette JH, et al. The prevalence and incidence of NAFLD worldwide: a systematic review and meta-analysis. Lancet Gastroenterol Hepatol. 2022;7:851-61. doi:10.1016/S2468-1253(22)00165-0
2. Scorletti E, Carr RM. A new perspective on NAFLD: Focusing on lipid droplets. J Hepatol. 2022;76:934-45. doi:10.1016/j.jhep.2021.11.009
3. Adams LA, Anstee QM, Tilg H, Targher G. Non-alcoholic fatty liver disease and its relationship with cardiovascular disease and other extrahepatic diseases. Gut. 2017;66:1138-53. doi:10.1136/gutjnl-2017-313884
4. Loomba R, Friedman SL, Shulman GI. Mechanisms and disease consequences of nonalcoholic fatty liver disease. Cell. 2021;184:2537-64. doi:10.1016/j.cell.2021.04.015

5. Ofosu A, Ramai D, Reddy M. Non-alcoholic fatty liver disease: controlling an emerging epidemic, challenges, and future directions. Ann Gastroenterol. 2018;31:288-95. doi:10.20524/aog.2018.0240
6. Eslam M, Newsome PN, Sarin SK, et al. A new definition for metabolic dysfunction-associated fatty liver disease: An international expert consensus statement. J Hepatol. 2020;73:202-09. doi:10.1016/j.jhep.2020.03.039
7. Eslam M, Sanyal AJ, George J; International Consensus Panel. MAFLD: A Consensus-Driven Proposed Nomenclature for Metabolic Associated Fatty Liver Disease. Gastroenterology. 2020;158:1999-2014.e1. doi:10.1053/j.gastro.2019.11.312
8. Leung H, Long X, Ni Y, et al. Risk assessment with gut microbiome and metabolite markers in NAFLD development. Sci Transl Med. 2022;14:eabk0855. doi:10.1126/scitranslmed.abk0855
9. Taliento AE, Dallio M, Federico A, Prati D, Valenti L. Novel Insights into the Genetic Landscape of Nonalcoholic Fatty Liver Disease. Int J Environ Res Public Health. 2019;16:2755. doi:10.3390/ijerph16152755
10. Masarone M, Troisi J, Aglitti A, et al. Untargeted metabolomics as a diagnostic tool in NAFLD: discrimination of steatosis, steatohepatitis and cirrhosis. Metabolomics. 2021;17:12. doi:10.1007/s11306-020-01756-1
11. Kim HY. Recent advances in nonalcoholic fatty liver disease metabolomics. Clin Mol Hepatol. 2021;27:553-59. doi:10.3350/cmh.2021.0127
12. Yamakado M, Tanaka T, Nagao K, et al. Plasma amino acid profile associated with fatty liver disease and co-occurrence of metabolic risk factors. Sci Rep. 2017;7:14485. doi:10.1038/s41598-017-14974-w
13. Gaggini M, Carli F, Rosso C, et al. Altered amino acid concentrations in NAFLD: Impact of obesity and insulin resistance. Hepatology. 2018;67:145-58. doi:10.1002/hep.29465
14. Hasegawa T, Iino C, Endo T, et al. Changed Amino Acids in NAFLD and Liver Fibrosis: A Large Cross-Sectional Study without Influence of Insulin Resistance. Nutrients. 2020;12:1450. doi:10.3390/nu12051450
15. Puri P, Daita K, Joyce A, et al. The presence and severity of nonalcoholic steatohepatitis is associated with specific changes in circulating bile acids. Hepatology. 2018;67:534-548. doi:10.1002/hep.29359
16. Skrivankova VW, Richmond RC, Woolf BAR, et al. Strengthening the Reporting of Observational Studies in Epidemiology Using Mendelian Randomization: The STROBE-MR Statement. JAMA. 2021;326:1614-21. doi:10.1001/jama.2021.18236
17. Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. Hum Mol Genet. 2014;23:R89-R98. doi:10.1093/hmg/ddu328
18. Greener JG, Kandathil SM, Moffat L, Jones DT. A guide to machine learning for biologists. Nat Rev Mol Cell Biol. 2022;23:40-55. doi:10.1038/s41580-021-00407-0
19. Hemani G, Zheng J, Elsworth B, et al. The MR-Base platform supports systematic causal inference across the human phenome. Elife. 2018;7:e34408. doi:10.7554/eLife.34408
20. Kurki MI, Karjalainen J, Palta P, et al. FinnGen provides genetic insights from a well-phenotyped isolated population. Nature. 2023;613:508-18. doi:10.1038/s41586-022-05473-8

21. Elsworth B, Lyon M, Alexander T, et al. The MRC IEU OpenGWAS data infrastructure[J]. BioRxiv, 2020: 2020.08. 10.244293
22. Chen Y, Lu T, Pettersson-Kymmer U, et al. Genomic atlas of the plasma metabolome prioritizes metabolites implicated in human diseases. Nat Genet. 2023;55:44-53. doi:10.1038/s41588-022-01270-1
23. Mukamal KJ, Stampfer MJ, Rimm EB. Genetic instrumental variable analysis: time to call mendelian randomization what it is. The example of alcohol and cardiovascular disease. Eur J Epidemiol. 2020;35:93-97. doi:10.1007/s10654-019-00578-3
24. Zhang H, Chen L, Fan Z, Lv G. The causal effects of inflammatory bowel disease on primary biliary cholangitis: A bidirectional two-sample Mendelian randomization study. Liver Int. 2023;43:1741-48. doi:10.1111/liv.15616
25. Bowden J, Spiller W, Del Greco M F, et al. Improving the visualization, interpretation and analysis of two-sample summary data Mendelian randomization via the Radial plot and Radial regression. Int J Epidemiol. 2018;47(4):1264-1278. doi:10.1093/ije/dyy101
26. Burgess S, Thompson SG. Interpreting findings from Mendelian randomization using the MR-Egger method. Eur J Epidemiol. 2017;32:377-89. doi:10.1007/s10654-017-0255-x
27. Verbanck M, Chen CY, Neale B, Do R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. Nat Genet. 2018;50:693-98. doi:10.1038/s41588-018-0099-7
28. Gribben C, Galanakis V, Calderwood A, et al. Acquisition of epithelial plasticity in human chronic liver disease. Nature. 2024;630:166-73. doi:10.1038/s41586-024-07465-2
29. Cruzat V, Macedo Rogero M, Noel Keane K, Curi R, Newsholme P. Glutamine: Metabolism and Immune Function, Supplementation and Clinical Translation. Nutrients. 2018;10:1564. doi:10.3390/nu10111564
30. Horvath S, Erhart W, Brosch M, et al. Obesity accelerates epigenetic aging of human liver. Proc Natl Acad Sci U S A. 2014;111:15538-43. doi:10.1073/pnas.1412759111
31. Ahrens M, Ammerpohl O, von Schönfels W, et al. DNA methylation analysis in nonalcoholic fatty liver disease suggests distinct disease-specific and remodeling signatures after bariatric surgery. Cell Metab. 2013;18:296-302. doi:10.1016/j.cmet.2013.07.004
32. Hancock JT, Khoshgoftaar TM. CatBoost for big data: an interdisciplinary review. J Big Data. 2020;7:94. doi:10.1186/s40537-020-00369-8
33. Chen J, Wang M, Zhao D, et al. MSINGB: A Novel Computational Method Based on NGBoost for Identifying Microsatellite Instability Status from Tumor Mutation Annotation Data. Interdiscip Sci. 2023;15:100-10. doi:10.1007/s12539-022-00544-w
34. Zhang J, Ma X, Zhang J, et al. Insights into geospatial heterogeneity of landslide susceptibility based on the SHAP-XGBoost model. J Environ Manage. 2023;332:117357. doi:10.1016/j.jenvman.2023.117357

35. Li X, Zhao Y, Zhang D, et al. Development of an interpretable machine learning model associated with heavy metals' exposure to identify coronary heart disease among US adults via SHAP: Findings of the US NHANES from 2003 to 2018. Chemosphere. 2023;311(Pt 1):137039. doi:10.1016/j.chemosphere.2022.137039
36. Hammoutene A, Rautou PE. Role of liver sinusoidal endothelial cells in non-alcoholic fatty liver disease. J Hepatol. 2019;70:1278-91. doi:10.1016/j.jhep.2019.02.012
37. Westi EW, Andersen JV, Aldana BI. Using stable isotope tracing to unravel the metabolic components of neurodegeneration: Focus on neuron-glia metabolic interactions. Neurobiol Dis. 2023;182:106145. doi:10.1016/j.nbd.2023.106145
38. Andersen JV, Schousboe A. Glial Glutamine Homeostasis in Health and Disease. Neurochem Res. 2023;48:1100-28. doi:10.1007/s11064-022-03771-1
39. Petrus P, Lecoutre S, Dollet L, et al. Glutamine Links Obesity to Inflammation in Human White Adipose Tissue. Cell Metab. 2020;31:375-90.e11. doi:10.1016/j.cmet.2019.11.019
40. Simón J, Martínez-Chantar ML, Delgado TC. Glutamine, fatty liver disease and aging. Aging (Albany NY). 2021;13:3165-66. doi:10.18632/aging.202666
41. Simon J, Nuñez-García M, Fernández-Tussy P, et al. Targeting Hepatic Glutaminase 1 Ameliorates Non-alcoholic Steatohepatitis by Restoring Very-Low-Density Lipoprotein Triglyceride Assembly. Cell Metab. 2020;31:605-22. doi:10.1016/j.cmet.2020.01.013
42. Radford-Smith DE, Patel PJ, Irvine KM, et al. Depressive symptoms in non-alcoholic fatty liver disease are identified by perturbed lipid and lipoprotein metabolism. PLoS One. 2022;17(1):e0261555. doi:10.1371/journal.pone.0261555
43. Du K, Chitneni SK, Suzuki A, et al. Increased Glutaminolysis Marks Active Scarring in Nonalcoholic Steatohepatitis Progression. Cell Mol Gastroenterol Hepatol. 2020;10:1-21. doi:10.1016/j.jcmgh.2019.12.006
44. Yan R, Cai H, Zhou X, Bao G, Bai Z, Ge RL. Hypoxia-inducible factor-2α promotes fibrosis in non-alcoholic fatty liver disease by enhancing glutamine catabolism and inhibiting yes-associated protein phosphorylation in hepatic stellate cells. Front Endocrinol (Lausanne). 2024;15:1344971. doi:10.3389/fendo.2024.1344971
45. Zheng Y, Ying H, Shi J, Li L, Zhao Y. Alanyl-Glutamine Dipeptide Attenuates Non-Alcoholic Fatty Liver Disease Induced by a High-Fat Diet in Mice by Improving Gut Microbiota Dysbiosis. Nutrients. 2023;15:3988. doi:10.3390/nu15183988
46. Thiele M, Suvitaival T, Trošt K, et al. Sphingolipids Are Depleted in Alcohol-Related Liver Fibrosis. Gastroenterology. 2023;164:1248-60. doi:10.1053/j.gastro.2023.02.023
47. Lieber CS. Alcoholic fatty liver: its pathogenesis and mechanism of progression to inflammation and fibrosis. Alcohol. 2004;34:9-19. doi:10.1016/j.alcohol.2004.07.008
48. Liu C, Sun X, Peng J, Yu H, Lu J, Feng Y. Association between dietary vitamin A intake from different sources and non-alcoholic fatty liver disease among adults. Sci Rep. 2024;14(1):1851. doi:10.1038/s41598-024-52077-5

49. Wang W, Zhao J, Gui W, et al. Tauroursodeoxycholic acid inhibits intestinal inflammation and barrier disruption in mice with non-alcoholic fatty liver disease. Br J Pharmacol. 2018;175:469-84. doi:10.1111/bph.14095
50. Luo WJ, Cheng TY, Wong KI, et al. Novel therapeutic drug identification and gene correlation for fatty liver disease using high-content screening: Proof of concept. Eur J Pharm Sci. 2018;121:106-17. doi:10.1016/j.ejps.2018.05.018
51. Li J, Li X, Liu D, et al. Phosphorylation of eIF2α signaling pathway attenuates obesity-induced non-alcoholic fatty liver disease in an ER stress and autophagy-dependent manner. Cell Death Dis. 2020;11:1069. doi:10.1038/s41419-020-03264-5
52. Filali-Mouncef Y, Hunter C, Roccio F, et al. The ménage à trois of autophagy, lipid droplets and liver disease. Autophagy. 2022;18:50-72. doi:10.1080/15548627.2021.1895658
53. Bhat N, Esteghamat F, Chaube BK, et al. TCF7L2 transcriptionally regulates Fgf15 to maintain bile acid and lipid homeostasis through gut-liver crosstalk. FASEB J. 2022;36:e22185. doi:10.1096/fj.202101607R
54. Long Z, Cao M, Su S, et al. Inhibition of hepatocyte nuclear factor 1b induces hepatic steatosis through DPP4/NOX1-mediated regulation of superoxide. Free Radic Biol Med. 2017;113:71-83. doi:10.1016/j.freeradbiomed.2017.09.016
55. Cholico GN, Fling RR, Sink WJ, Nault R, Zacharewski T. Inhibition of the urea cycle by the environmental contaminant 2,3,7,8-tetrachlorodibenzo-p-dioxin increases serum ammonia levels in mice. J Biol Chem. 2024;300:105500. doi:10.1016/j.jbc.2023.105500
56. Petrus P, Lecoutre S, Dollet L, et al. Glutamine Links Obesity to Inflammation in Human White Adipose Tissue. Cell Metab. 2020;31:375-90. doi:10.1016/j.cmet.2019.11.019
57. Fan Z, Wang S, Meng Y, Wen C, Xu M, Li X. Butyrate Alleviates High-Fat-Induced Metabolic Disorders Partially through Increasing Systematic Glutamine. J Agric Food Chem. 2024;72:449-60. doi:10.1021/acs.jafc.3c08926
58. Stettner N, Rosen C, Bernshtein B, et al. Induction of Nitric-Oxide Metabolism in Enterocytes Alleviates Colitis and Inflammation-Associated Colon Cancer. Cell Rep. 2018;23:1962-76. doi:10.1016/j.celrep.2018.04.053
59. Han K, Singh K, Meadows AM, et al. Boosting NAD preferentially blunts Th17 inflammation via arginine biosynthesis and redox control in healthy and psoriasis subjects. Cell Rep Med. 2023;4:101157. doi:10.1016/j.xcrm.2023.101157
60. Hsieh SY, Hsu CY, He JR, et al. Identifying apoptosis-evasion proteins/pathways in human hepatoma cells via induction of cellular hormesis by UV irradiation. J Proteome Res. 2009;8:3977-86. doi:10.1021/pr900289g.

Table 1. The Mendelian randomization results of glutamine to alanine ratio and glutamine levels

Exposure	Outcome	nSNP	Method	pval	beta	Database
Glutamine levels	NAFLD	24	IVW	0.010	-0.102	IEU
Glutamine to alanine ratio	NAFLD	30	IVW	0.004	-0.144	IEU
Glutamine to alanine ratio	NAFLD	37	IVW	0.015	-0.175	FinnGen

MR: Mendelian randomization; NAFLD: non-alcoholic fatty liver disease; IVW: inverse variance weighted; SNP: single nucleotide polymorphism; pval: IVW p value.

Table 2. Sensitivity analysis results

Exposure	Outcome	pleiotropy test pval	Cochran's Q test pval	
			MR Egger	IVM
Glutamine levels (IEU)	NAFLD	0.055	0.753	0.563
Glutamine to alanine ratio (IEU)	NAFLD	0.888	0.018	0.023
Glutamine to alanine ratio (FinnGen)	NAFLD	0.139	0.097	0.069

NAFLD: non-alcoholic fatty liver disease; IVW: inverse variance weighted.

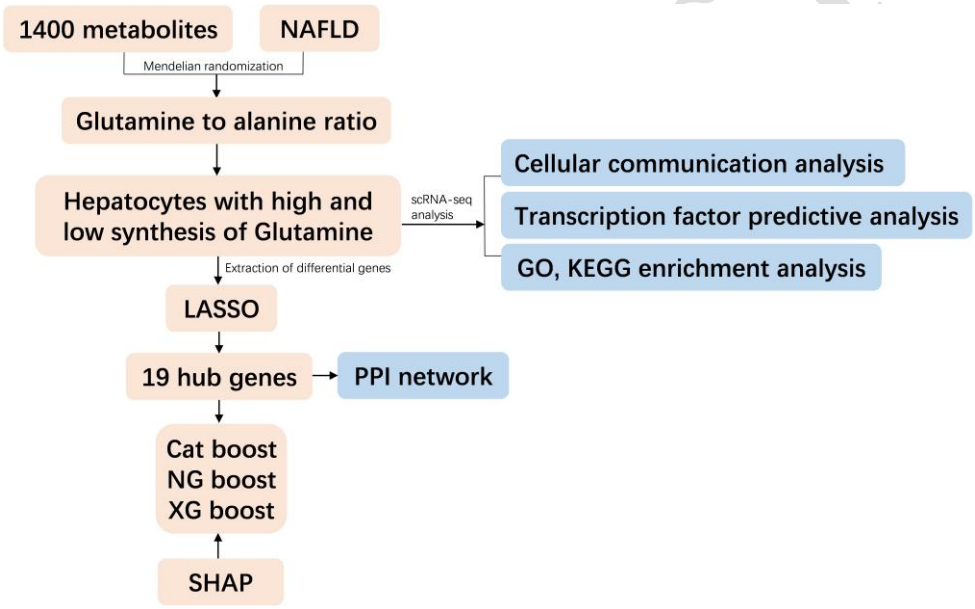

Figure 1. The research design's flow chart. LASSO: least absolute shrinkage and selection operator; PPI: protein-protein interaction

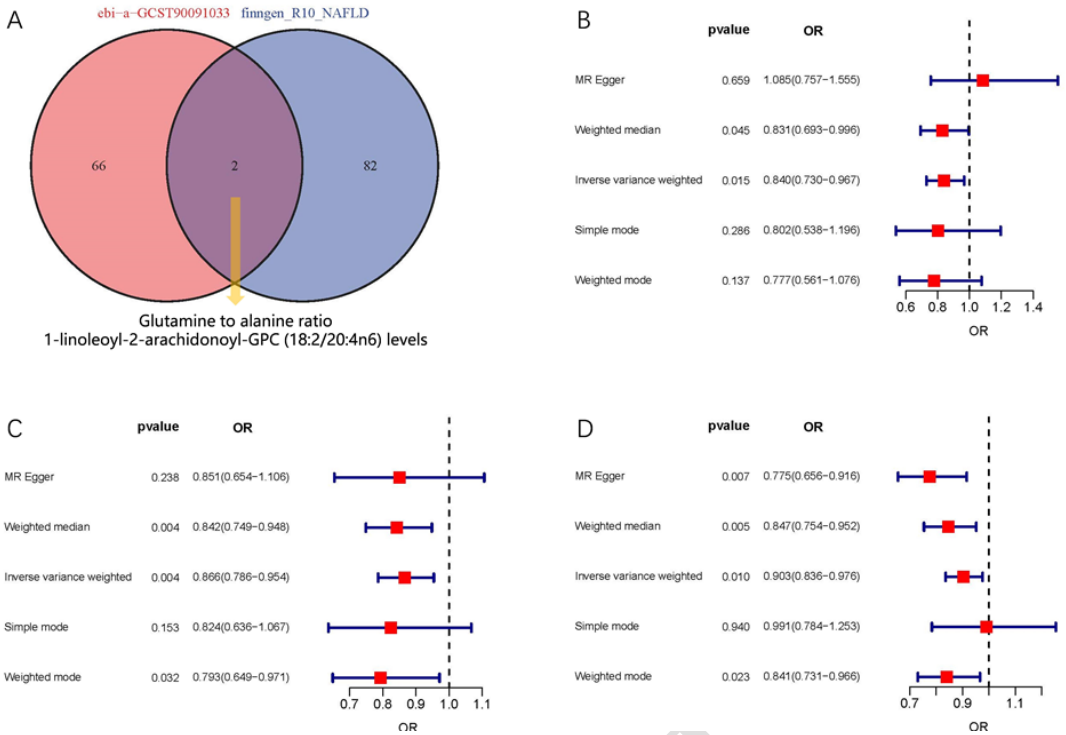

Figure 2. Result of mendelian randomization. (A) Venn diagram of 1400 serum metabolites mendelian randomization results of ebi-a-GCST90091033 and finngen_R10_NAFLD. (B) Forest plot of the results of mendelian randomization of Glutamine to alanine ratio and NAFLD in finngen_R10_NAFLD. (C) Forest plot of the results of mendelian randomization of Glutamine to alanine ratio and NAFLD in ebi-a-GCST90091033. (D) Forest plot of the results of mendelian randomization of Glutamine levels and NAFLD in ebi-a-GCST90091033. NAFLD: non-alcoholic fatty liver disease

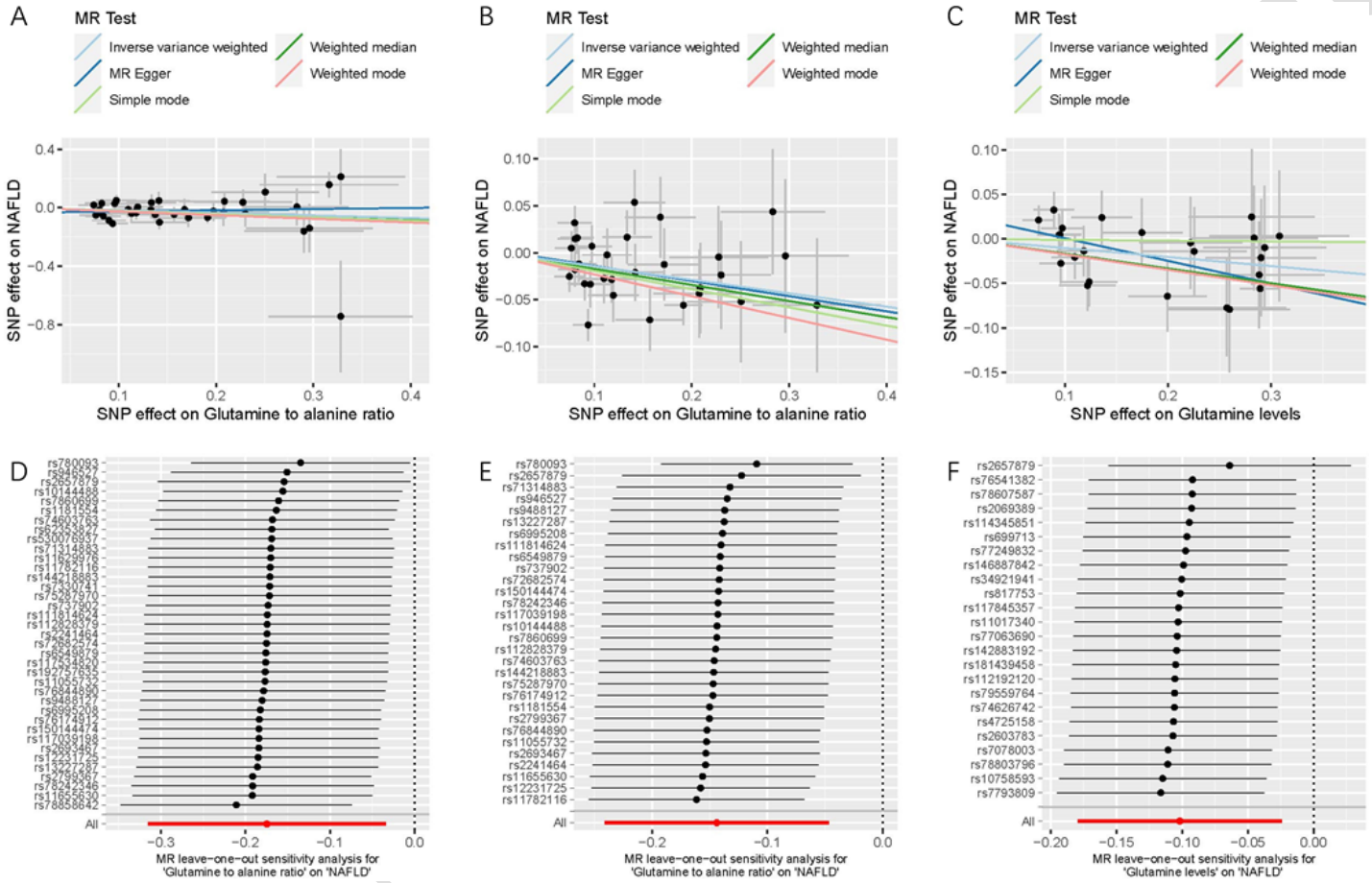

Figure 3. Scatter plot and leaveoneout plot. (A) Scatter plot of Glutamine to alanine ratio and NAFLD in finngen_R10_NAFLD. (B) Scatter plot of Glutamine to alanine ratio and NAFLD in ebi-a-GCST90091033. (C) Scatter plot of Glutamine levels and NAFLD in ebi-a-GCST90091033. (D) Leaveoneout plot of Glutamine to alanine ratio and NAFLD in finngen_R10_NAFLD. (E) Leaveoneout plot of Glutamine to alanine ratio and NAFLD in ebi-a-GCST90091033. (F) Leaveoneout plot of Glutamine levels and NAFLD in ebi-a-GCST90091033

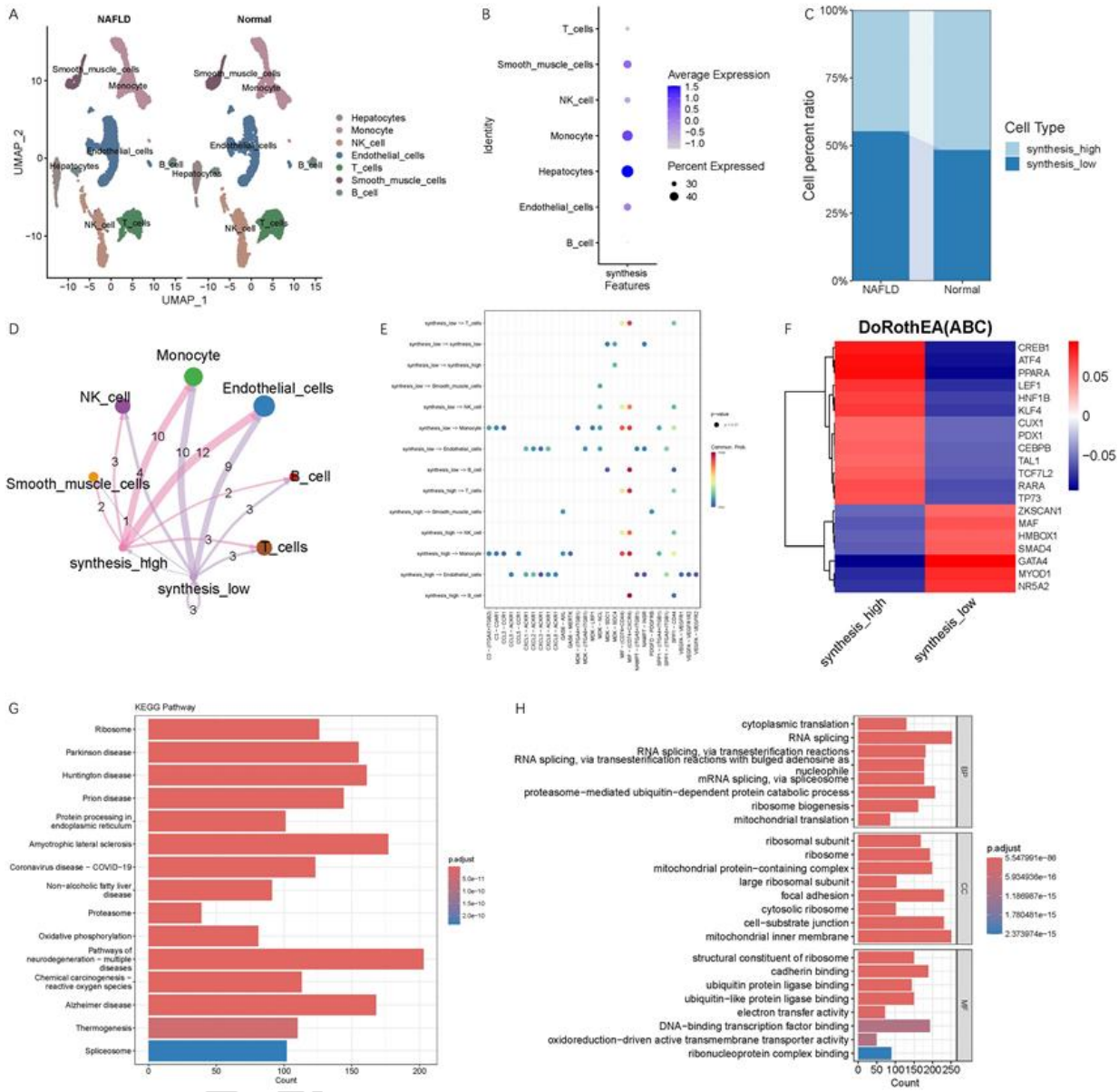

Figure 4. Results of scRNA-seq analysis. (A) UMAP of scRNA-seq; (B) The score of glutamine synthesis in different cells. Hepatocytes had the highest score; (C) The proportion of synthesis_high and synthesis_low in NAFLD group and Normal group. The proportion of synthesis_high in the NAFLD group was higher than that in the Normal group; (D)-(E) Cell communication of synthesis_high and synthesis_low; (F) Transcription factor activity prediction of synthesis_high and synthesis_low; (G)-(H) KEGG and GO enrichment analysis of differentially expressed genes in synthesis_high and synthesis_low. Synthesis_high: hepatocytes with high glutamine synthesis. Synthesis_low: hepatocytes with low glutamine synthesis. NAFLD: non-alcoholic fatty liver disease

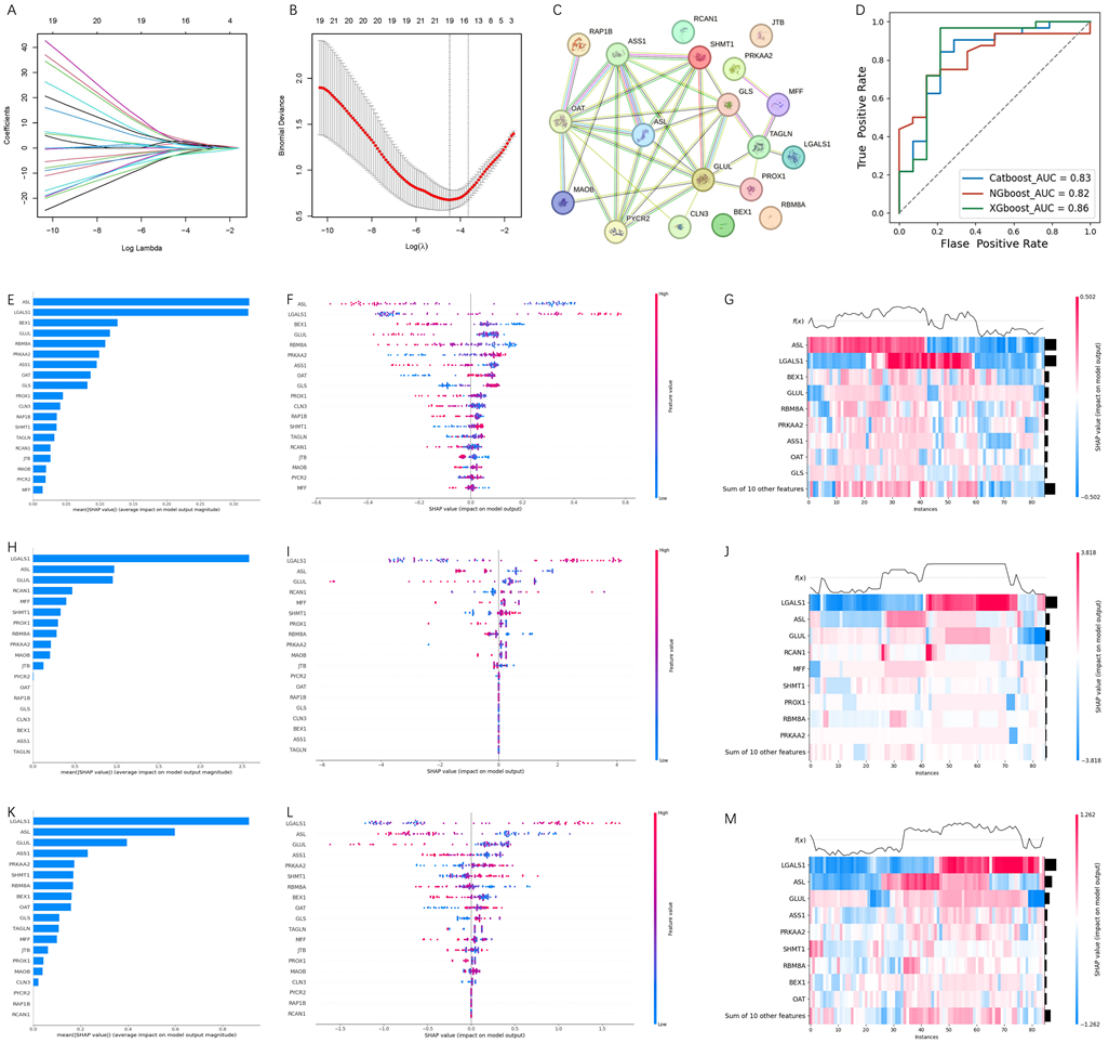

Figure 5. Machine learning and SHAP. (A) Path diagram of LASSO coefficients of hub genes in the training set. (B) Cross validation plot of LASSO regression. (C) PPI network of hub genes. (D) AUC curves of three machine learning methods in the validation set. The AUC of Catboost is 0.83, NGboost is 0.82, and XGboost is 0.86. (E)-(G) Summary Plots, Beeswarm Plots, and Heatmap Plots of Catboost. The gene with the largest contribution in Catboost is ASL. (H)-(J) Summary Plots, Beeswarm Plots, and Heatmap Plots of NGboost. The gene with the largest contribution in NGboost is LGALS1. (K)-(M) Summary Plots, Beeswarm Plots, and Heatmap Plots of XGboost. The gene with the largest contribution in XGboost is LGALS1