ORIGINAL RESEARCH

Bioinformatic Analysis of Key Biomarkers and Infiltrating Immune Cells in Nonalcoholic Fatty Liver Disease

Ge Chen, MM; Shunshun Jiang, MM; Lingling Kong, MD; Ting Yang, MM; Xutong Sun, MM; Tingting Xu, MM; Liqing Fan, MM; Qing Xie, MD; Hong Zhao, MD

ABSTRACT

Context • The liver is both the largest metabolic and the largest immune organ and is closely related to the mechanisms of disease development. Clarifying the immune environment of the NAFLD liver to determine its interactions with biomarkers would be beneficial in exploring the mechanisms of disease development.

Objective • The study aimed to identify biomarkers and immune cells associated with nonalcoholic fatty liver disease (NAFLD) and to analyze the correlation between key genes and immune cells in NAFLD, to improve the understanding of the mechanisms underlying NAFLD and provide potential therapeutic targets.

Design • The research team performed a genetic study.

Setting • The study took place at Qingdao, Shandong Province, China.

Outcome Measures • The research team: (1) obtained the NAFLD-related datasets GSE63067, GSE48452, and GSE89632 from the Gene Expression Omnibus (GEO) database; (2) analyzed immune-cell infiltrates using single-sample gene set enrichment analysis (ssGSEA) to determine the hub immune cells; (3) selected the differentially expressed genes (DEGs) between the NAFLD and normal samples and screened them to identify the hub genes; (4) evaluated the efficiency of the hub genes using receiver operating characteristic (ROC) curves; and (5) analyzed the correlations between hub genes and immune cells.

Results • The research team: (1) found 28 differential immune cells; (2) identified monocytes as the hub immune cells; (3) identified 55 DEGs; (4) comparing the top 10 genes, identified five hub genes: S100 calcium binding proteins A12 (S100A12), S100A9, S100A8, selectin L (SELL), and sex hormone binding globulin (SHBG); (5) for all five, the area under the ROC curve (AUC) was greater than 0.6—training set: AUCSA00A12 = 0.699, AUCSELL = 0.743, AUCS100A9 = 0.735, AUCSHBG = 0.752, and AUCS100A8 = 0.703; and validation set: AUCSA00A12 = 0.852, AUCSELL = 0.905, AUCS100A9 = 0.819, AUCSHBG = 0.830, and AUCS100A8 = 0.822; (6) negatively correlated SHBG with immune cells ($P > .05$, r=-0.09); and (7) positively correlated S100A12, S100A9, S100A8, and SELL with immune cells—rS100A8 = 0.40, rS100A9 = 0.50, rS100A12 = 0.38, and rSELL = 0.42, respectively.

Conclusions • Based on bioinformatic analyses, the progression of NAFLD may involve monocytes through promotion of liver inflammation. The hub genes S100A12, S100A9, S100A8, SELL, and SHBG are potential biomarkers that may be useful as diagnostic tools or therapeutic targets for NAFLD. (*Altern Ther Health Med.* 2024;30(6):276-283).

Ge Chen, MM, Qingdao Medical College, Qingdao University, and MM, Resident Physician, the First Department of Gastroenterology, Affiliated Qingdao Central Hospital of Qingdao University, Qingdao Cancer Hospital, Qingdao, Shandong, China. **Shunshun Jiang,** MM, Attending Physician; **Lingling Kong,** MD, Attending Physician; **Ting Yang,** MM, Head Nurse; **Xutong Sun,** MM, Resident Physician; **Tingting Xu,** MM, Attending Physician; **Liqing Fan,** MM, Attending Physician; **Hong Zhao,** MD, Professor, Chief Physician; the First Department of Gastroenterology, Affiliated Qingdao Central Hospital of Qingdao University, Qingdao Cancer Hospital, Qingdao, Shandong, China. **Qing Xie,** MD, Professor, Chief Physician, Department of Infectious Diseases, Ruijin Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai, China.

Corresponding author: Hong Zhao, MD
E-mail: zhaohong1968qd@163.com
Corresponding author: Qing Xie, MD
E-mail: xieqingrjh@163.com

Nonalcoholic fatty liver disease (NAFLD) is the pathological accumulation of hepatocellular fat in more than 5% of liver tissue in the absence of viral liver disease, drug intake, alcohol consumption, or other secondary cause.[1-3] Over the past century, NAFLD has emerged as an important

cause of liver disease and is likely to become a major cause of end-stage liver disease.[4,5]

NAFLD encompasses a wide range of liver dysfunctions, from simple fatty liver to inflammation and fibrosis and even cirrhosis, which increases the risk of hepatocellular carcinoma by more than 2.5-fold.[6] Unhealthy habits, such as a sedentary lifestyle and excess calorie intake, are important risk factors for NAFLD. With the increasing rate of obesity, the prevalence of NAFLD has increased worldwide from 15% in 2005 to 25% in 2010.[7,8] NAFLD is an important personal and public healthcare issue because of its high incidence and complex clinical management.[9] However, the mechanism underlying the progression of NAFLD isn't well understood.[10-12]

Treatment modalities for NAFLD include specific pharmacological treatments, such as vitamin E and pioglitazone; glucagon-like peptide 1(GLP-1) agonists; and farnesoid X receptor (FXR) and peroxisome proliferator-activated receptor (PPAR) ligands.[13] At the same time, physicians propose a healthy diet and increased physical activity as a means of disease control.

Immune System

In the presence of excess fat deposition in the enterohepatic axis and liver, the development of NAFLD is associated with the immune system. The liver is both the largest metabolic and the largest immune organ and is closely related to the mechanisms of disease development.[14-17]

Immune cells make up 10-20% of the total number of hepatocytes and of several types of the liver's non-stromal cells, including Kupffer cells, natural killer cells, and T lymphocytes; they play a crucial role in the pathogenesis of NAFLD.[18,19] Changes in the immune system, including immune-cell infiltration and the expression levels of related genes, may be important diagnostic tools that physicians could use to identify patients at risk of developing NAFLD.

The liver's immune cells process external antigen- and pathogen-rich blood from the gastrointestinal tract to maintain an immune-tolerant microenvironment.[15] Deposition of excess fat in hepatocytes alters the liver's metabolic status and immune microenvironment. Clarifying the immune environment of the NAFLD liver to determine its interactions with biomarkers would be beneficial in exploring the mechanisms of disease development.

Monocytes are highly plastic, heterogeneous immune cells that play key roles in tissue homeostasis and host defense. Peripheral blood monocytes continuously flow into a healthy liver.[20] In an injured liver, monocyte recruitment increases, promoting inflammation and fibrosis.[21,22] Gadd et al and Liaskou et al found that the number of liver monocytes was elevated in the early stages of NAFLD.[23,24] Yozgat et al found that the monocyte to high-density lipoprotein (HDL) ratio significantly increased in NAFLD and was correlated with insulin resistance.[25]

Adipose tissue is an endocrinal organ that regulates inflammation and energy.[26,27] Stanton et al found that adipose-tissue inflammation was a prerequisite for NAFLD progression in mice.[28] The chronic inflammation associated with obesity and insulin resistance influences the development of NAFLD. Infiltration of immune cells into adipose tissue can control the secretion of anti-inflammatory and pro-inflammatory factors and induce NAFLD.[28,29] Therefore, exploring the molecular mechanisms and biomarkers associated with the immune system in the liver are essential for preventing and treating NAFLD.

Sumarac-Dumanovic et al found that obesity was related to increased production of interleukin-17A (IL-17A).[30] Zuniga et al found that IL-17A knockout mice gained weight but that abnormal glucose metabolism didn't affect them.[31] Harley et al found that the IL-17 pathway was a causal factor in the development of NAFLD to steatohepatitis and that neutralization of IL17A could significantly decrease obesity-driven hepatocellular damage in wild-type mice.[32]

Genetic Influences

Genetic influences in NAFLD include variants: (1) of transmembrane 6 superfamily member 2 (TM6SF2), (2) of membrane bound O-acyltransferase domain containing 7 (MBOAT7), (3) of glucokinase regulator (GCKR), and (4) of "isoleucine to methionine substitution at position 148 in the patatin-like phospholipase domain containing 3 protein" (I148M PNPLA3).[33] Yet few studies have occurred examining the correlation between immune cells and the markers of disease.

Potential Biomarkers

S100 calcium binding proteins A9 (S100A9). S100A9 is a calcium-binding S100 protein that has two calcium-binding N-terminal E helix-C-terminal F helix. (EF)-hand motifs linked by a central hinge region. Tilg et al found that S100A9 was a diagnostic marker for noninfectious inflammatory diseases,[34] and Cai et al found a link between S100A9 in NAFLD patients and liver injury from inflammation related to diabetes mellitus.[35] Gonzalez et al suggested that other S100 proteins, such as S100A8 and S100A12, are also potential biomarkers of inflammatory diseases and obesity.[36]

Peroxisome proliferator-activated receptors (PPARs). PPARs play important roles in NAFLD.[37] However, the mechanism of PPARs and signaling pathways in the pathogenesis of NAFLD is unclear.[8,9] PPAR pathway-related genes include perilipin 1 (PLIN1), fatty acid desaturase 2 (FADS2), and fatty acid binding protein 5 (FABP5).

Carr et al found that PLIN1 was modestly higher in the livers of patients with NAFLD compared to that of healthy controls.[38] FADS2 is a rate-limiting enzyme in the synthesis of long-chain polyunsaturated fatty acids. Arendt et al found that the levels of n-3 and n-6, long-chain polyunsaturated fatty acids were significantly lower in the livers of patients with nonalcoholic steatohepatitis than in the livers of patients with simple steatosis, which may have been due to the overexpression of FADS2.[39]

Westerbacka et al found that FABP5 expression in the liver was correlated with hepatic fatty infiltration in patients with

NAFLD.[40] Thus, the PPAR signaling pathway may play a critical role in NAFLD by regulating PLIN1, FADS2, and FABP5.

Selectin L (SELL). SELL is a member of the selectin family of adhesion molecules and several leukocytes express it, especially monocytes, neutrophils, and lymphocytes, which mediate the initial attachment and subsequent rolling of leukocytes on activated endothelial cells.[41,42]

Martinez-Mier et al found that blockade of SELL could result in decreased liver enzymes in a mouse model of liver ischemia reperfusion, suggesting its critical role in liver diseases.[43] Drescher et al found that a high expression of SELL was correlated with disease activity in nonalcoholic steatohepatitis and that SELL knockout protected against diet-induced steatohepatitis in mice.[44] Therefore, SELL is a promising target for therapeutic interventions in NAFLD.

Sex hormone binding globulin (SHBG). SHBG is a liver glycoprotein associated with the regulation and transport of circulating androgens. In recent years, Toljan et al reported that SHBG was a biomarker for several diseases, such as: obesity, metabolic syndrome, polycystic ovary syndrome, osteoporosis, breast and prostate cancer.[45] In particular, Yamazaki et al found that low serum levels of SHBG were related to high levels of inflammation markers,[46] and Deswal et al found that SHBG was related to risk of obesity and NAFLD.[47] Deswal et al also suggested that re-establishment of normal SHBG levels was negatively associated with reduced hepatic fat accumulation, ,even though they have had difficulty in elucidating the mechanism by which this occurs.

Current Study

The current study aimed to identify biomarkers and immune cells associated with NAFLD and to analyze the correlation between key genes and immune cells in NAFLD, to improve the understanding of the mechanisms underlying NAFLD and provide potential therapeutic targets.

Methods

The research team performed a genetic study, which took place at Qingdao, Shandong Province, China.

Outcome measures. The research team: (1) obtained the NAFLD-related datasets GSE63067[48], GSE48452[49], and GSE89632[50] from the Gene Expression Omnibus (GEO) database[51]; (2) analyzed immune-cell infiltrates using single-sample gene set enrichment analysis (ssGSEA) to determine the hub immune cells; (3) selected the differentially expressed genes (DEGs) between the NAFLD and normal samples and screened them to identify the hub genes; (4) evaluated the efficiency of the hub genes using receiver operating characteristic (ROC) curves; and (5) analyzed the correlations between hub genes and immune cells.

Data source. Download the GSE63067, GSE48452 and GSE89632 datasets. Sequencing platforms were GPL570 [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array, GPL11532 [HuGene-1_1-st] Affymetrix Human Gene 1.1 ST Array and GPL14951 Illumina HumanHT-12 WG-DASL V4.0 R2 expression beadchip. Species in all datasets were Homo sapiens.

GSE63067 contains data from 11 disease samples and seven normal samples; GSE48452 contains data from 14 disease samples and 14 normal samples; and GSE89632 contains data from 20 disease samples and 24 normal samples.

Data pre-processing. The research team: (1) initially acquired the Series Matrix Files containing the datasets; (2) subsequently, merged datasets GSE63067 and GSE48452 using SilicoMerging 3.0, a component of the R package[52]; and (3) found a batch effect and then eliminated it employing Johnson et al's proposed technique,[53] interweaving them to form a matrix, a UMAP, that served as the training set,

Acquisition of immune-cell data. The research team: (1) used the single-sample gene set enrichment analysis (ssGSEA) algorithm, a built-in algorithm of the Gene Set Variation Analysis (GSVA) package, to convert the gene expression profile of each sample in the final expression matrix into an enrichment profile of the immune gene set; (2) selected 28 differential immune cell types from the NAFLD and normal groups, and (2) analyzed the groups' samples using the ssGSEA algorithm in the R package (GSEA 3.15)[54] to identify the infiltrating immune cells and calculate their abundance.

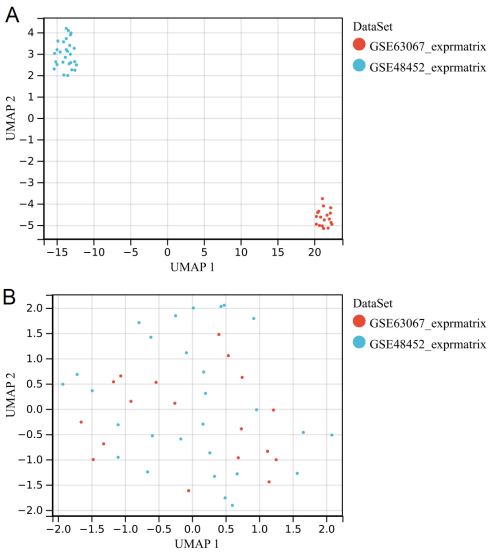
Hub immune-cell identification. The research team used three methods to screen the identified immune cells: (1) based on the analysis of immune-cell infiltration in the NAFLD and normal groups, performed a t-test in R4.2.1 and selected immune cells with $P < .05$; (2) conducted a LASSO analysis of the immune cells using the glmnet, version 4.1-4, R package[55]; and (3) screened the immune cells using the support vector machine-recursive feature elimination (SVM-RFE) algorithm. The team considered the intersection of the data from the three methods to be the hub immune cells.

Identification of differentially expressed genes (DEGs). Based on the expression matrix obtained above, the research team used the R package limma 3.10.3[56] to compare gene expression between the NAFLD and normal groups and obtain the genes' corresponding p values and log2FCs. The team set the thresholds for differential expression at $P < .05$ and $|\log2FC| \geq 0.585$. The team divided the samples into NAFLD group and Normal sample group based on the expression levels of the DEGs.

Function analysis of the DEGs. The research team: (1) used the Kyoto Encyclopedia of Genes and Genomes (KEGG) rest application programming interface (API) to obtain gene annotations using the latest KEGG pathways; (2) using the pathways as the background, mapped the DEGs to the background set for enrichment analysis using the R package clusterProfiler 3.15[57]; (3) also downloaded c5.go.bp.v7.4.symbols.gmt from the Molecular Signatures Database[58] to use as the gene ontology (GO) background; and (4) mapped the DEGs to the background for enrichment analysis using the R package clusterProfiler 3.15. $P < .05$ was the threshold to select significantly enriched pathways and functions.

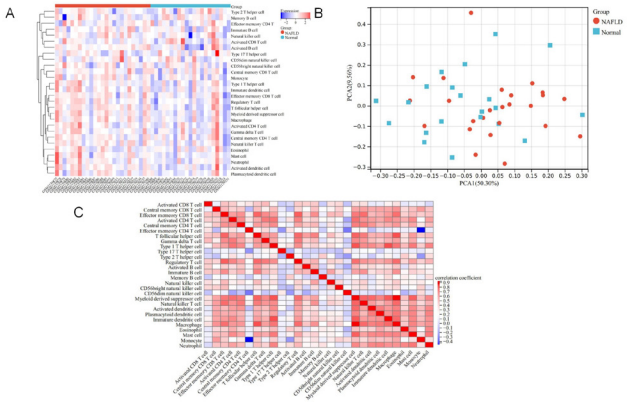
Hub gene identification. The research team: (1) constructed a protein-protein interaction (PPI) network of pathway genes using the Search Tool for the Retrieval of Interacting Genes/Proteins[59] and visualized the network

Figure 1. Density Map Before (**A**) and After (**B**) Removal of the Batch Effect



Abbreviations: GSE, gene set enrichment.

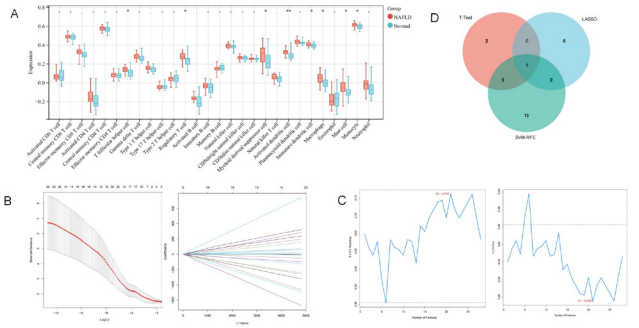
Figure 2. Heatmap of Immune Cell Clustering and Correlation of 28 Immune Cells. **A** shows the cluster heatmap; **B** shows the PCA plot; and **C** shows the correlation heatmap. PCA1 and PCA2 are the names of the matrices of different dimensions in the space formed after dimensionality reduction of the data, and 50.30% and 9.56% denote the contribution of the principal components. The values of the horizontal and vertical axes denote the range of the loading coefficients.



Abbreviations: CD4, cluster of differentiation 4; NAFLD, nonalcoholic fatty liver disease; PCA, principal component analysis.

using Cytoscape 3.8.2[60]; (2) selected the top 10 important genes in the network using four topological algorithms in cytoHubba: maximal clique centrality (MCC), degree, maximum neighborhood component (MNC), and edge percolated component (EPC); (3) identified the intersecting genes from the four algorithms as hub genes; (4) verified the differential expression of the candidate hub genes using a *t* test for both the training and validation sets; and (5) visualized the expression levels of the five hub genes in the training and verification sets using heatmaps and box plots.

Figure 3. NAFLD-associated Differential Immune Cells. Figure 3A shows the diagram of differences in immune cells as detected by *t* test; Figure 3B shows the LASSO assay parameter diagrams for immune cell screening; Figure 3C shows the immune cell screening by SVM-RFE; and Figure 3D shows the Venn diagram of immune cells obtained from *t* test, LASSO, and SVM-RFE.



Abbreviations: CD, cluster of differentiation; LASSO, least absolute shrinkage and selection operator; NAFLD, nonalcoholic fatty liver disease; PCA, principal component analysis; SVM-RFE, support vector machine-recursive feature elimination.

Efficacy evaluation of hub genes. The research team: (1) to evaluate the accuracy of the gene signature, drew ROC curves for the hub genes using pROC 1.7.2[61] in R and (2) calculated Pearson's correlation coefficients to determine the correlations between the hub genes and hub immune cells.

RESULTS

Data Preprocessing

The UMAP showed that the initial sample distributions of the two datasets were significantly different, indicating a batch effect (Figure 1). After its removal, the samples from the two datasets were interwoven in UMAP.

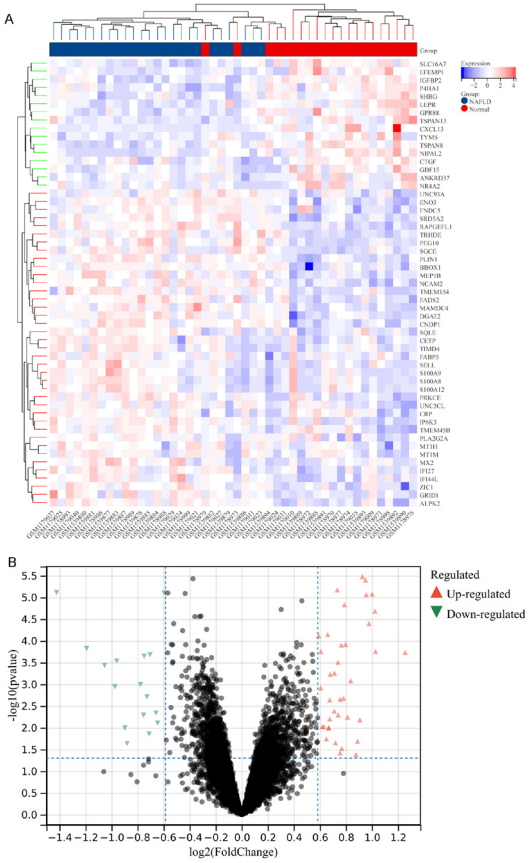
Immune-cell Acquisition

Figures 2A, 2B, and 2C show the cluster heatmap, principal component analysis (PCA) diagram, and correlation heatmap of the 28 differential immune-cell types, respectively. The cluster heatmap shows the different expression levels for the different immune cells in the NAFLD and normal groups. In the PCA diagram, the immune cells in each group were relatively clustered with good repetition in the respective groups. PCA1 and PCA2 are the names of the matrices of different dimensions in the space formed after dimensionality reduction of the data, and 50.30% and 9.56% denote the contribution of the principal components. The correlation heatmap shows that the 28 immune cells were positively correlated.

Hub Immune Cells

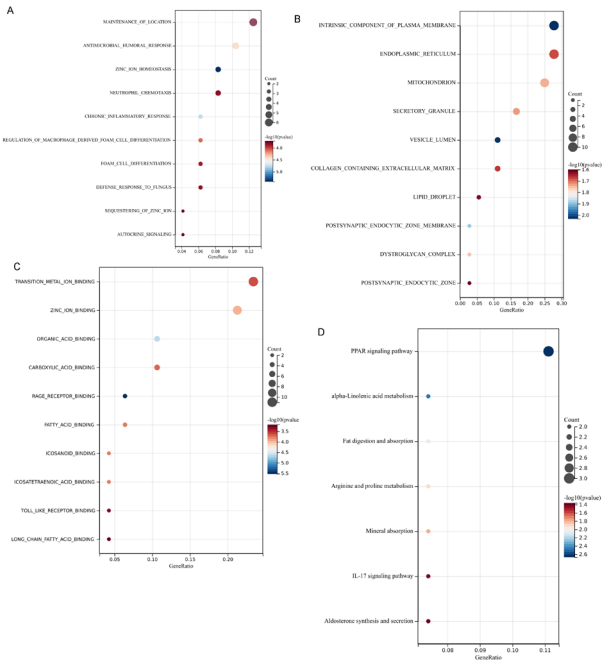
The T-test analysis identified eight differential, immune-cell types between the NAFLD and normal groups (Figure 3A). Figure 3B shows the nine immune cell types from the LASSO analysis. Figure 3C shows the 21 immune-cell types from the analysis using the SVM-RFE algorithm. The combined three methods identified only one cell type, monocytes (Figure 3D).

Figure 4. Demonstration of Differential Gene Expression. **A** shows the heatmap of differentially expressed genes (DEGS), and **B** shows the Volcano plot of DEGS.



Abbreviations: ALPK2, alpha kinase 2; ANKRD37, ankyrin repeat domain 37; BBOX1, gamma-butyrobetaine hydroxylase 1; CETP, cholesteryl ester transfer protein; CNDP1, carnosine dipeptidase 1; CRP, C-reactive protein; CTGF, connective tissue growth factor; CXCL13, CXC motif chemokine ligand 13; DGAT2, diacylglycerol O-acyltransferase 2; EFEMP1, EGF containing fibulin extracellular matrix protein 1; ENO3, enolase 3; FADS2, fatty acid desaturase 2; FABP5, peroxidase; FNDC5, fibronectin type III domain-containing protein 5; GDF15, growth/differentiation factor-15; GPR88, G-protein coupled receptor 88; GRID1, glutamate receptor, ionotropic, delta 1; IF127, interferon alpha-inducible protein 27; IF144L, interferon induced protein 44 like; IGFBP2, insulin-like growth factor-binding protein 2; IP6K3, inositol hexakisphosphate kinase 3; LEPR, leptin receptor; MAMDC4, MAM domain containing 4; MEP1B, meprin A subunit beta; MT1H, metallothionein 1H; MT1M, metallothionein 1M; MX2, myxovirus resistance 2; NAFLD, nonalcoholic fatty liver disease; NCAM2, neural cell adhesion molecule 2; NIPAL2, nuclear interacting partner of anaplastic lymphoma (NIPAL)-like domain containing 2; NR4A2, nuclear receptor subfamily 4 group A member 2; P4HA1, prolyl 4-hydroxylase subunit alpha 1; PEG10, paternally expressed gene 10; PLA2G2A, phospholipase A2 group IIA; PLIN1, perilipin 1; PRKCE, protein kinase C epsilon; RAPGEFL1, rap guanine nucleotide exchange factor 1; S100A8, S100 calcium-binding protein A8; S100A9, S100 calcium-binding protein A9; S100A12, S100 calcium-binding protein A12; SELL, selectin L; SGCE, sarcoglycan epsilon; SHBG, sex hormone binding globulin; SLC16A7, solute carrier family 16 member 7; SQLE, squalene epoxidase; SRD5A2, steroid 5-alpha reductase 2; TIMD4, T-cell immunoglobulin and mucin domain containing 4; TMEM45B, transmembrane protein 45B; TMEM154, transmembrane protein 154; TRHDE, thyrotropin releasing hormone degrading enzyme; TSPAN8, tetraspanin-8; TSPAN8, tetraspanin-13; TYMS, thymidylate synthase; UNC5CL, UNC-5, Unc-5 homolog C; UNC93A, Unc-93 homolog A; ZIC1, Zic family member 1.

Figure 5. Enrichment Analysis of Differentially Expressed Genes (DEGs). **A** shows the biological processes; **B** shows the cellular components; **C** shows the molecular functions; and **D** shows the pathways enriched based on the DEGs.



Abbreviations: PPAR, peroxisome proliferator-activated receptors

DEGs and Function Analysis

Differential analysis identified 55 DEGs. Figure 4A shows a heat map of the genes. Figure 4B shows a volcano plot, with 39 upregulated and 16 downregulated DEGs.

Functional enrichment analysis showed that the genes were significantly associated with 16 biological processes, such as maintenance of location, antimicrobial humoral response, neutrophil chemotaxis, and chronic inflammatory response. We have selected the top 10 pathways to showcase (Figure 5A). 299 cellular components, such as intrinsic component of plasma membrane, endoplasmic reticulum, and mitochondrion. We have selected the top 10 pathways to showcase (Figure 5B). 63 molecular functions, such as zinc ion binding, transition metal ion binding, and organic acid binding. We have selected the top 7 pathways to showcase (Figure 5C). The DEGs participated in seven pathways, including the PPAR signaling pathway, alpha-linolenic acid metabolism, interleukin 17 (IL-17) signaling pathway, and fat digestion and absorption (Figure 5D).

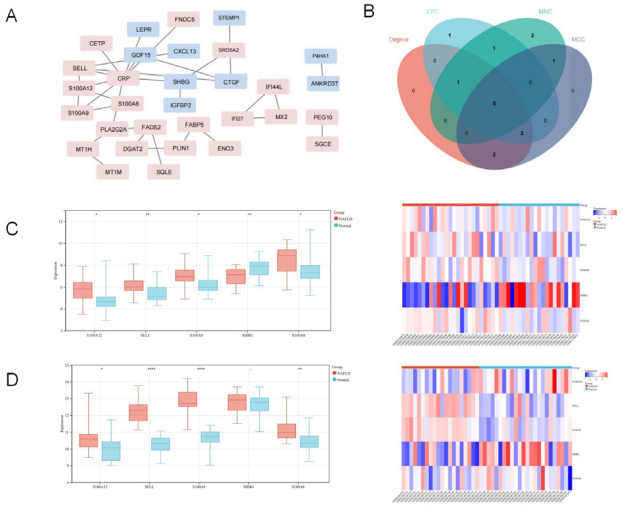
Hub Identification and Validation

The constructed PPI network included 31 nodes (Figure 6A and Table 1). Figure 6B shows that the comparison of the top 10 genes from the four topological analysis algorithms found five hub genes: S100A12, S100A9, S100A8, SELL, and SHBG.

PPI Network

As Figures 6C and 6D show, the genes' expression levels were significantly different between the NAFLD and normal groups, groups in both the training and verification sets.

Figure 6. Four Topological Analysis Algorithms. **A** shows the PPI network in NAFLD. The red squares represent upregulated genes, and the blue represent downregulated genes. **B** shows the Venn diagram of the genes identified using the four topology analysis algorithms; **C** shows the box plots and heatmaps of hub gene expression in the training sets; and Figure 6D shows the box plots and heatmaps of hub gene expression in the validation sets.



Abbreviations: ANKRD37, ankyrin repeat domain 37; CETP, cholesteryl ester transfer protein; CRP, C-reactive protein; CTGF, connective tissue growth factor; CXCL13, CXC motif chemokine ligand 13; DGAT2, diacylglycerol O-acyltransferase 2; EFEMP1, EGF containing fibulin extracellular matrix protein 1; ENO3, enolase 3; EPC, edge percolated component; FABP5, peroxidase; FADS2, fatty acid desaturase 2; FNDC5, fibronectin type III domain-containing protein 5; GDF15, growth/differentiation factor-15; IF127, interferon alpha-inducible protein 27; IF144L, interferon induced protein 44 like; IGFBP2, insulin-like growth factor-binding protein 2; LEPR, leptin receptor; MCC, maximal clique centrality; MNC, maximum neighborhood component; MT1H, metallothionein 1H; MT1M, metallothionein 1M; MX2, myxovirus resistance 2; NAFLD, nonalcoholic fatty liver disease; PEG10, paternally expressed gene 10; PLA2G2A, phospholipase A2 group IIA; P4HA1, prolyl 4-hydroxylase subunit alpha 1; PLIN1, perilipin 1; S100A8, S100 calcium-binding protein A8; S100A9, S100 calcium-binding protein A9; S100A12, S100 calcium-binding protein A12; SELL, selectin L; SHBG, sex hormone binding globulin; SGCE, sarcoglycan epsilon; SQLE, squalene epoxidase; SRD5A2, steroid 5-alpha reductase 2.

Efficacy of Hub Genes

Figure 7A shows that the areas under the curve (AUC) were greater than 0.6 for all five genes in the training set: S100A12 (0.699), S100A9 (0.735), S100A8 (0.703), SELL (0.743), and SHBG (0.752). In the validation set, the AUCs for the five genes were greater than 0.8: S100A12 (0.852), S100A9 (0.819), S100A8 (0.822), SELL (0.905), and SHBG (0.830). (Figure 7B).

Correlation of hub genes and immune cells

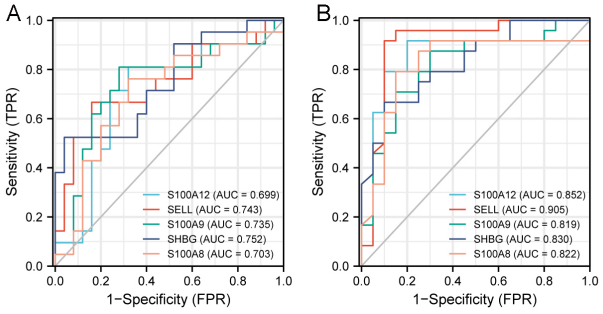
Figure 8 shows the correlation analysis of the selected hub genes and immune cells. SHBG was significantly negatively correlated with immune cells ($P > .05$, r=-0.09), whereas the S100A12, S100A9, S100A8, and SELL were significantly

Table 1. The Nodes in the Protein-protein Interaction (PPI) Network

Name	MCC	MNC	Degree	EPC
ANKRD37	1	1	1	1.497
CETP	1	1	1	4.988
CRP	10	3	9	8.847
CTGF	3	1	3	6.062
CXCL13	1	1	1	4.368
DGAT2	2	1	2	3.668
EFEMP1	1	1	1	3.537
ENO3	1	1	1	2.09
FABP5	2	1	2	2.636
FADS2	3	1	3	4.881
FNDC5	1	1	1	5.176
GDF15	2	1	2	6.353
IFI27	2	2	2	2.239
IFI44L	2	2	2	2.234
IGFBP2	1	1	1	3.537
LEPR	1	1	1	5.163
MT1H	2	1	2	4.124
MT1M	1	1	1	2.694
MX2	2	2	2	2.217
P4HA1	1	1	1	1.497
PEG10	1	1	1	1.461
PLA2G2A	3	1	3	6.314
PLIN1	2	1	2	2.991
S100A12	6	4	4	7.765
S100A8	2	2	2	6.055
S100A9	4	3	3	7.001
SELL	4	2	4	7.408
SGCE	1	1	1	1.461
SHBG	3	1	3	5.619
SQLE	1	1	1	3.09
SRD5A2	1	1	1	3.291

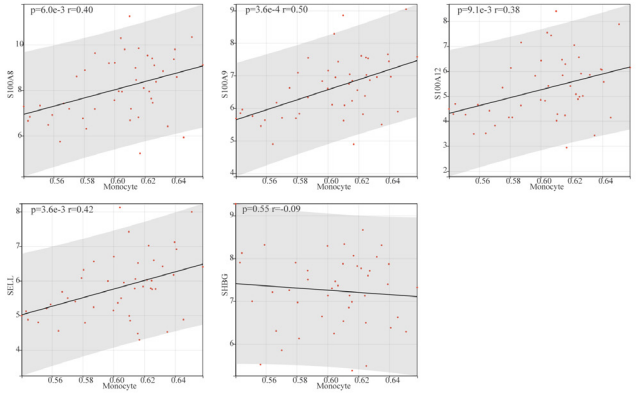
Abbreviations: ANKRD37, ankyrin repeat domain 37; CETP, cholesteryl ester transfer protein; CRP, C-reactive protein; CTGF, connective tissue growth factor; CXCL13, CXC motif chemokine ligand 13; DGAT2, diacylglycerol O-acyltransferase 2; EFEMP1, EGF containing fibulin extracellular matrix protein 1; ENO3, enolase 3; EPC, edge percolated component; FABP5, fatty acid-binding protein, epidermal; FADS2, fatty acid desaturase 2; FNDC5, fibronectin type III domain-containing protein 5; GDF15, growth/differentiation factor-15; IFI27, interferon alpha-inducible protein 27; IFI44L, interferon induced protein 44 like; IGFBP2, insulin-like growth factor-binding protein 2; LEPR, leptin receptor; MCC, maximal clique centrality; MNC, maximum neighborhood component; MT1H, metallothionein 1H; MT1M, metallothionein 1M; MX2, myxovirus resistance 2; P4HA1, prolyl 4-hydroxylase subunit alpha 1; PEG10, paternally expressed gene 10; PLA2G2A, phospholipase A2 group IIA; PLIN1, perilipin 1; S100A12, S100 calcium-binding protein A12; S100A8, S100 calcium-binding protein A8; S100A9, S100 calcium-binding protein A9; SELL, selectin L; SGCE, sarcoglycan epsilon; SHBG, sex hormone binding globulin; SQLE, squalene epoxidase; SRD5A2, steroid 5-alpha reductase 2.

Figure 7. ROC Curves of the Hub Genes the training set (**A**) and the validation set (**B**)



Abbreviations: FPR, false positive rate; S100A8, S100 calcium-binding protein A8; S100A9, S100 calcium-binding protein A9; S100A12, S100 calcium-binding protein A12; SELL, selectin L; SHBG, sex hormone binding globulin; TPR, true positive rate.

Figure 8. Correlation Analysis of Hub Genes—S100A12, S100A9, S100A8, SELL, and SHBG—and Monocytes**.** SHBG was negatively correlated with immune cells ($P > .05$, r = -0.9), whereas the four other genes were positively correlated with immune cells ($P < .05$, r > 0.30).



Abbreviations: S100A8, S100 calcium-binding protein A8; S100A9, S100 calcium-binding protein A9; S100A12, S100 calcium-binding protein A12; SELL, selectin L; SHBG, sex hormone binding globulin

positively correlated with immune cells, with all $P < .05$, rS100A8=0.40, rS100A9=0.50, rS100A12=0.38, rSELL=0.42).

DISCUSSION

In accordance with the prior studies,[34-39] the current study found that monocytes were higher in NAFLD samples compared to the levels in the normal controls. The results suggest a critical role for monocytes in NAFLD development and progression.

Among the four genes with significant positive correlations with monocytes, S100A9 and S100A8 were significantly enriched for the IL-17 signaling pathway.[62]

Thus, the current research hypothesizes that S100A9 and S100A8 may be involved in NAFLD through the IL-17 signaling pathway. However, the mechanism of influence still needs to be further verified in clinical or animal experiments. The present study also found that SELL was upregulated in the NAFLD group and was positively correlated with monocytes.

In addition to the four positively correlated genes, the current study found that SHBG was negatively correlated with monocytes, with the SHBG expression being lower in the NAFLD group than in with the normal group, suggesting that reduced SHBG expression may be a biomarker of NAFLD.

There were a few limitations to this study: first, it was only a preliminary analysis from GEO database, no further mechanism data were shown. In the second place, the inherent limitations of the database used to select platforms and microarray data sets might have led to some targets being missed. The majority of our research results were based on GEO public databases and already published data, however, because of sample size and platform limitations, further biological studies and clinical trials were required to confirm our findings.

CONCLUSIONS

Based on bioinformatic analyses, the progression of NAFLD may involve monocytes through promotion of liver inflammation. The hub genes S100A12, S100A9, S100A8, SELL, and SHBG are potential biomarkers that may be useful as diagnostic tools or therapeutic targets for NAFLD.

AVAILABILITY OF DATA AND MATERIALS

All data included in this study are available upon request by contacting the corresponding author.

AUTHORS' DISCLOSURE STATEMENT

The authors declare that they have no conflicts of interest related to the study.

REFERENCES

1. Bellentani S. The epidemiology of non-alcoholic fatty liver disease. [J]. *Liver Int*. 2017;37(S1) (suppl 1):81-84. doi:10.1111/liv.13299
2. Friedman SL, Neuschwander-Tetri BA, Rinella M, Sanyal AJ. Mechanisms of NAFLD development and therapeutic strategies. [J]. *Nat Med*. 2018;24(7):908-922. doi:10.1038/s41591-018-0104-9
3. Younossi Z, Anstee QM, Marietti M, et al. Global burden of NAFLD and NASH: trends, predictions, risk factors and prevention. [J]. *Nat Rev Gastroenterol Hepatol*. 2018;15(1):11-20. doi:10.1038/nrgastro.2017.109
4. Saab S, Manne V, Nieto J, Schwimmer JB, Chalasani NP. Nonalcoholic fatty liver disease in Latinos [J]. *Clin Gastroenterol Hepatol*. 2016;14(1):5-12. doi:10.1016/j.cgh.2015.05.001
5. Kanwal F, Kramer J R, DUAN Z, et al. Trends in the burden of nonalcoholic fatty liver disease in a United States cohort of veterans [J]. Clinical Gastroenterology and *Hepatology*, 2016, 14(2): 301-8. e2.
6. Neuschwander-Tetri BA. Hepatic lipotoxicity and the pathogenesis of nonalcoholic steatohepatitis: the central role of nontriglyceride fatty acid metabolites. [J]. *Hepatology*. 2010;52(2):774-788. doi:10.1002/hep.23719
7. Younossi ZM, Koenig AB, Abdelatif D, Fazel Y, Henry L, Wymer M. Global epidemiology of nonalcoholic fatty liver disease-Meta-analytic assessment of prevalence, incidence, and outcomes. [J]. *Hepatology*. 2016;64(1):73-84. doi:10.1002/hep.28431
8. Younossi Z, Anstee QM, Marietti M, et al. Global burden of NAFLD and NASH: trends, predictions, risk factors and prevention. [J]. *Nat Rev Gastroenterol Hepatol*. 2018;15(1):11-20. doi:10.1038/nrgastro.2017.109
9. Lindenmeyer CC, McCullough AJ. The natural history of nonalcoholic fatty liver disease—an evolving view [J]. *Clin Liver Dis*. 2018;22(1):11-21. doi:10.1016/j.cld.2017.08.003
10. Anstee QM, Targher G, Day CP. Progression of NAFLD to diabetes mellitus, cardiovascular disease or cirrhosis. [J]. *Nat Rev Gastroenterol Hepatol*. 2013;10(6):330-344. doi:10.1038/nrgastro.2013.41
11. Loomba R, Friedman SL, Shulman GI. Mechanisms and disease consequences of nonalcoholic fatty liver disease. [J]. *Cell*. 2021;184(10):2537-2564. doi:10.1016/j.cell.2021.04.015
12. Pafili K, Roden M. Nonalcoholic fatty liver disease (NAFLD) from pathogenesis to treatment concepts in humans. [J]. *Mol Metab*. 2021;50:101122. doi:10.1016/j.molmet.2020.101122
13. Paternostro R, Trauner M. Current treatment of non-alcoholic fatty liver disease. [J]. *J Intern Med*. 2022;292(2):190-204. doi:10.1111/joim.13531
14. Heymann F, Tacke F. Immunology in the liver--from homeostasis to disease. [J]. *Nat Rev Gastroenterol Hepatol*. 2016;13(2):88-110. doi:10.1038/nrgastro.2015.200
15. Jenne CN, Kubes P. Immune surveillance by the liver. [J]. *Nat Immunol*. 2013;14(10):996-1006. doi:10.1038/ni.2691
16. Reinke H, Asher G. Circadian Clock Control of Liver Metabolic Functions. [J]. *Gastroenterology*. 2016;150(3):574-580. doi:10.1053/j.gastro.2015.11.043
17. Trefts E, Gannon M, Wasserman DH. The liver. [J]. *Curr Biol*. 2017;27(21):R1147-R1151. doi:10.1016/j.cub.2017.09.019
18. Racanelli V, Rehermann B. The liver as an immunological organ. [J]. *Hepatology*. 2006;43(2) (suppl 1):S54-S62. doi:10.1002/hep.21060
19. Feng D. The alteration of immune cells in the pathogenesis of non-alcoholic fatty liver disease and non-alcoholic steatohepatitis [J]. *Liver Res*. 2020;4(1):23-27. doi:10.1016/j.livres.2020.02.003
20. Zimmermann HW, Bruns T, Weston CJ, et al. Bidirectional transendothelial migration of monocytes across hepatic sinusoidal endothelium shapes monocyte differentiation and regulates the balance between immunity and tolerance in liver. [J]. *Hepatology*. 2016;63(1):233-246. doi:10.1002/hep.28285
21. Melino M, Gadd VL, Alexander KA, et al. Spatiotemporal characterization of the cellular and molecular contributors to liver fibrosis in a murine hepatotoxic-injury model [J]. *Am J Pathol*. 2016;186(3):524-538. doi:10.1016/j.ajpath.2015.10.029
22. Ramachandran P, Pellicoro A, Vernon MA, et al. Differential Ly-6C expression identifies the recruited macrophage phenotype, which orchestrates the regression of murine liver fibrosis. [J]. *Proc Natl Acad Sci USA*. 2012;109(46):E3186-E3195. doi:10.1073/pnas.1119964109
23. Gadd VL, Skoien R, Powell EE, et al. The portal inflammatory infiltrate and ductular reaction in human nonalcoholic fatty liver disease. [J]. *Hepatology*. 2014;59(4):1393-1405. doi:10.1002/hep.26937
24. Liaskou E, Zimmermann HW, Li KK, et al. Monocyte subsets in human liver disease show distinct phenotypic and functional characteristics. [J]. *Hepatology*. 2013;57(1):385-398. doi:10.1002/hep.26016
25. Yozgat A, Ekmen N, Kasapoglu B, Unsal Y, Sokmen FC, KEKiLLi M. Monocyte/HDL ratio in non-alcoholic hepatic steatosis [J]. *Arq Gastroenterol*. 2021;58(4):439-442.
26. Mathis D. Immunological goings-on in visceral adipose tissue. [J]. *Cell Metab*. 2013;17(6):851-859. doi:10.1016/j.cmet.2013.05.008
27. Zwick RK, Guerrero-Juarez CF, Horsley V, Plikus MV. Anatomical, Physiological, and Functional Diversity of Adipose Tissue. [J]. *Cell Metab*. 2018;27(1):68-83. doi:10.1016/j.cmet.2017.12.002
28. Stanton MC, Chen S-C, Jackson JV, et al. Inflammatory Signals shift from adipose to liver during high fat feeding and influence the development of steatohepatitis in mice. [J]. *J Inflamm (Lond)*. 2011;8(1):8. doi:10.1186/1476-9255-8-8
29. Poitou C, Perret C, Mathieu F, et al. Bariatric surgery induces disruption in inflammatory signaling pathways mediated by immune cells in adipose tissue: a RNA-Seq study [J]. *PLoS One*. 2015;10(5):e0125718. doi:10.1371/journal.pone.0125718
30. Sumarac-Dumanovic M, Stevanovic D, Ljubic A, et al. Increased activity of interleukin-23/interleukin-17 proinflammatory axis in obese women. [J]. *Int J Obes (Lond)*. 2009;33(1):151-156. doi:10.1038/ijo.2008.216

31. Zúñiga LA, Shen WJ, Joyce-Shaikh B, et al. IL-17 regulates adipogenesis, glucose homeostasis, and obesity. [J]. *J Immunol*. 2010;185(11):6947-6959. doi:10.4049/jimmunol.1001269
32. Harley IT, Stankiewicz TE, Giles DA, et al. IL-17 signaling accelerates the progression of nonalcoholic fatty liver disease in mice. [J]. *Hepatology*. 2014;59(5):1830-1839. doi:10.1002/hep.26746
33. Eslam M, Valenti L, Romeo S. Genetics and epigenetics of NAFLD and NASH: clinical impact. [J]. *J Hepatol*. 2018;68(2):268-279. doi:10.1016/j.jhep.2017.09.003
34. Tilg H, Moschen AR, Roden M. NAFLD and diabetes mellitus. [J]. *Nat Rev Gastroenterol Hepatol*. 2017;14(1):32-42. doi:10.1038/nrgastro.2016.147
35. Cai Q, Zhu J, Cui X, et al. S100A9 promotes inflammatory response in diabetic nonalcoholic fatty liver disease. [J]. *Biochem Biophys Res Commun*. 2022;618:127-132. doi:10.1016/j.bbrc.2022.06.026
36. Gonzalez LL, Garrie K, Turner MD. Role of S100 proteins in health and disease. *Biochim Biophys Acta Mol Cell Res*. 2020;1867(6):118677. doi:10.1016/j.bbamcr.2020.118677
37. Li Y, Wang C, Lu J, et al. PPAR δ inhibition protects against palmitic acid-LPS induced lipidosis and injury in cultured hepatocyte L02 cell. [J]. *Int J Med Sci*. 2019;16(12):1593-1603. doi:10.7150/ijms.37677
38. Carr RM, Dhir R, Mahadev K, Comerford M, Chalasani NP, Ahima RS. Perilipin staining distinguishes between steatosis and nonalcoholic steatohepatitis in adults and children [J]. *Clin Gastroenterol Hepatol*. 2017;15(1):145-147. doi:10.1016/j.cgh.2016.08.023
39. Arendt BM, Comelli EM, Ma DW, et al. Altered hepatic gene expression in nonalcoholic fatty liver disease is associated with lower hepatic n-3 and n-6 polyunsaturated fatty acids. [J]. *Hepatology*. 2015;61(5):1565-1578. doi:10.1002/hep.27695
40. Westerbacka J, Kolak M, Kiviluoto T, et al. Genes involved in fatty acid partitioning and binding, lipolysis, monocyte/macrophage recruitment, and inflammation are overexpressed in the human fatty liver of insulin-resistant subjects. [J]. *Diabetes*. 2007;56(11):2759-2765. doi:10.2337/db07-0156
41. Fotis L, Giannakopoulos D, Stamogiannou L, Xatzipsalti M. Intercellular cell adhesion molecule-1 and vascular cell adhesion molecule-1 in children. Do they play a role in the progression of atherosclerosis? [J]. *Hormones (Athens)*. 2012;11(2):140-146. doi:10.14310/horm.2002.1340
42. Ivetic A, Hoskins Green HL, Hart SJ. L-selectin: a major regulator of leukocyte adhesion, migration and signaling [J]. *Front Immunol*. 2019;10:1068. doi:10.3389/fimmu.2019.01068
43. Martinez-Mier G, Toledo-Pereyra LH, McDuffie E, Warner RL, Ward PA. L-Selectin and chemokine response after liver ischemia and reperfusion. [J]. *J Surg Res*. 2000;93(1):156-162. doi:10.1006/jsre.2000.5954
44. Drescher HK, Schippers A, Rosenhain S, et al. L-selectin/CD62L is a key driver of non-alcoholic steatohepatitis in mice and men [J]. *Cells*. 2020;9(5):1106. doi:10.3390/cells9051106
45. Toljan K, Grgić F, Pavičić Baldani D, et al. Sex hormone binding globulin (SHBG) as a marker of clinical disorders [J]. *Coll Antropol*. 2016;40(3):199-209.
46. Yamazaki H, Kushiyama A, Sakoda H, et al. Protective effect of sex hormone-binding globulin against metabolic syndrome: in vitro evidence showing anti-inflammatory and lipolytic effects on adipocytes and macrophages [J]. *Mediators of inflammation*, 2018, 2018.
47. Deswal R, Yadav A, Dang AS. Sex hormone binding globulin - an important biomarker for predicting PCOS risk: A systematic review and meta-analysis. [J]. *Syst Biol Reprod Med*. 2018;64(1):12-24. doi:10.1080/19396368.2017.1410591
48. Ye QH, Qin LX, Forgues M, et al. Predicting hepatitis B virus-positive metastatic hepatocellular carcinomas using gene expression profiling and supervised machine learning. *Nat Med*. 2003;9(4):416-423. doi:10.1038/nm843
49. 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. [J]. *Cell Metab*. 2013;18(2):296-302. doi:10.1016/j.cmet.2013.07.004
50. Arendt BM, Comelli EM, Ma DW, et al. Altered hepatic gene expression in nonalcoholic fatty liver disease is associated with lower hepatic n-3 and n-6 polyunsaturated fatty acids. *Hepatology*. 2015;61(5):1565-1578. doi:10.1002/hep.27695
51. Taminau J, Meganck S, Lazar C, et al. Unlocking the potential of publicly available microarray data using inSilicoDb and inSilicoMerging R/Bioconductor packages. *BMC Bioinformatics*. 2012;13(1):335. https://www.ncbi.nlm.nih.gov/geo/ doi:10.1186/1471-2105-13-335
52. Taminau J, Meganck S, Lazar C, et al. Unlocking the potential of publicly available microarray data using inSilicoDb and inSilicoMerging R/Bioconductor packages. [J]. *BMC Bioinformatics*. 2012;13(1):335. doi:10.1186/1471-2105-13-335
53. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. [J]. *Biostatistics*. 2007;8(1):118-127. doi:10.1093/biostatistics/kxj037
54. Reimand J, Isserlin R, Voisin V, et al. Pathway enrichment analysis and visualization of omics data using g:Profiler, GSEA, Cytoscape and EnrichmentMap. [J]. *Nat Protoc*. 2019;14(2):482-517. doi:10.1038/s41596-018-0103-9
55. Simon N, Friedman J, Hastie T, Tibshirani R. Regularization paths for Cox's proportional hazards model via coordinate descent [J]. *J Stat Softw*. 2011;39(5):1-13. doi:10.18637/jss.v039.i05
56. Ritchie M E, PHIPSON B, WU D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies [J]. *Nucleic acids research*, 2015, 43(7): e47-e.
57. Wu T, Hu E, Xu S, et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. [J]. *Innovation (Camb)*. 2021;2(3):100141. doi:10.1016/j.xinn.2021.100141
58. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. *Bioinformatics*. 2011;27(12):1739-1740. Internet. doi:10.1093/bioinformatics/btr260
59. Szklarczyk D, Kirsch R, Koutrouli M, et al. The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. [J]. *Nucleic Acids Res*. 2023;51(D1):D638-D646. doi:10.1093/nar/gkac1000
60. Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. [J]. *Genome Res*. 2003;13(11):2498-2504. doi:10.1101/gr.1239303
61. Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. [J]. *BMC Bioinformatics*. 2011;12(1):77. doi:10.1186/1471-2105-12-77
62. Sumarac-Dumanovic M, Stevanovic D, Ljubic A, et al. Increased activity of interleukin-23/interleukin-17 proinflammatory axis in obese women. [J]. *Int J Obes (Lond)*. 2009;33(1):151-156. doi:10.1038/ijo.2008.216