ACS OMEGA

http://pubs.acs.org/journal/acsodf

Article

Exploring Key Genes with Diagnostic Value for Nonalcoholic Steatohepatitis Based on Bioinformatics Analysis

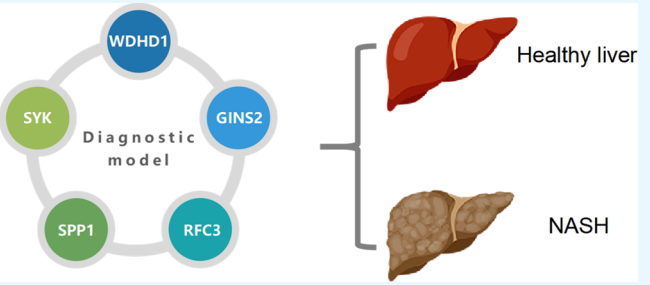
Wenchun Zeng, Xiangwei Xu, Fang Xu, Fang Zhu, Yuecui Li,* and Ji Ma*

Cite This: *ACS Omega* 2023, 8, 20959−20967

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: We aimed to screen specific genes in liver tissue samples of patients with nonalcoholic steatohepatitis (NASH) with clinical diagnostic value based on bioinformatics analysis. The datasets of liver tissue samples from healthy individuals and NASH patients were retrieved for consistency cluster analysis to obtain the NASH sample typing, followed by verification of the diagnostic value of sample genotyping-specific genes. All samples were subjected to logistic regression analysis, followed by the establishment of the risk model, and then, the diagnostic value was determined by receiver operating characteristic curve analysis. NASH samples could be divided into cluster 1, cluster 2, and cluster 3, which could predict the nonalcoholic fatty liver disease activity score of patients. A total of 162 sample genotyping-specific genes were extracted from patient clinical parameters, and the top 20 core genes in the protein interaction network were obtained for logistic regression analysis. Five sample genotyping-specific genes (WD repeat and HMG-box DNA-binding protein 1 [WDHD1], GINS complex subunit 2 [GINS2], replication factor C subunit 3 (RFC3), secreted phosphoprotein 1 [SPP1], and spleen tyrosine kinase [SYK]) were extracted to construct the risk models with high diagnostic value in NASH. Compared with the low-risk group, the high-risk group of the model showed increased lipoproduction and decreased lipolysis and lipid β oxidation. The risk models based on WDHD1, GINS2, RFC3, SPP1, and SYK have high diagnostic value in NASH, and this risk model is closely related to lipid metabolism pathways.



1. INTRODUCTION

Nonalcoholic steatohepatitis (NASH), a manifestation of nonalcoholic fatty liver disease (NAFLD), is diagnosed based on the extent of steatosis, hepatocyte ballooning, lobular inflammation, and degree of fibrosis.[1,2] NASH is related to dysregulated lipid metabolism and hepatic inflammation.[3] Due to the rising prevalence of NAFLD, NASH may become the most common cause of cirrhosis and end-stage liver disease in the coming decades.[4] Moreover, the treatment effect of current drug therapies on NASH is limited, and no drug is approved by the Food and Drug Administration (FDA).[5] Thus, exploring a novel target biomarker for treating NASH is urgently needed.

In recent years, multiple bioinformatic analyses have been adopted to uncover the crucial genes involved in the progression of NASH.[6] WD repeat and HMG-box DNA-binding protein 1 (WDHD1), also known as AND-1, CHTF4, CTF4, and AND1, is the mammalian orthologue of budding yeast (*Saccharomyces cerevisiae*) CTF4, and WDHD1 expression is increased in lung cancer tissues (relative to adjacent tissues), which is negatively correlated with patient prognosis.[7] Besides, highly expressed WDHD1 is also detected in cholangiocarcinoma.[8] GINS complex subunit 2 (GINS2) is identified as a potential hub gene in NASH, which shares a correlation with patients' prognosis.[9] Replication factor C subunit 3 (RFC3) is distinctly upregulated in hepatocellular carcinoma (HCC) tissues and cells, and its downregulation is capable of suppressing HCC progression.[10] A recent study has also indicated that seven hub genes, including secreted phosphoprotein 1 (SPP1), are correlated with fibrosis progression in NASH.[11] SPP1, known as osteopontin (OPN), is implicated in inflammation and liver fibrosis in NAFLD.[12] Another study has also validated that the SPP1 expression is elevated in NASH liver tissues.[4] In the liver, spleen tyrosine kinase (SYK) expression has been detected in both parenchymal (hepatocytes) and non-parenchymal cells (hepatic stellate cells and Kupffer cells) and shares a positive correlation with the disease severity.[13]

Herein, in the present study, we analyzed the Gene Expression Omnibus (GEO) database to identify differentially expressed genes (DEGs). Then, weighted gene coexpression network analysis (WGCNA) and protein−protein interaction (PPI) networks were characterized to explore the impact of

Received: March 14, 2023
Accepted: April 27, 2023
Published: May 30, 2023

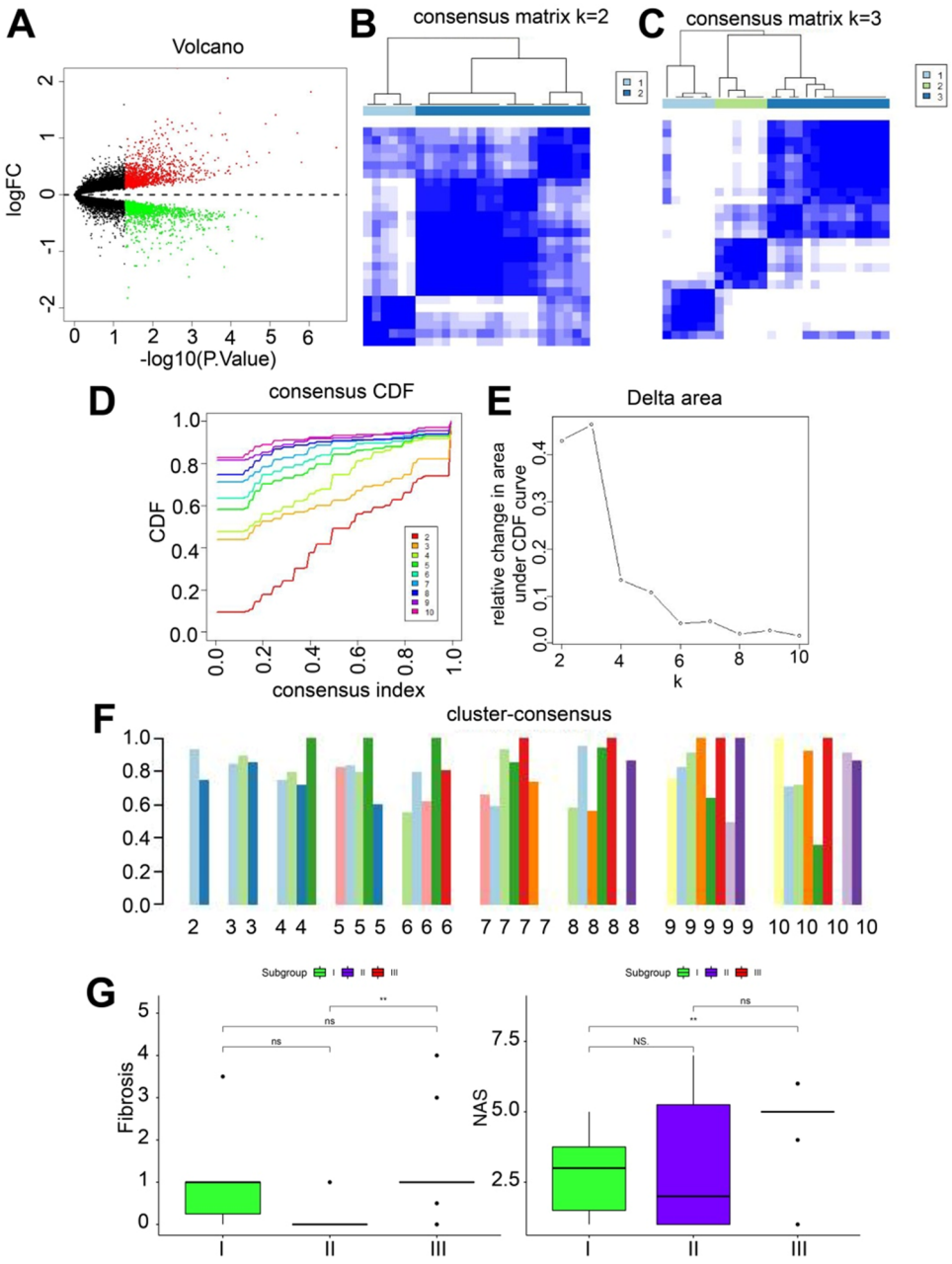


Figure 1. Sample typing of NASH patients based on the GSE48452 dataset. (A) Volcano map of differential analysis of the GSE48452 dataset. The red dots indicate upregulated genes, green dots indicate downregulated genes, and black dots indicate genes without significant expression (control, n = 11; NASH, n = 17). (B) Consistency clustering matrix of k = 2. (C) Consistency clustering matrix of k = 3. (D) Consensus cumulative distribution map. Consensus cluster CDF from k = 2 to k = 10. (E) Delta area plot from k = 2 to k = 10. Relative change of the area under the CDF curve. (F) Cluster−consensus plot from k = 2 to k = 10. (G) Comparison of the fibrosis index and NAS between sample typing; cluster 1, n = 6; cluster 2, n = 6; cluster 3, and n = 14. *p < 0.05; **p < 0.01.

DEGs on NASH. This study aimed at screening potential genes for establishing the risk model with diagnostic value for NASH.

2. MATERIALS AND METHODS

2.1. NASH-Related Transcriptome Sequencing Data. The microarray dataset GSE48452 of liver tissue samples from NASH cases and healthy individuals was downloaded through the Gene Expression Omnibus (GEO) database. After the removal of the samples after bariatric surgery, 28 normal control samples (control, n = 11, healthy obese, n = 17), 9 liver fat deposition samples, and liver tissue samples from 17 NASH patients were retained. As these data were obtained from publicly available databases, ethics committee approval was not required.

2.2. Screening of DEGs. Genes with significantly elevated expressions were selected using the R language "limma" package based on liver tissue sample data from 11 healthy individuals and 17 NASH patients with p < 0.05 as the screening condition.

2.3. Consistency Cluster Analysis. Consistency clustering analysis, the unsupervised clustering method, uses the resampling method to extract a dataset of certain samples, specify the number of clusters k, and calculate the rationality

https://doi.org/10.1021/acsomega.3c01709
ACS Omega 2023, 8, 20959−20967

A

Table: GSEA Results Summary	
Dataset	subgroupI-con
Phenotype	NoPhenotypeAvailable
Upregulated in class	na_ pos
GeneSet	SUBGROUP_I
Enrichment Score (ES)	0.5847791
Normalized Enrichment Score (NES)	1.3725479
Nominal p-value	0.0
FDR q-value	0.0
FWER p-Value	0.0

B

Table: GSEA Results Summary	
Dataset	subgroupII-con
Phenotype	NoPhenotypeAvailable
Upregulated in class	na _pos
GeneSet	SUBGROUP_II
Enrichment Score (ES)	0.6488242
Normalized Enrichment Score (NES)	1. 567163
Nominal p-value	0.0
FDR q-value	0.0
FWER p-Value	0.0

C

Table: GSEA Results Summary	
Dataset	subgroupII-con
Phenotype	NoPhenotypeAvailable
Upregulated in class	na_ pos
GeneSet	SUBGROUP_III
Enrichment Score (ES)	0 6635872
Normalized Enrichment Score (NES)	2.1460953
Nominal p-value	0.0
FDR q-value	0.0
FWER p-Value	0.0

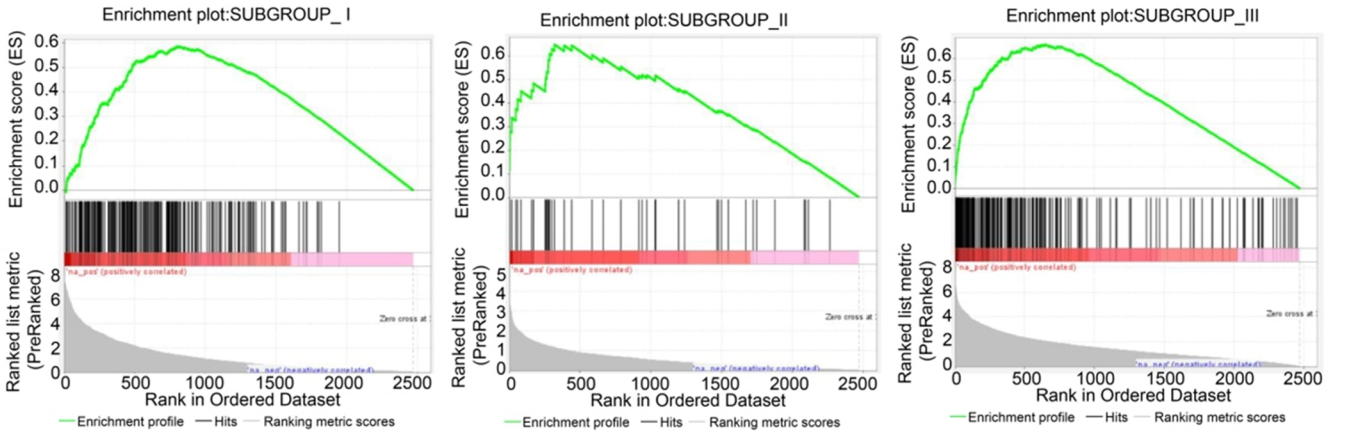


Figure 2. GSEA results of comparison between specific genes for each sample and normal samples. (A) Enrichment plot of comparison between cluster 1 specific genes and normal samples. (B) Enrichment plot of comparison between cluster 2 specific genes and normal samples. (C) Enrichment plot of comparison between cluster 3-specific genes and normal samples. The black vertical line represents the specific genes in the sample typing; the abscissa indicates the difference between the sample typing and the normal samples. The genes closer to the left are the genes with a more significant difference.

under a different number of clusters. In this study, the ConsensusClusterPlus function was performed using R software (v4.0.5). By calculating the consensus cumulative distribution map, the CDF plots were used to help determine the optimal k value, and the delta area plot was drawn. For each of the k, k^{-1} was compared, which was the relative change in the area under the CDF curve. The point with an insignificant increase was selected as the best k value. Moreover, the cluster−consensus plot exhibited the cluster−consensus value for each classification at different k values. A higher value represents higher stability. It can be used to judge the cluster−consensus value at the same k value and between different k values. Furthermore, through the multiway analysis of variance, the relationship between sample typing and clinical parameters was more accurately measured to verify the typing effect.

2.4. Gene Set Enrichment Analysis (GSEA). GSEA is a computational method that assesses whether a priori-defined set of genes shows statistically significant, concordant differences between two biological states. For GSEA, based on genes with significantly different genes compared to normal control samples, the expression matrix was formed to generate an ordered gene list. The genes with significantly different genes compared to other typed samples served as the reference gene set. Each analysis was subjected to 1000 genome permutations, and an enrichment plot was plotted.

2.5. WGCNA. WGCNA, a systematic bioinformatics algorithm,[14] was performed using the R language "WGCNA" package. "pickSoftThreshold" was used to select the suitable soft threshold β. Subsequently, the adjacency matrix was clustered using the topological overlap measure (TOM) to construct the hierarchical clustering tree. The similar gene expression was divided into different modules with minimum gene numbers of 20 in each module. To combine possible similar modules, 0.25 was defined as the threshold of cutting height. At last, each module expression's first principal component was summarized as module eigengenes (MEs). The correlation between ME and traits was calculated, and the most relevant module was selected. In addition, heat maps of the expression of module genes in normal samples and cluster 1, cluster 2, and cluster 3 samples were drawn using the R language "heatmap" package.

2.6. Protein Interaction Analysis. The interaction relationship between 162 sample typing-specific gene-coding proteins was obtained by the STRING website, and the interaction relationship network of genes was visualized using Cytoscape (v3.8.2) software. The depth of color indicates the degree value (the number of connections between proteins and other proteins). A darker color suggests a greater degree value.

2.7. Logistic Regression Analysis. The logistic regression model is the most common multivariate quantification analysis method applied in the regression analysis of dichodependent variables (i.e., $y = 1$ or $y = 0$). In this study, the R language package glmnet was used for logistic regression analysis. The 54 clinical samples were randomly divided into the training set (train group) and the test set (test group), the former for analysis and the latter for validation. The receiver operating characteristic (ROC) curves were drawn using the R language package pROC. The ROC curve and the area under the ROC curve (AUC values) were used to assess the sensitivity and specificity of the risk models or genes in predicting NASH.

2.8. Correlation Analysis. The R language "corrplot" package was used to analyze the correlation between the SPP1 expression and expression of lipid metabolism-related genes and the correlation between values of the risk model and Fibrosis index and nonalcoholic fatty liver disease activity score (NAS).

https://doi.org/10.1021/acsomega.3c01709
ACS Omega 2023, 8, 20959−20967

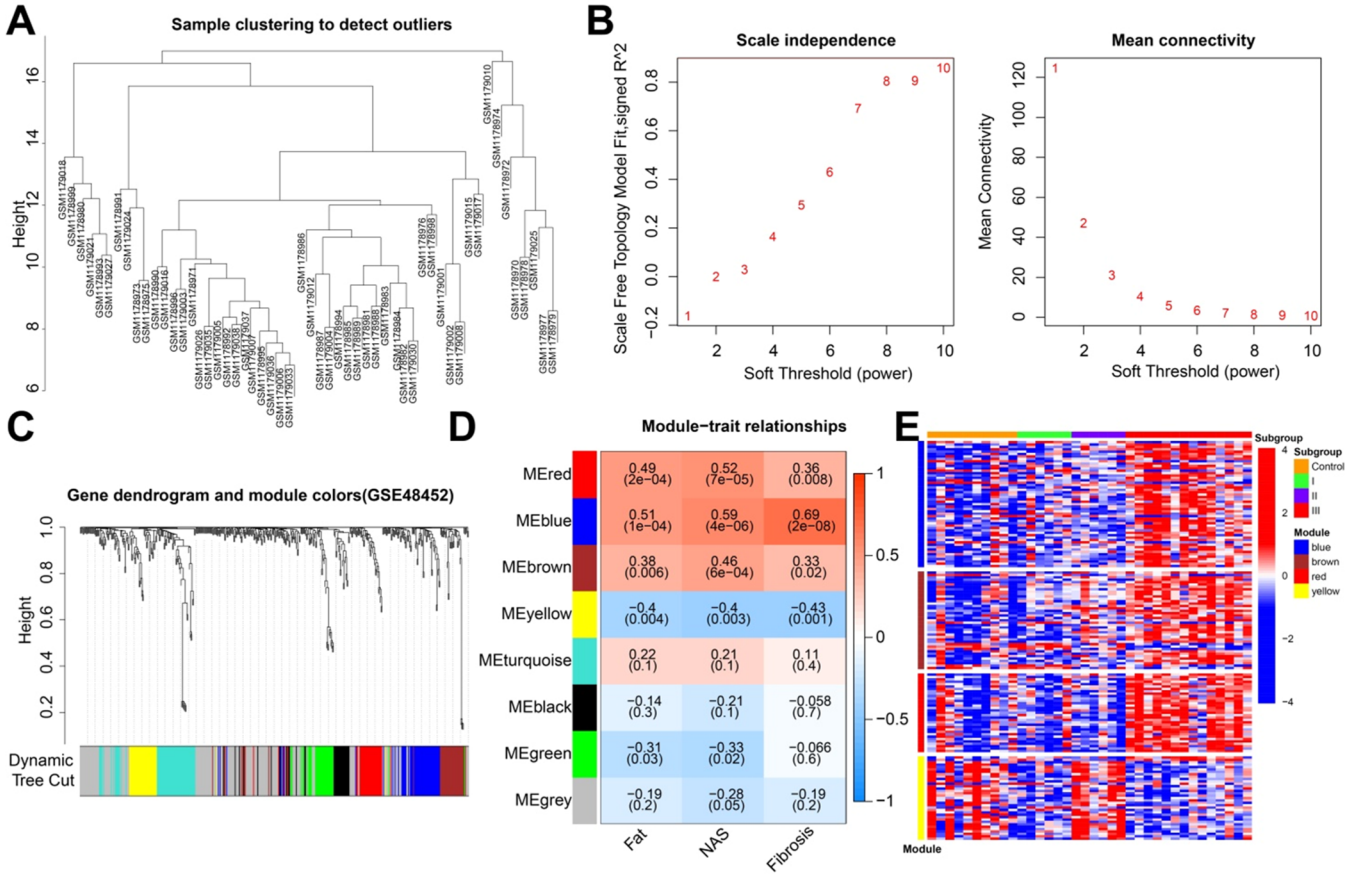


Figure 3. Modules associated with the clinical parameters of NASH based on WGCNA. (A) Cluster dendrogram of liver tissue samples from 26 NASH patients. (B) Scale-free fitting index (left) and average connectivity (right) of various soft threshold power β. The red line indicates the correlation coefficient (0.9). (C) Cluster dendrogram of the coexpressed genes. Each leaf on the cluster dendrogram corresponds to a different gene module. (D) Heat map of correlation between modules and patient clinical parameters (Fat, NAS, and fibrosis index). Each cell contains the corresponding correlation and *P*-values. (E) Heat map of the expression of the red, blue, brown, and yellow module genes in the normal samples, cluster 1, cluster 2, and cluster 3 samples.

3. RESULTS

3.1. Typing of Liver Tissue Samples in NASH Patients. The GSE48452 dataset was retrieved and downloaded from the GEO database with $p < 0.05$ as the threshold for differential analysis, and 2472 DEGs were screened out (Figure 1A). Moreover, liver tissue samples from 9 patients with liver fat deposition and 17 NASH patients were further selected for consistency clustering analysis, which showed that when $k = 3$, the typing results of clusters 1, 2, and 3 had the best effect and when $k \geq 3$, the coefficient dropped sharply, while the classification results of class 4 or over 4 were relatively poor. There was no obvious boundary between the categories (Figure 1B−F). The comparison results of patient clinical parameters (fibrosis and NAS) indicated that the fibrosis index of cluster 3 samples was significantly higher than that of cluster 2 (Figure 1G), and the NAS of cluster 3 samples was significantly higher than that of cluster 1 (Figure 1G). In addition, the multivariate ANOVA results were statistically different ($p < 0.05$), which indicated that the sample typing was a significant effect independent of age, which predicted the NAS of patients (Table S1).

3.2. Specific Genes in NASH are Significantly Different Compared with Normal Samples. Each typing sample was compared with other typing samples with |meaControliffl > 0.2 and adj.*P*.value < 0.05 as the threshold. A total of 210 specific genes in cluster 1, 41 specific genes in cluster 2, and 213 specific genes in cluster 3 were selected (Table S2). Further GSEA was performed; genes that were significantly different from normal control samples served as the input gene dataset and genes that were significantly different compared with other typing samples served as the reference gene dataset. The gene enrichment was compared to test whether these genes were also specific when compared with normal samples, which showed that three samples were significantly different, namely NES > 1, nominal p-value < 0.05, FDR q-value < 0.05, and FWER p-value < 0.05 (Figure 2A−C). These results indicated that the specific genes were also significantly different compared with the normal samples.

3.3. 162 Specific Genes are Closely Related to Clinical Features Based on WGCNA. WGCNA was further performed to screen sample genotyping-specific genes closely related to clinical characteristics in patients with NASH. The samples were first hierarchically clustered, and after removal of two outlier samples was removed (Figure 3A), $\beta = 10$ (no scale $R^2 = 0.9$) was selected as a soft threshold to establish a scale-free network (Figure 3B). Eight coexpression modules were identified (Figure 3C). Significant correlations were found between the red module (33 genes), blue module (53 genes), brown module (41 genes), and yellow module (35 genes) with clinical parameters of patients (fat, NAS, and fibrosis index), among which the red, blue, and brown modules were positively

https://doi.org/10.1021/acsomega.3c01709
ACS Omega 2023, 8, 20959−20967

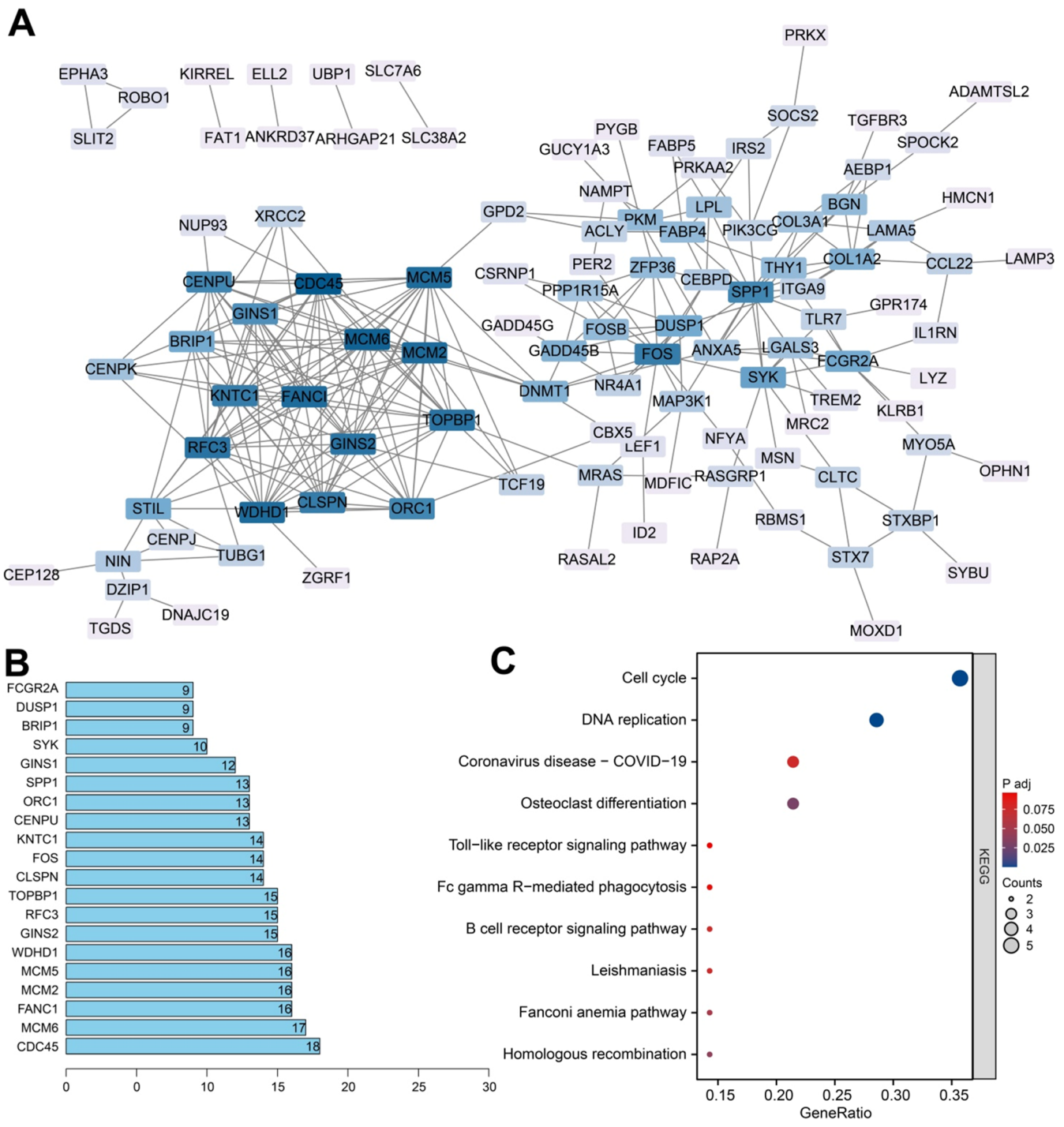


Figure 4. Protein interaction network of 162 sample genotyping-specific genes. (A) Protein interaction network of 162 sample genotyping-specific genes. (B) Ranking diagram of degree values (top 20). (C) Bubble plot of KEGG pathway enrichment results for the top 20 genes; the horizontal coordinate indicates GeneRatio, the vertical coordinate indicates the pathway entry name, the bubble color indicates *P* adj, and the bubble size indicates counts.

correlated with the patient clinical parameters, while the yellow module was negatively correlated with patient clinical parameters. The above-mentioned four modules contained 162 genes. These genes are also specific genes (Figure 3D). As shown in Figure 3E, for red, blue, and brown modules, cluster 3-specific gene expression was significantly higher than that of normal samples, and for yellow modules, cluster 1- and cluster 3-specific gene expression was significantly lower than that of normal samples. These findings indicated that the sample genotyping-specific genes included in the four modules were closely related to the clinical characteristics of NASH patients. These 162 sample genotyping-specific genes were used for subsequent analysis.

3.4. Twenty Specific Genes are the Core Genes in the Protein Interaction Network. We then constructed the protein interaction network through the STRING database to screen out the core genes. A total of 162 specific genes were imported into the STRING database, with human as the

https://doi.org/10.1021/acsomega.3c01709
ACS Omega 2023, 8, 20959−20967

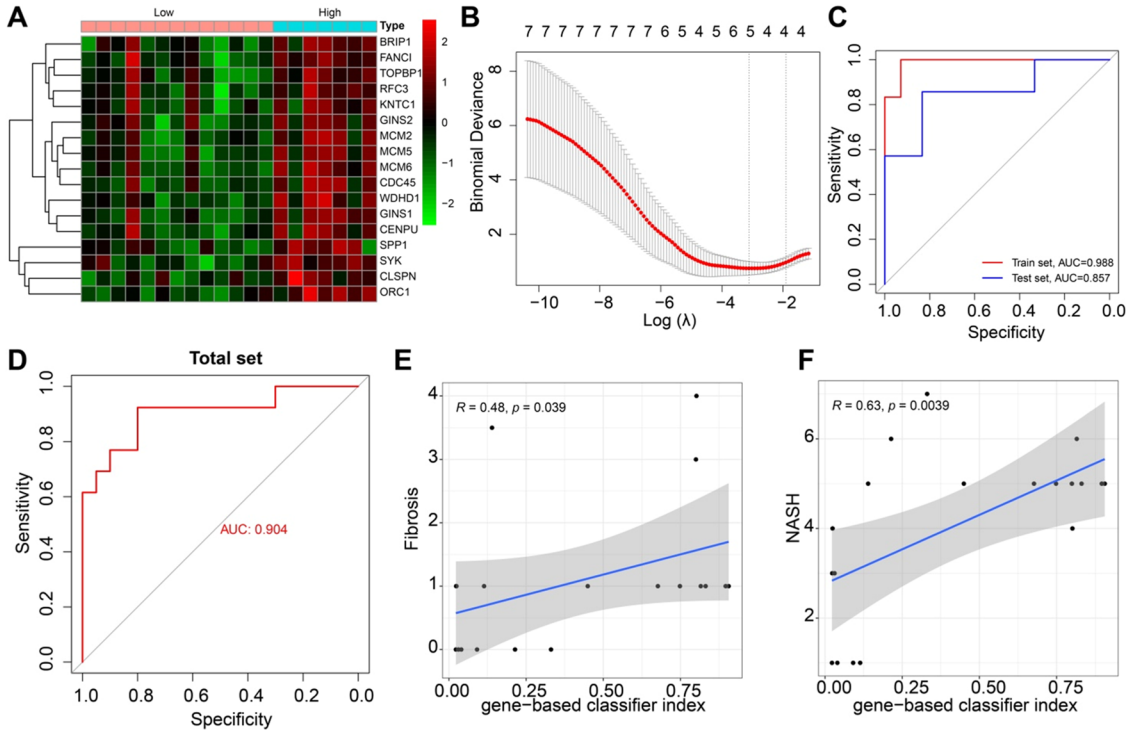


Figure 5. Risk model construction and clinical correlation analysis. (A) Heat map of DEGs in the training group. (B) Abscissa represents the log(λ) value, the ordinate represents the binomial deviance, the upper part is the number of genes retained by the corresponding log(λ) value for the calculation, and the dashed line indicates the log(λ) value and the number of genes. (C) ROC analysis of the constructed risk model and the NASH occurrence in the train and test groups. (D) ROC analysis of the risk model constructed and NASH occurrence in all samples. (E) Correlation between the risk model value and fibrosis index. (F) Correlation between the risk model value and NAS.

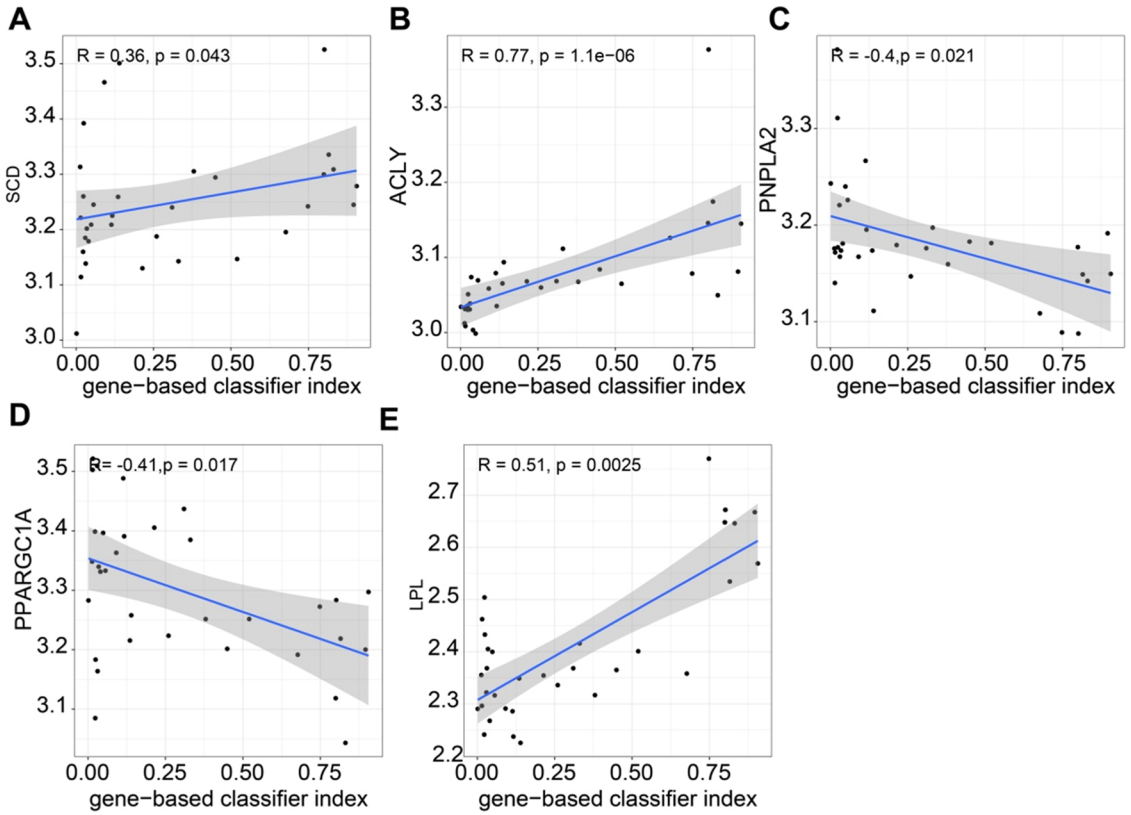


Figure 6. Diagnostic value of five sample typing-specific genes and correlation with lipid metabolism. (A–E) Correlation between the risk model values with SCD (A), ACLY (B), PNPLA2 (C), PGC1α (PPARGC1A) (D), and LPL (E) expression.

https://doi.org/10.1021/acsomega.3c01709
ACS Omega 2023, 8, 20959−20967

species. The generated protein interaction network was further visualized in Cytoscape software. The degree value indicated the number of connections to other proteins in the network, which reflected the core degree of the protein in the network. As shown in Figure 4A, darker blue indicated a larger degree value. We ranked the top 20 genes by degree values as CDC45, MCM6, FANCI, MCM2, MCM5, WDHD1, GINS2, RFC3, TOPBP1, CLSPN, FOS, KNTC1, CENPU, ORC1, SPP1, GINS1, SYK, BRIP1, DUSP1, and FCGR2A (Figure 4B). We believed that the above-mentioned 20 genes were the core genes in the protein interaction network. KEGG pathway enrichment analysis based on core genes showed that core genes were mainly involved in the inflammation-related signaling pathway, including the cell cycle and toll-like receptor signaling pathway and B cell receptor signaling pathway (Figure 4C).

3.5. WDHD1, GINS2, RFC3, SPP1, SYK, and Their Constructed Risk Models have Good Diagnostic Value in NASH. All samples were divided into the train and test groups. There was no difference in the GeControler and age distribution between the two groups, which showed that the test group could be used for validation (Table S3). Based on the samples in the training group, the above-mentioned 20 specific genes were first differentially analyzed, and 8 DEGs were selected (Figure 5A). The logistic regression analysis was performed on these eight genes, and we found that as the number of genes decreased, the error value gradually decreased, and the final number of candidate genes with the smallest error was 5, and the corresponding $\log(\lambda)$ value was −3. The result showed that the risk model of the five genes was successfully constructed (Figure 5B).

ROC analysis showed that the risk model had good diagnostic value in the train (AUC = 0.988) and test (AUC = 0.857) groups (Figure 5C). Combining the two results, the risk model had very high diagnostic value in NASH (AUC = 0.904) (Figure 5D and Table S4). Further analyzing the correlation between the risk model and clinical parameters showed that the value of the risk model was significantly positively correlated with the fibrosis index (R = 0.48, P = 0.039) (Figure 5E) and significantly positively correlated with the NAS (R = 0.63, P = 0.0039) (Figure 5F).

The ROC analysis for five genes was conducted, which showed that all five genes had good diagnostic value, including SPP1 (AUC = 0.817), RFC3 (AUC = 0.801), WDHD1 (AUC = 0.787), SYK (AUC = 0.754), and GINS2 (AUC = 0.752) (Figure S1A). In addition, the five genes were significantly and positively correlated with the NAS, which was ranked as SPP1 (R = 0.502, p < 0.0.001), WDHD1 (R = 0.051, p < 0.001), RFC3 (R = 0.427, p = 0.001), SYK (R = 0.385, p = 0.004), and GINS2 (R = 0.061, p < 0.007) (Figure S1B−F). The above-mentioned results showed that the risk model constructed based on five sample genotyping-specific genes (WDHD1, GINS2, RFC3, SPP1, SYK) had very high diagnostic value in NASH.

3.6. Risk Model Constructed with Five Specific Genes is Closely Related to Lipid Metabolism. Finally, correlation analysis was conducted between the value of risk models constructed with WDHD1, GINS2, RFC3, SPP1, and SYK and the expression of genes related to lipid metabolism. The results showed that the value of the risk model was significantly positively correlated with the expression of lipoproduction-related genes SCD and ACLY (Figure 6A,B), negatively correlated with the expression of lipidolysis-related genes PNPLA2 (Figure 6C), negatively correlated with the expression of lipid β oxidation-related gene PGC1α (PPARGC1A) (Figure 6D), and positively correlated with the expression of LPL, a key enzyme in triglyceride metabolism (Figure 6E). These results showed that compared with the low-risk group, the high-risk group of the model showed increased lipoproduction and decreased lipolysis and lipid β oxidation.

4. DISCUSSION

NASH is characterized by chronic inflammation and accumulation of fat in liver tissue, leading to end-stage liver failure or HCC.[3,15] Therefore, finding novel therapy for hepatic lipid metabolism in NASH remains the focus of the currently relevant research. In our current study, we found 2472 common DEGs by searching the NASH datasets from GEO. Following consistency clustering analysis, we found that when k = 3, the typing results of clusters 1, 2, and 3 had the best effect. We further validated that cluster 3 displayed an obvious clinical correlation with NASH patients. Subsequent analysis confirmed five hub genes (WDHD1, GINS2, RFC3, SPP1, and SYK) showing high diagnostic value in NAFLD. These genes and the risk model constructed with the five genes may have an important ability to predict the risk of NASH.

WDHD1 is a DNA-binding protein in the nucleoplasm and can modulate the cell response to DNA damage and repair and bears great potential in malignant tumors.[7] Higher WDHD1 expression is detected in lung and esophageal carcinogenesis.[16] However, very little research about WDHD1 in the liver can be found, especially in NASH.

GINS2, a member of the GINS family, exerts a crucial role in DNA duplication, and its overexpression causes unfavorable outcomes in diverse tumors, such as non-small-cell lung cancer (NSCLC), breast cancer, HCC, and cervical cancer.[17−20] Importantly, recent studies highlighted GINS2 as the hub gene involved in NASH.[9,21,22] GINS2-overexpressed HCC patients showed poorer overall survival.[23]

It has been documented that decreased RFC3 can limit the multiplication of cancer cells.[24] More importantly, possible roles of RFC3 in liver, breast, esophageal, and ovarian cancers have been highly characterized, and its downregulation is capable of inhibiting cancer cell malignant features.[10,25−27]

In addition, SPP1 had good diagnostic value in NASH, which is closely related to lipid metabolism in NASH. It is universally acknowledged that the fatty acid composition of the diet is a key factor capable of affecting the hepatic lipid metabolism, which affects the pathogenesis of liver diseases.[28] SPP1 is involved in multiple liver diseases by promoting inflammatory reactions.[29,30] SPP1 is upregulated in NASH mice induced by the MCD diet, which is related to the lipid droplet area and inflammation.[4] SPP1 could promote hepatic lipid accumulation and aggravate NASH.[31]

SYK is a cytoplasmic nonreceptor tyrosine kinase, and its overexpression has been detected in balloon hepatocytes containing Mallory−Denk bodies, a hallmark of chronic liver diseases such as metabolism-associated fatty liver diseases and alcoholic liver diseases.[32] Besides, SYK is capable of facilitating liver fibrosis.[33] Specifically, SYK is also confirmed as the hub gene implicated in NASH.[34]

In conclusion, based on the observations and evaluations made during the study, it was suggested that the risk models based on WDHD1, GINS2, RFC3, SPP1, and SYK had high diagnostic value in NASH, and this risk model was closely

https://doi.org/10.1021/acsomega.3c01709
ACS Omega 2023, 8, 20959−20967

related to lipid metabolism pathways. Thus, studying the risk model constructed with the above-mentioned five genes will stimulate a greater interest in prospective studies, leading to the discovery of novel and effective therapeutic targets for NASH treatment.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.3c01709.

Diagnostic value of five specific genes and their correlation with clinical parameters; multivariate ANOVA; details for specific genes in clusters 1, 2, and 3; comparison of clinical parameters between the train group and test group, and risk models based on five genes (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Yuecui Li − *Department of Infectious Liver Disease, The First People's Hospital of Yongkang, Affiliated to Hangzhou Medical College, Jinhua 321300, P. R. China*; Phone: +86-13819909255; Email: yklycwh@126.com

Ji Ma − *Department of Gastroenterology, The First People's Hospital of Yongkang, Affiliated to Hangzhou Medical College, Jinhua 321300, P. R. China*; orcid.org/0000-0001-5049-2179; Phone: +86-13868957815; Email: ma71261698@163.com

Authors

Wenchun Zeng − *Department of Gastroenterology, The First People's Hospital of Yongkang, Affiliated to Hangzhou Medical College, Jinhua 321300, P. R. China*

Xiangwei Xu − *Department of Pharmacy, The First People's Hospital of Yongkang, Affiliated to Hangzhou Medical College, Jinhua 321300, P. R. China*; orcid.org/0000-0003-0035-4303

Fang Xu − *Department of Gastroenterology, The First People's Hospital of Yongkang, Affiliated to Hangzhou Medical College, Jinhua 321300, P. R. China*

Fang Zhu − *Department of Gastroenterology, The First People's Hospital of Yongkang, Affiliated to Hangzhou Medical College, Jinhua 321300, P. R. China*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.3c01709

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was funded by the General Science and Technology Project of Zhejiang Provincial Education Department (Y202146053).

■ REFERENCES

(1) Huby, T.; Gautier, E. L. Immune cell-mediated features of non-alcoholic steatohepatitis. *Nat. Rev. Immunol.* **2022**, *22*, 429−443.

(2) Romero-Gomez, M. Non-alcoholic steatohepatitis. *Med. Clin.* **2022**, *159*, 388−395.

(3) Fraile, J. M.; Palliyil, S.; Barelle, C.; Porter, A. J.; Kovaleva, M. Non-Alcoholic Steatohepatitis (NASH) - A review of a crowded clinical landscape, driven by a complex disease. *Drug Des., Dev. Ther.* **2021**, *Volume 15*, 3997−4009.

(4) Wang, W.; Liu, X.; Wei, P.; Ye, F.; Chen, Y.; Shi, L.; Zhang, X.; Li, J.; Lin, S.; Yang, X. SPP1 and CXCL9 promote non-alcoholic steatohepatitis progression based on bioinformatics analysis and experimental studies. *Front. Med.* **2022**, *9*, No. 862278.

(5) Muthiah, M. D.; Sanyal, A. J. Current management of non-alcoholic steatohepatitis. *Liver Int.* **2020**, *40*, 89−95.

(6) He, W.; Huang, C.; Zhang, X.; Wang, D.; Chen, Y.; Zhao, Y.; Li, X. Identification of transcriptomic signatures and crucial pathways involved in non-alcoholic steatohepatitis. *Endocrine* **2021**, *73*, 52−64.

(7) Gong, L.; Xiao, M.; He, D.; Hu, Y.; Zhu, Y.; Xiang, L.; Bao, Y.; Liu, X.; Zeng, Q.; Liu, J.; Zhou, M.; Zhou, Y.; Cheng, Y.; Zhang, Y.; Deng, L.; Zhu, R.; Lan, H.; Cao, K. WDHD1 leads to cisplatin resistance by promoting MAPRE2 ubiquitination in lung adenocarcinoma. *Front. Oncol.* **2020**, *10*, 461.

(8) Liu, B.; Hu, Y.; Qin, L.; Peng, X. B.; Huang, Y. X. MicroRNA-494-dependent WDHDI inhibition suppresses epithelial-mesenchymal transition, tumor growth and metastasis in cholangiocarcinoma. *Dig. Liver Dis.* **2019**, *51*, 397−411.

(9) Wu, C.; Zhou, Y.; Wang, M.; Dai, G.; Liu, X.; Lai, L.; Tang, S. Bioinformatics analysis explores potential hub genes in nonalcoholic fatty liver disease. *Front. Genet.* **2021**, *12*, No. 772487.

(10) Yao, Z.; Hu, K.; Huang, H.; Xu, S.; Wang, Q.; Zhang, P.; Yang, P.; Liu, B. shRNA-mediated silencing of the RFC3 gene suppresses hepatocellular carcinoma cell proliferation. *Int. J. Mol. Med.* **2015**, *36*, 1393−1399.

(11) Wang, Z.; Zhao, Z.; Xia, Y.; Cai, Z.; Wang, C.; Shen, Y.; Liu, R.; Qin, H.; Jia, J.; Yuan, G. Potential biomarkers in the fibrosis progression of nonalcoholic steatohepatitis (NASH). *J. Endocrinol. Invest.* **2022**, *45*, 1379−1392.

(12) Song, Z.; Chen, W.; Athavale, D.; Ge, X.; Desert, R.; Das, S.; Han, H.; Nieto, N. Osteopontin takes center stage in chronic liver disease. *Hepatology* **2021**, *73*, 1594−1608.

(13) Kurniawan, D. W.; Storm, G.; Prakash, J.; Bansal, R. Role of spleen tyrosine kinase in liver diseases. *World J. Gastroenterol.* **2020**, *26*, 1005−1019.

(14) Langfelder, P.; Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinf.* **2008**, *9*, 559.

(15) Dewidar, B.; Kahl, S.; Pafili, K.; Roden, M. Metabolic liver disease in diabetes-From mechanisms to clinical trials. *Metabolism* **2020**, *111*, No. 154299.

(16) Sato, N.; Koinuma, J.; Fujita, M.; Hosokawa, M.; Ito, T.; Tsuchiya, E.; Kondo, S.; Nakamura, Y.; Daigo, Y. Activation of WD repeat and high-mobility group box DNA binding protein 1 in pulmonary and esophageal carcinogenesis. *Clin. Cancer Res.* **2010**, *16*, 226−239.

(17) Chi, F.; Wang, Z.; Li, Y.; Chang, N. Knockdown of GINS2 inhibits proliferation and promotes apoptosis through the p53/GADD45A pathway in non-small-cell lung cancer. *Biosci. Rep.* **2020**, *40*, No. BSR20193949.

(18) Rantala, J. K.; Edgren, H.; Lehtinen, L.; Wolf, M.; Kleivi, K.; Vollan, H. K.; Aaltola, A. R.; Laasola, P.; Kilpinen, S.; Saviranta, P.; Iljin, K.; Kallioniemi, O. Integrative functional genomics analysis of sustained polyploidy phenotypes in breast cancer cells identifies an oncogenic profile for GINS2. *Neoplasia* **2010**, *12*, 877−888.

(19) Tian, W.; Yang, X.; Yang, H.; Zhou, B. GINS2 functions as a key gene in lung adenocarcinoma by WGCNA co-expression network analysis. *OncoTargets Ther.* **2020**, *13*, 6735−6746.

(20) Ouyang, F.; Liu, J.; Xia, M.; Lin, C.; Wu, X.; Ye, L.; Song, L.; Li, J.; Wang, J.; Guo, P.; He, M. GINS2 is a novel prognostic biomarker and promotes tumor progression in early-stage cervical cancer. *Oncol. Rep.* **2017**, *37*, 2652−2662.

(21) Zhang, D.; Liu, J.; Xie, T.; Jiang, Q.; Ding, L.; Zhu, J.; Ye, Q. Oleate acid-stimulated HMMR expression by CEBPalpha is associated with nonalcoholic steatohepatitis and hepatocellular carcinoma. *Int. J. Biol. Sci.* **2020**, *16*, 2812−2827.

(22) Meng, Q.; Li, X.; Xiong, X. Identification of hub genes associated with non-alcoholic steatohepatitis using integrated bioinformatics analysis. *Front. Genet.* **2022**, *13*, No. 872518.

(23) Li, Z.; Song, G.; Guo, D.; Zhou, Z.; Qiu, C.; Xiao, C.; Wang, X.; Wang, Y. Identification of GINS2 prognostic potential and involvement in immune cell infiltration in hepatocellular carcinoma. *J. Cancer* **2022**, *13*, 610−622.
(24) Li, Y.; Gan, S.; Ren, L.; Yuan, L.; Liu, J.; Wang, W.; Wang, X.; Zhang, Y.; Jiang, J.; Zhang, F.; Qi, X. Multifaceted regulation and functions of replication factor C family in human cancers. *Am. J. Cancer Res.* **2018**, *8*, 1343−1355.
(25) He, Z. Y.; Wu, S. G.; Peng, F.; Zhang, Q.; Luo, Y.; Chen, M.; Bao, Y. Up-Regulation of RFC3 promotes triple negative breast cancer metastasis and is associated with poor prognosis Via EMT. *Transl. Oncol.* **2017**, *10*, 1−9.
(26) Lockwood, W. W.; Thu, K. L.; Lin, L.; Pikor, L. A.; Chari, R.; Lam, W. L.; Beer, D. G. Integrative genomics identified RFC3 as an amplified candidate oncogene in esophageal adenocarcinoma. *Clin. Cancer Res.* **2012**, *18*, 1936−1946.
(27) Shen, H.; Xu, J.; Zhao, S.; Shi, H.; Yao, S.; Jiang, N. ShRNA-mediated silencing of the RFC3 gene suppress ovarian tumor cells proliferation. *Int. J. Clin. Exp. Pathol.* **2015**, *8*, 8968−8975.
(28) Ferramosca, A.; Zara, V. Modulation of hepatic steatosis by dietary fatty acids. *World J. Gastroenterol.* **2014**, *20*, 1746−1755.
(29) Wang, G.; Chen, S.; Zhao, C.; Li, X.; Zhao, W.; Yang, J.; Chang, C.; Xu, C. Comparative analysis of gene expression profiles of OPN signaling pathway in four kinds of liver diseases. *J. Genet.* **2016**, *95*, 741−750.
(30) Wang, C.; He, M.; Peng, J.; Li, S.; Long, M.; Chen, W.; Liu, D.; Yang, G.; Zhang, L. Increased plasma osteopontin levels are associated with nonalcoholic fatty liver disease in patients with type 2 diabetes mellitus. *Cytokine* **2020**, *125*, No. 154837.
(31) Huang, B.; Wen, W.; Ye, S. TSH-SPP1/TRbeta-TSH positive feedback loop mediates fat deposition of hepatocyte: Crosstalk between thyroid and liver. *Front. Immunol.* **2022**, *13*, No. 1009912.
(32) Luci, C.; Vieira, E.; Bourinet, M.; Rousseau, D.; Bonnafous, S.; Patouraux, S.; Lefevre, L.; Larbret, F.; Prod'homme, V.; Iannelli, A.; Tran, A.; Anty, R.; Bailly-Maitre, B.; Deckert, M.; Gual, P. SYK-3BP2 pathway activity in parenchymal and myeloid cells is a key pathogenic factor in metabolic steatohepatitis. *Cell. Mol. Gastroenterol. Hepatol.* **2022**, *13*, 173−191.
(33) Qu, C.; Zheng, D.; Li, S.; Liu, Y.; Lidofsky, A.; Holmes, J. A.; Chen, J.; He, L.; Wei, L.; Liao, Y.; Yuan, H.; Jin, Q.; Lin, Z.; Hu, Q.; Jiang, Y.; Tu, M.; Chen, X.; Li, W.; Lin, W.; Fuchs, B. C.; Chung, R. T.; Hong, J. Tyrosine kinase SYK is a potential therapeutic target for liver fibrosis. *Hepatology* **2018**, *68*, 1125−1139.
(34) Kurniawan, D. W.; Jajoriya, A. K.; Dhawan, G.; Mishra, D.; Argemi, J.; Bataller, R.; Storm, G.; Mishra, D. P.; Prakash, J.; Bansal, R. Therapeutic inhibition of spleen tyrosine kinase in inflammatory macrophages using PLGA nanoparticles for the treatment of non-alcoholic steatohepatitis. *J. Controlled Release* **2018**, *288*, 227−238.

https://doi.org/10.1021/acsomega.3c01709
ACS Omega 2023, 8, 20959−20967