



OPEN ACCESS

EDITED BY

Habiba Alsafar,
Khalifa University, United Arab Emirates

REVIEWED BY

Geoffrey William Mccaughan,
The University of Sydney, Australia
Manuela Neuman,
University of Toronto, Canada

*CORRESPONDENCE

Xiaoguang Zou,
✉ zxgks@163.com
Fangping He,
✉ hefp5@mail.sysu.edu.cn
Wenjie Chen,
✉ chenwj5@mail.sysu.edu.cn

†These authors have contributed equally to this work and share first authorship

SPECIALTY SECTION

This article was submitted to Genetics of Common and Rare Diseases, a section of the journal Frontiers in Genetics

RECEIVED 10 October 2022

ACCEPTED 21 February 2023

PUBLISHED 06 March 2023

CITATION

Yang F, Ni B, Lian Q, Qiu X, He Y, Zhang Q, Zou X, He F and Chen W (2023), Key genes associated with non-alcoholic fatty liver disease and hepatocellular carcinoma with metabolic risk factors. *Front. Genet.* 14:1066410. doi: 10.3389/fgene.2023.1066410

COPYRIGHT

© 2023 Yang, Ni, Lian, Qiu, He, Zhang, Zou, He and Chen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Key genes associated with non-alcoholic fatty liver disease and hepatocellular carcinoma with metabolic risk factors

Fan Yang^{1,2,3†}, Beibei Ni^{4†}, Qinghai Lian^{4†}, Xiusheng Qiu⁴, Yizhan He², Qi Zhang^{1,2}, Xiaoguang Zou^{1*}, Fangping He^{5*} and Wenjie Chen^{2,4*}

¹Department of Infectious Diseases, The First People's Hospital of Kashi, The Kashi Affiliated Hospital, Sun Yat-Sen University, Kashi, China, ²Biotherapy Centre, The Third Affiliated Hospital, Sun Yat-Sen University, Guangzhou, China, ³Postdoctoral Research Station, Xinjiang Medical University, Ürümqi, China, ⁴Cell-Gene Therapy Translational Medicine Research Centre, The Third Affiliated Hospital, Sun Yat-Sen University, Guangzhou, China, ⁵Department of Hepatobiliary and Pancreatic Surgery, The Eighth Affiliated Hospital, Sun Yat-Sen University, Shenzhen, Guangdong, China

Background: Hepatocellular carcinoma (HCC) has become the world's primary cause of cancer death. Obesity, hyperglycemia, and dyslipidemia are all illnesses that are part of the metabolic syndrome. In recent years, this risk factor has become increasingly recognized as a contributing factor to HCC. Around the world, non-alcoholic fatty liver disease (NAFLD) is on the rise, especially in western countries. In the past, the exact pathogenesis of NAFLD that progressed to metabolic risk factors (MFRs)-associated HCC has not been fully understood.

Methods: Two groups of the GEO dataset (including normal/NAFLD and HCC with MFRs) were used to analyze differential expression. Differentially expressed genes of HCC were verified by overlapping in TCGA. In addition, functional enrichment analysis, modular analysis, Receiver Operating Characteristic (ROC) analysis, LASSO analysis, and Genes with key survival characteristics were analyzed.

Results: We identified six hub genes (FABP5, SCD, CCL20, AGPAT9(GPAT3), PLIN1, and IL1RN) that may be closely related to NAFLD and HCC with MFRs. We constructed survival and prognosis gene markers based on FABP5, CCL20, AGPAT9(GPAT3), PLIN1, and IL1RN. This gene signature has shown good diagnostic accuracy in both NAFLD and HCC and in predicting HCC overall survival rates.

Conclusion: As a result of the findings of this study, there is some guiding significance for the diagnosis and treatment of liver disease associated with NAFLD progression.

KEYWORDS

hepatocellular carcinoma (HCC), metabolic risk factors (MRFs), non-alcoholic fatty liver disease (NAFLD), differentially expressed genes (DEGs), diagnosis, prognosis

1 Introduction

Hepatocellular carcinoma (HCC) accounts for about 90% of all liver cancers, making it the second leading cause of cancer death worldwide (Llovet et al., 2021, M. A; Morse et al., 2019). The development of HCC usually follows a background of chronic low-grade inflammation characterised by chronic liver damage followed by inflammation,

hepatocellular necrosis, and regeneration. HCCs are predominantly caused by Hepatitis B (HBV), Hepatitis C (HCV) infection, alcohol consumption, as well as metabolic perturbations leading to non-alcoholic fatty liver disease (NAFLD) (R. E. Ericksen et al., 2019, C; Trierweiler et al., 2016). Due to universal vaccination and antiviral therapy, viral HCC prevalence is decreasing. It will be necessary to modify strategies for cancer prevention, prediction, and surveillance for HCC (Awosika and Sohal, 2022, S. F; Huang et al., 2018). The metabolic syndrome refers to a group of disorders that include dyslipidemia, hyperglycemia, and obesity, has received increasing attention as a novel risk factor for HCC (K. Akahoshi et al., 2016). There was a greater than 100-fold increase in risk for HBV or HCV carriers who also had diabetes or obesity, which suggests synergistic effects of metabolic factors and hepatitis (C. L. Chen et al., 2008). Patients with metabolic risk factors (MRFs) may be at greater risk of developing hepatocarcinogenesis when FABP4 is overexpressed in HSCs (N. Chiyonobu et al., 2018).

According to statistics, 25 per cent of the population worldwide suffers from NAFLD (A. Lonardo et al., 2016). NAFLD is becoming more prevalent worldwide, especially in western countries (Mancina et al., 2016). Consequently, NAFLD has become an economic and health concern worldwide. There is no doubt that NAFLD is a hepatic manifestation of metabolic syndrome (MS) and is often associated with dyslipidemia, obesity, and T2DM (J. Wattacheril, 2020). As a result of liver lipid accumulation, NAFLD can cause inflammation and damage to the hepatocytes (J. Wattacheril, 2020). The liver biopsy usually shows milder forms (steatosis) to severe conditions (non-alcoholic steatohepatitis (NASH), advanced fibrosis, cirrhosis) (Pouwels et al., 2022).

In the past, the exact pathogenesis of NAFLD that progressed to MFRs-associated HCC has not been fully understood. High-throughput gene chips and transcriptome sequencing have entirely changed the previous systematic analysis methods for disease research (M. Bustoros et al., 2020). RNA sequencing and high-throughput microarrays help to identify reliable biological markers, classify diseases, and reveal mechanisms of disease development. The discovery of new biomarkers can be helpful in predicting risk and determining which treatment is most suitable for an individual patient. Thus, the prediction of candidate genes may also be based on NAFLD-HCC with MRFs pathogenesis.

This study aims to identify the key genes involved in NAFLD and HCC with MRFs and to provide a reference for further study of the transformation of MFRs-associated HCC and a molecular-targeted approach to cancer treatment. In this study, we analyzed microarray data comprehensively, selecting normal tissues and NAFLD samples and microarray data of MFRs-associated HCC and adjacent normal tissues, and separately analysed the differentially expressed genes (DEGs) in both groups of chips. Combining the GEO DEG data of human HCC with MFRs and normal liver tissue with chip data to determine key DEGs that directly affect the diagnosis and treatment of NAFLD. Afterwards, further functional enrichment analysis was conducted to determine how DEGs regulate the main biological functions. Furthermore, by using protein-protein interaction (PPI) networks and survival analysis of patient data, key genes are identified that affect the diagnosis, treatment, and prognosis of patients with NAFLD.

2 Methods

2.1 Profiles of gene expression

GSE63067, GSE89632, and GSE102079 datasets were downloaded from Gene Expression Omnibus (GEO), an open-access database that provides gene expression profiles. GSE63067 (Frades et al., 2015) and GSE102079 (N. Chiyonobu, S. Shimada, Y. Akiyama, K. Mogushi, M. Itoh, K. Akahoshi, S. Matsumura, K. Ogawa, H. Ono, Y. Mitsunori, D. Ban, A. Kudo, S. Arii, T. Suganami, S. Yamaoka, Y. Ogawa, M. Tanabe and S. Tanaka, 2018) are both based on the GPL570 [(HG-U133_Plus_2) Affymetrix Human Genome U133 Plus 2.0 Array]. GSE89632 (B. M. Arendt et al., 2015) is based on [(GPL14951) Illumina HumanHT-12 WG-DASL V4.0 R2 expression bead chip]. The title of the GSE63067 data set is “Expression data from human non-alcoholic fatty liver disease stages”. The data contained the gene expression profiles of 11 NAFLD patients and seven non-NAFLD controls. The title of the GSE102079 data set is “FABP4 overexpressed in intratumoral hepatic stellate cells within hepatocellular carcinoma with metabolic risk factors”. Between 2006 and 2011, 152 patients who underwent curative hepatic resection for HCC at Tokyo Medical and Dental University Hospital participated in an integrated gene expression microarray study. In the control group, 14 adjacent liver tissues were obtained from patients with metastases of colorectal cancer without chemotherapy. The validation data set was from GSE89632 and The Cancer Genome Atlas (TCGA) data set. The title of the GSE89632 data set is “Genome-wide analysis of hepatic gene expression in patients with non-alcoholic fatty liver disease and healthy donors with hepatic fatty acid composition and other nutritional factors”. A cross-sectional study included 20 patients with simple steatosis (SS), 19 non-alcoholic steatohepatitis (NASH), and 24 healthy liver donors. The TCGA database of liver hepatocellular carcinoma (LIHC) contains RNA-Seq data for 374 HCC patients and 50 normal tissues (<https://portal.gdc.cancer.gov/>) for gene expression and immune system infiltrates.

2.2 Analysis of differentially expressed genes (DEGs) in NAFLD and HCC with MRFs

A comparison of DEGs between NAFLD and normal controls, HCC patients with MRFs, and corresponding controls was performed using the limma R package “complexheatmap” and “ggplot2” to generate heat maps and volcano maps, respectively, which is an efficient analysis method in bioinformatics (M. E. Ritchie et al., 2015). In NAFLD datasets, the selected criteria were p -value <0.05 and $|\log_2FC| > 1$. In HCC datasets, the selected criteria were p -value <0.05 and $|\log_2FC| > 1$. Additionally, the overlapping DEGs between NAFLD and HCC with MRFs were determined by Venn diagrams using the Venn platform (<http://bioinformatics.psb.ugent.be/webtools/Venn/>). A subsequent analysis was performed on these overlapping DEGs.

2.3 Functional classification and pathway enrichment for DEGs

GO function enrichment analyses were conducted on the above overlapping DEGs. It consisted of biological process (BP), cellular component (CC), and molecular function (MF) (2006). The analysis of KEGG signaling pathway enrichment using a package called “clusterProfiler” (M. Kanehisa et al., 2016). GO terms and KEGG pathways enriched with adjusted *p*-value of 0.05 were selected for analysis.

2.4 Establishment of protein-protein interactions and identification of hub genes

In order to further investigate the interactions between the above-mentioned common genes, a search tool called the Search Tool for Retrieval of Interacting Genes (STRING) has been developed for PPI network construction (D. Szklarczyk et al., 2015). Interaction scores of at least 0.4 were considered significant. Subsequently, PPI network visualisation was conducted with Cytoscape software. Then, the Maximal Clique Centrality (MCC), Density of Maximum Neighborhood Component (DMNC), Maximum Neighborhood Component (MNC), Degree, and Edge Percolated Component (EPC), algorithms in the cytoHubba plug-in (<http://hub.iis.sinica.edu.tw/cytohubba/>) was applied to identify PPI hub genes with high connectivity.

2.5 Comparing the hub gene expression degree and analysing the prognosis

Based on the TCGA database, the six hub genes expression in HCC normal tissues and tumor tissues was investigated. In LIHC, 374 HCC specimens with normal adjacent tissues and HCC tissue (50 each) were compared with neighbouring normal tissues. GEPIA was used to investigate the prognostic significance of hub genes (<http://gepia.cancer-pku.cn/index.html>) (Z. Tang et al., 2017). Survival analyses were considered significant when log-rank *p* < 0.05 was used.

2.6 Developing signatures and evaluating their reliability

Based on the training dataset, hub genes associated with prognosis were identified and assessed against other datasets for their predictive performance. Half of TCGA is set as the training set. The other half of TCGA is set as a validation set. The entire TCGA cohort is a verification set. Using univariate Cox proportional hazard regression analysis, it was evaluated whether hub genes are associated with overall survival (OS) in the training set. In the “glmnet” package, the Latent Selection Operator penalised Cox proportional hazard regression using Cox proportional hazards models. A prediction formula for gene characteristics was devised. The formula for the model is as follows: risk score = gene1 $\times\beta$ 1 (gene one expression level) + gene 2 $\times\beta$ 2 (gene two expression level) + . . . gene n $\times\beta$ n (gene n expression level). In this formula, genes are combined with gene expression values and regression coefficients from multiple Cox proportional hazards regression models (George et al., 2014;

Zhang J. et al., 2021). Using the Kaplan–Meier (K–M) survival curves, survival comparisons were performed between low- and high-risk groups *via* the R package “survival”. Furthermore, a time-dependent receiver operating characteristic (ROC) analysis (including 1-, 3-, and 5-year survival) was conducted to evaluate hub gene sensitivity and specificity using the R package “survival ROC” (P. J. Heagerty et al., 2000). It is critical to consider the area of the AUC curve when trying to predict clinical outcomes. Prognosis is better when AUC >0.5; the closer AUC is to 1, the better.

2.7 The expression of hub genes is correlated with the presence of immune cells in tumor

Tumor contains a large number of immune cells, and the prognosis of high-grade HCC patients with high subtype of dominant immunity is obviously better (Y. Kurebayashi et al., 2018). To examine whether the expression of hub genes is correlated with the presence of immune cells in HCC, we examined the correlation between hub gene mRNA expression and tumor-infiltrating immune cells. The web tool TIMER was used (<https://cistrome.shinyapps.io/timer/>) (T. Li et al., 2017). Six tumor-infiltrating cell subsets were analysed, such as B cells, CD8⁺ T cells, CD4⁺ T cells, macrophages, neutrophils, and dendritic cells.

2.8 Statistical analysis

ROC curves for hub genes were constructed using the pROC package (X. Robin et al., 2011). To measure the effectiveness of the model, we calculated the area under the curves (AUC). These results showed the usefulness of genes for diagnostic purposes. *p*-values <0.05 were considered statistically significant.

3 Results

3.1 Identification of DEGs in NAFLD and HCC

The series GSE63067 dataset about NAFLD and the series GSE102079 dataset about HCC from the NCBI GEO database was downloaded. Based on a *p*-value of 0.05 and |log₂FC| of >1.0, 125 DEGs were identified in GSE63067, and 726 DEGs were identified in GSE102079 using the “limma” package in R software. Volcano plots and heatmaps were used to visualise the DEGs of the two data sets shown in Figures 1A,B and Figures 2A,B, respectively. Using the Venn Diagram online tool, 26 common genes of two diseases were identified and are shown in Figure 1C.

3.2 Analysis of pathways and functional roles associated with overlapping DEGs

Functional enrichment and KEGG pathway analyses of 26 common NAFLD and HCC genes were performed at a

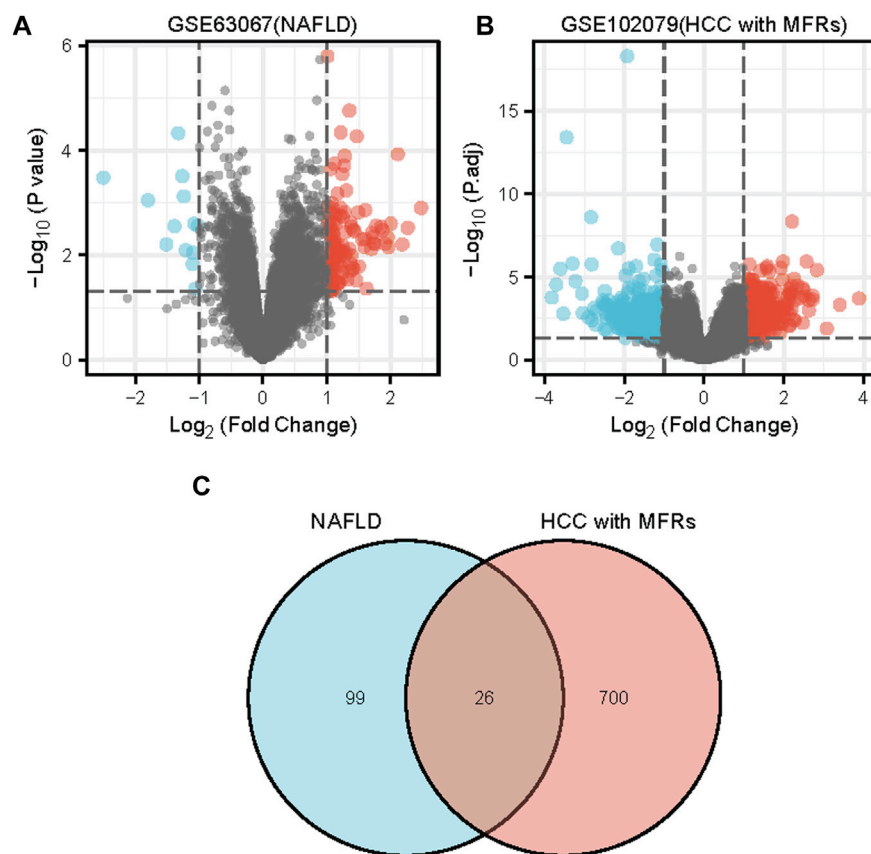


FIGURE 1

Differentially expressed genes (DEGs) shown in a volcano plot and Venn diagram (A) An analysis of the differential genes in GSE63067 using a volcano map. (B) An analysis of the differential genes in GSE102079 using a volcano map. (C) Venn diagram of DEGs in GSE63067 and GSE102079 data sets. Abbreviations: DEGs, differentially expressed genes; NAFLD, non-alcoholic fatty liver; HCC, Hepatocellular Carcinoma, MFRs, metabolic risk factors.

threshold of p -value <0.05 . The results showed that DEGs were enriched in biological processes, including cellular response to environmental stimuli, cellular response to abiotic stimuli, cellular response to ionising radiation, unsaturated fatty acid biosynthetic process, and response to zinc ions (Figure 3A). Regarding molecular function, DEGs were principally associated with receptor ligand activity, cytokine activity, monocarboxylic acid binding, and fatty acid binding (Figure 3B). The KEGG pathways of DEGs were enriched in the PPAR signalling pathway (Figure 3C).

3.3 Analysing the PPI network and selecting hub genes

The PPI network was first performed based on the STRING database to investigate how DEGs interact with one another. Afterwards, the results were imported into Cytoscape to be analysed (Figure 4A). A Cytoscape plug-in, Cytohubba, was used to analyse the PPI network and identify hub genes. We got top 10 genes from protein-protein network ranked by five different algorithms of cytohubba including MCC, DMNC, MNC, Degree, and EPC (Table 1). In this study, the genes with

the top six values were considered as hub genes. Based on the five algorithms, the top six genes were determined to be hub genes: FABP5, SCD, CCL20, AGPAT9, PLIN1, IL1RN (Figure 4B).

3.4 The diagnostic value of hub genes has been validated

To evaluate the diagnostic value of the top six hub genes obtained from the above analysis, ROC curves were constructed and their corresponding area under the curve (AUC) was calculated. Figure 5A shows the result of NAFLD. The AUC for FABP5, SCD, CCL20, AGPAT9, PLIN1, and IL1RN in NAFLD patients and normal controls were 0.828, 0.818, 0.883, 0.857, 0.961, 0.818 at NAFLD GSE63067 dataset. Figure 5B shows the ROC curves in HCC patients and normal controls. The AUC for FABP5, SCD, CCL20, AGPAT9, PLIN1, and IL1RN in HCC and the normal controls were 0.717, 0.785, 0.776, 0.821, 0.901, 0.839 at HCC with MRFs GSE102079 dataset. For validation, The AUC for FABP5, SCD, CCL20, AGPAT9, and IL1RN were 0.561, 0.746, 0.858, 0.981, 0.907 in NAFLD based on GSE89632 (Figure 5C). The AUC for FABP5, SCD, CCL20, AGPAT9, PLIN1, and

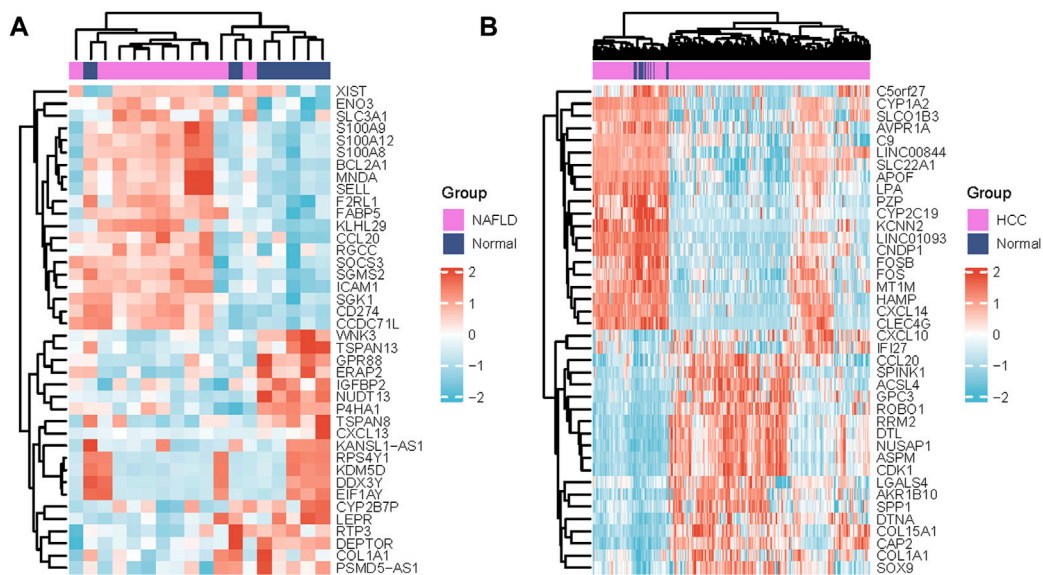


FIGURE 2 Differentially expressed genes heatmaps (A) Heat map of DEGs in GSE63067 (NAFLD) and (B) Heatmap of DEGs in GSE102079(HCC with MFRs). DEGs in red indicate upregulation, DEGs in blue indicate downregulation, and DEGs in white indicate no significant changes. Abbreviations: NAFLD, non-alcoholic fatty liver; HCC, Hepatocellular Carcinoma.

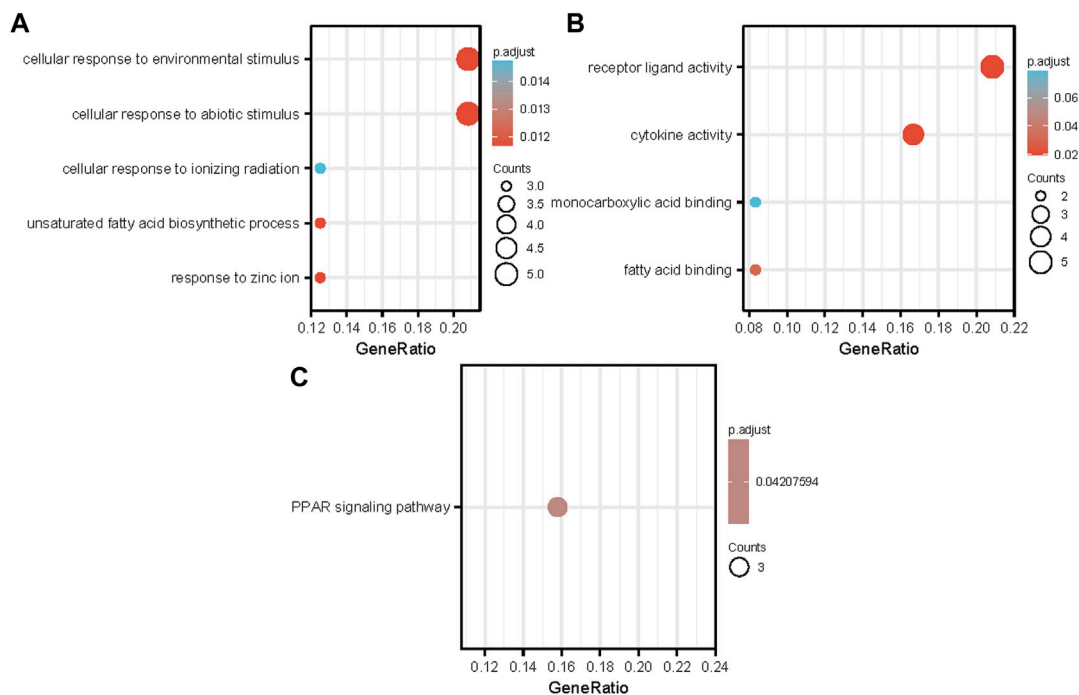


FIGURE 3 Analyses of functional enrichment between two groups of DEGs. (A) Enrichment results for GO biological processes; (B) Enrichment results for GO molecular function processes; (C) Enrichment results for KEGG pathways. A bubble's size represents the number of genes associated with each term. A term's bubble size represents how many genes are associated with it. Each bubble's color indicates the adjusted *p*-value. Abbreviations: GO, Gene Ontology; BP, biological process; MF, molecular function; KEGG, Kyoto Encyclopedia of Genes and Genomes.

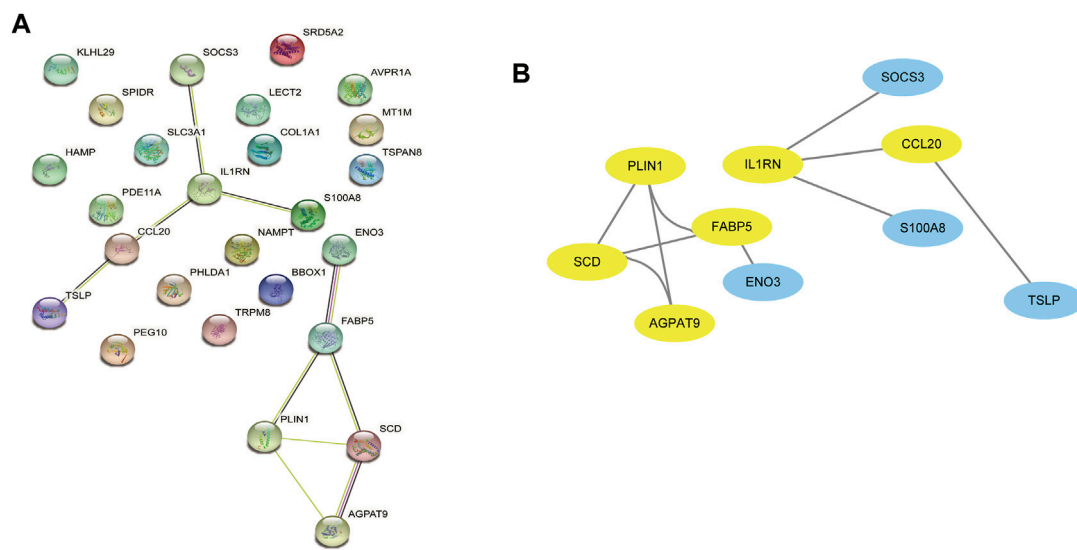


FIGURE 4
The PPI network analysis highlights the most significant modules related to DEGs. (A) 26 DEGs were used in the construction of this PPI network. (B) The most significant module of the PPI network includes 6 hub genes (yellow circles). DEGs differentially expressed genes; PPI, Protein-Protein interaction.

TABLE 1 Results for analysis by Cytohubba.

node_name	MCC	DMNC	MNC	Degree	EPC
S100A8	1	0	1	1	1.947
SOCS3	1	0	1	1	1.998
FABP5	3	0.30779	2	3	2.985
ENO3	1	0	1	1	2.051
TSLP	1	0	1	1	1.853
IL1RN	3	0	1	3	2.56
CCL20	2	0	1	2	2.318
SCD	4	0.30898	3	3	3.027
PLIN1	4	0.30898	3	3	3.055
AGPAT9	2	0.30779	2	2	2.75

MCC, maximal clique centrality; DMNC, density of maximum neighborhood component; MNC, maximum neighborhood component; EPC, edge percolated component

IL1RN in HCC and the normal controls were 0.900, 0.614, 0.807, 0.589, 0.725, 0.769 at HCC TCGA dataset (Figure 5D).

3.5 An evaluation of the expression patterns and survival analysis of six hub genes

Consistent with the results in GEO datasets, the mRNA expression of AGPAT9, PLIN1, and IL1RN was significantly downregulated. At the same time, that of FABP5, SCD, and CCL20 were upregulated considerably in TCGA HCC compared with non-tumor tissues (Figures 6A–L). According to the GEPIA web tool, FABP5 and PLIN1 mRNA expression are significantly linked to overall survival (OS) (Figures 6M–Q).

3.6 An analysis of the correlation between hub gene expression levels

The correlation of expression levels of hub genes was captured using GEPIA. An analysis of correlation was performed on any two of FABP5, SCD, CCL20, AGPAT9(GPAT3), PLIN1, and IL1RN and six hub genes. The above data indicate that upregulation of one of them will decrease the high expression of other genes. IL1RN and GPAT3 (Figure 7A), CCL20 and SCD (Figure 7B), CCL20 and FABP5 (Figure 7C), SCD and PLIN1 (Figure 7D), IL1RN and CCL20 (Figure 7E), and IL1RN and PLIN1 (Figure 7F) are all positively related to each other. This may indicate that there is a common transcription factor as well as epigenetic modifications controlling them all.

3.7 The construction of the hub gene prognostic signature

To avoid overfitting by LASSO regression, lambda. Min was selected, resulting in a more accurate prediction rate. We used the multivariate Cox proportional hazards regression analysis. Five prognostic genes were developed, including ABP5, CCL20, GPAT3, PLIN1, and IL1RN in order to calculate the risk score for each patient, the following formula was used: risk score = (0.170124209 × FABP5) + (0.073621309 × CCL20) + (0.011005683 × GPAT3) + (−0.056212587 × PLIN1) + (−0.100588077 × IL1RN). The LASSO coefficient for SCD is equal to 0. Hub gene risk scores were used to determine whether HCC patients were low-risk or high-risk (Figure 8A). A significantly worse OS was observed in high-risk patients compared to low-risk patients (Figure 8B, training set $p = 0.008$, validation set $p = 0.0026$, entire TCGA set $p = 0.001$). The reliability of hub genes was subsequently

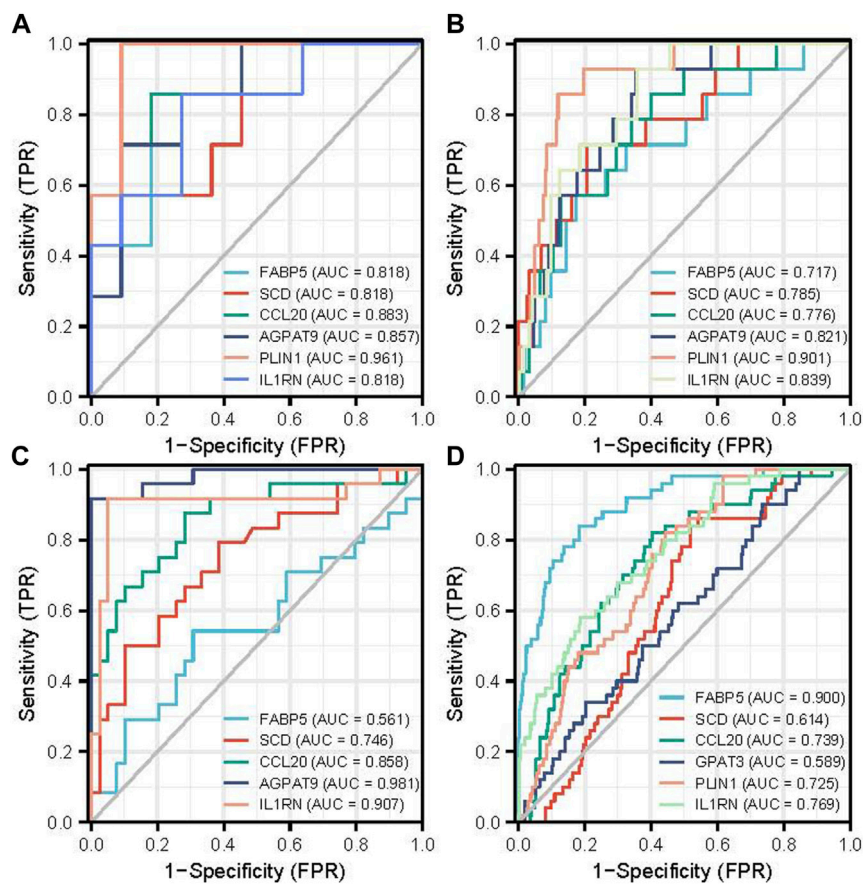


FIGURE 5

The diagnostic value of the top six hub genes with ROC curves in NAFLD and HCC. (A) The diagnostic value of the top six hub genes on ROC curves in NAFLD is based on the GSE63067 data set. (B) The diagnostic value of the top six hub genes on ROC curves in HCC with metabolic risk factors is based on the GSE102079 data set. (C) The diagnostic value of the top six hub genes on ROC curves in NAFLD is based on GSE89632 for validation. (D) The diagnostic value of the top six hub genes on ROC curves in HCC is based on TCGA data. Abbreviations: TPR, True Positive Rate; FPR, False Positive Rate.

assessed using time-dependent ROC curves (Figure 8C). As a result, the area under the curve (AUC) was 0.738, 0.612, and 0.695 for 1-year, 3-year, and 5-year survival, respectively for the training set. The AUC was 0.611, 0.633, and 0.664 for 1-year, 3-year, and 5-year survival, respectively, for the validation set. These curves were also applied in the entire TCGA set. The AUC was 0.696, 0.634, and 0.673 for 1-year, 3-year, and 5-year survival, respectively.

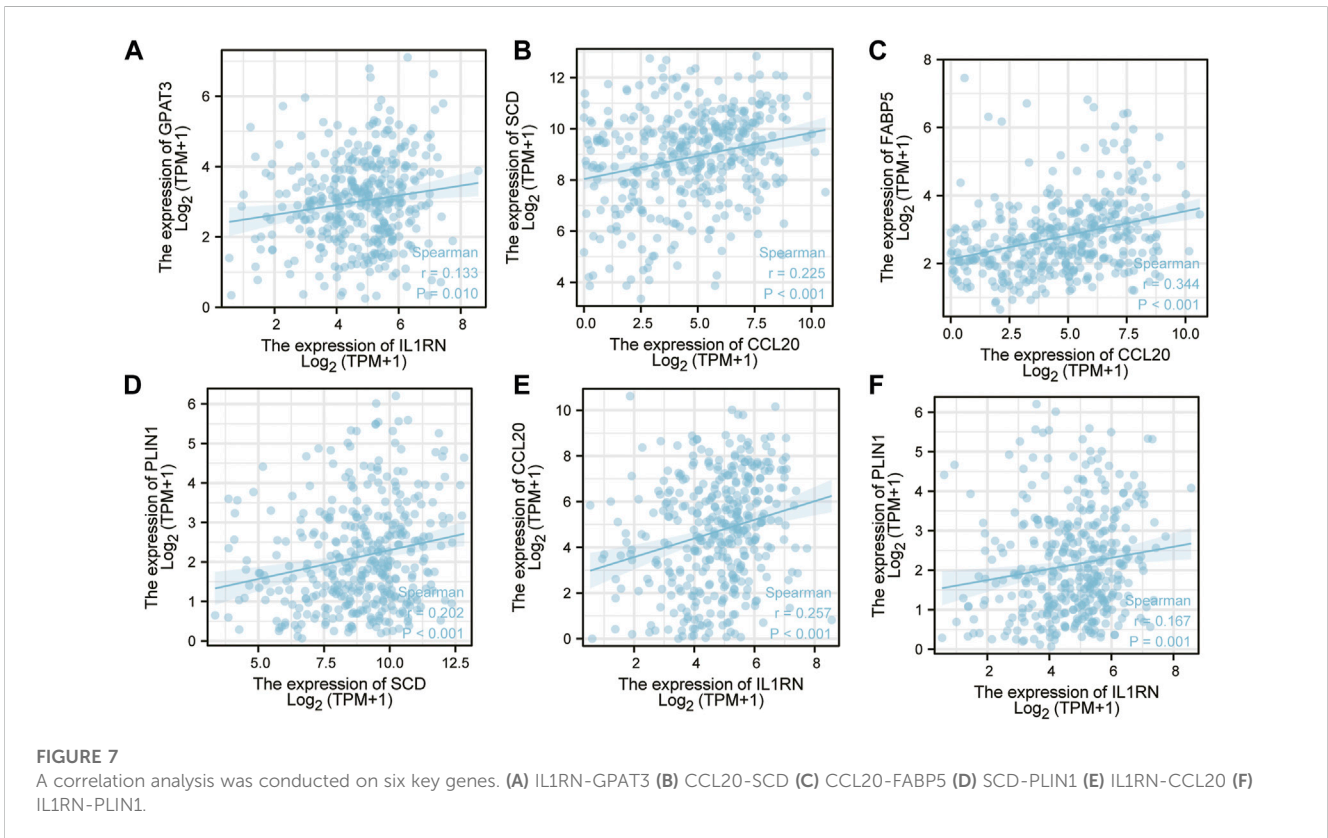
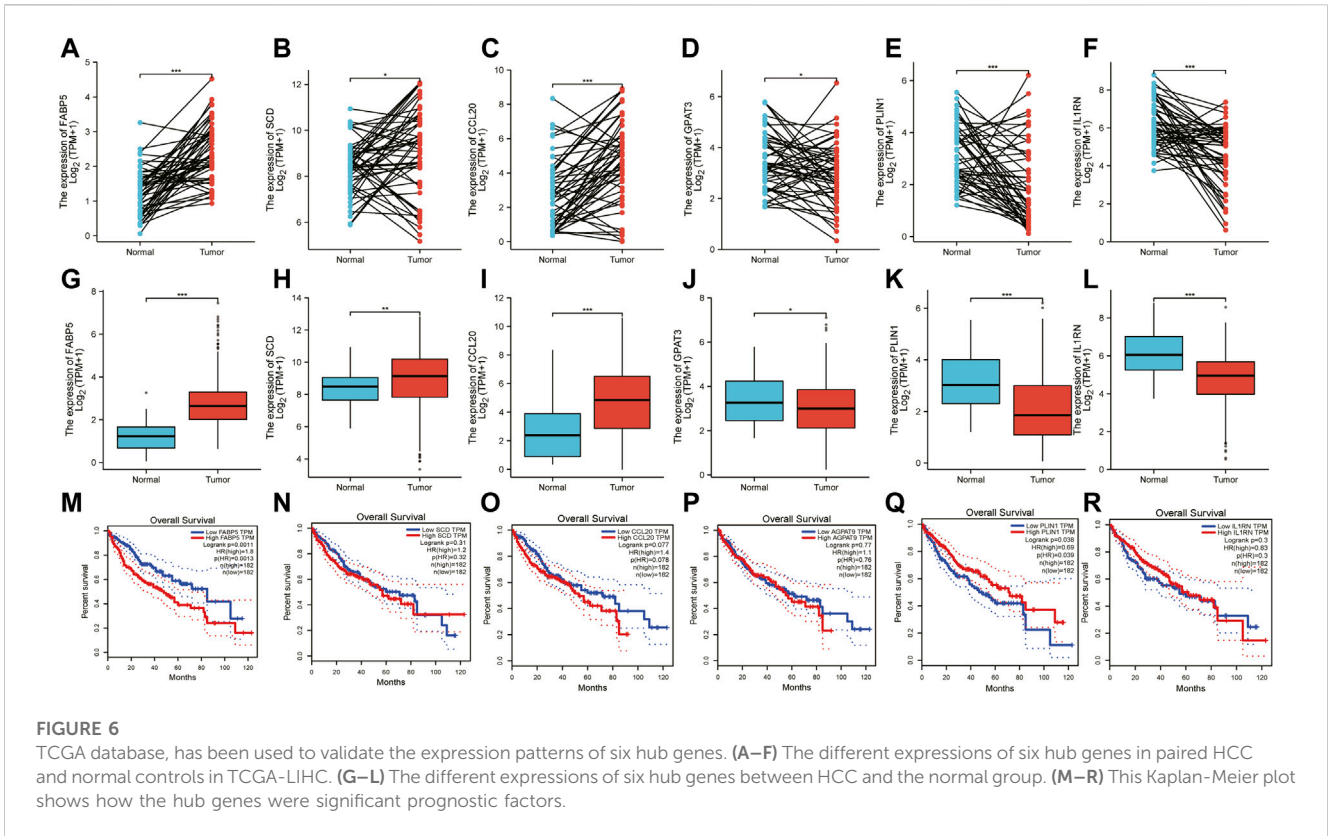
3.8 Correlation analysis of hub gene mRNA levels with tuours-infiltrating immune cells

There are three kinds of cells in the tumor microenvironment: tumor cells, stromal cells, and immune cells that infiltrate the tumor. The TIMER web tool showed that the expression of all six hub genes was associated with infiltrating immune subsets, and the expression of NAFLD and HCC showed the most significant correlation with them. For B-cells, CD4⁺ T-cells, CD8⁺ T-cells, neutrophils, macrophages, and dendritic cells, the expression of FABP5 showed the most significant correlation with them (Figures 9A–F).

4 Discussion

In recent years, increasingly studies have confirmed the link between NAFLD and HCC. A higher risk of HCC has been associated with metabolic syndrome (Agosti et al., 2018, Y. P; Lin et al., 2022, Y; Tan et al., 2019). In the clinic, development and transformation of NAFLD are governed by common law, and its transformation process is also typical of HCC transformation. So far, the mechanism linking NAFLD and HCC remains unclear. Therefore, exploring the molecular mechanisms between NAFLD and other diseases and early identifying and intervening are likely to have significant clinical significance. Bioinformatics analyses comprehensively concentrate primarily on DEGs screening, the development of related protein interaction networks, the screening of genes, and the study of gene associations.

In this study, through searching the datasets of NAFLD and HCC with MRFs from the GEO database, we found 26 common DEGs between these diseases. The results of GO enrichment analysis indicated that the DEGs were mainly enriched in receptor-ligand activity, cytokine activity, monocarboxylic acid binding, and fatty acid binding. Based on the KEGG



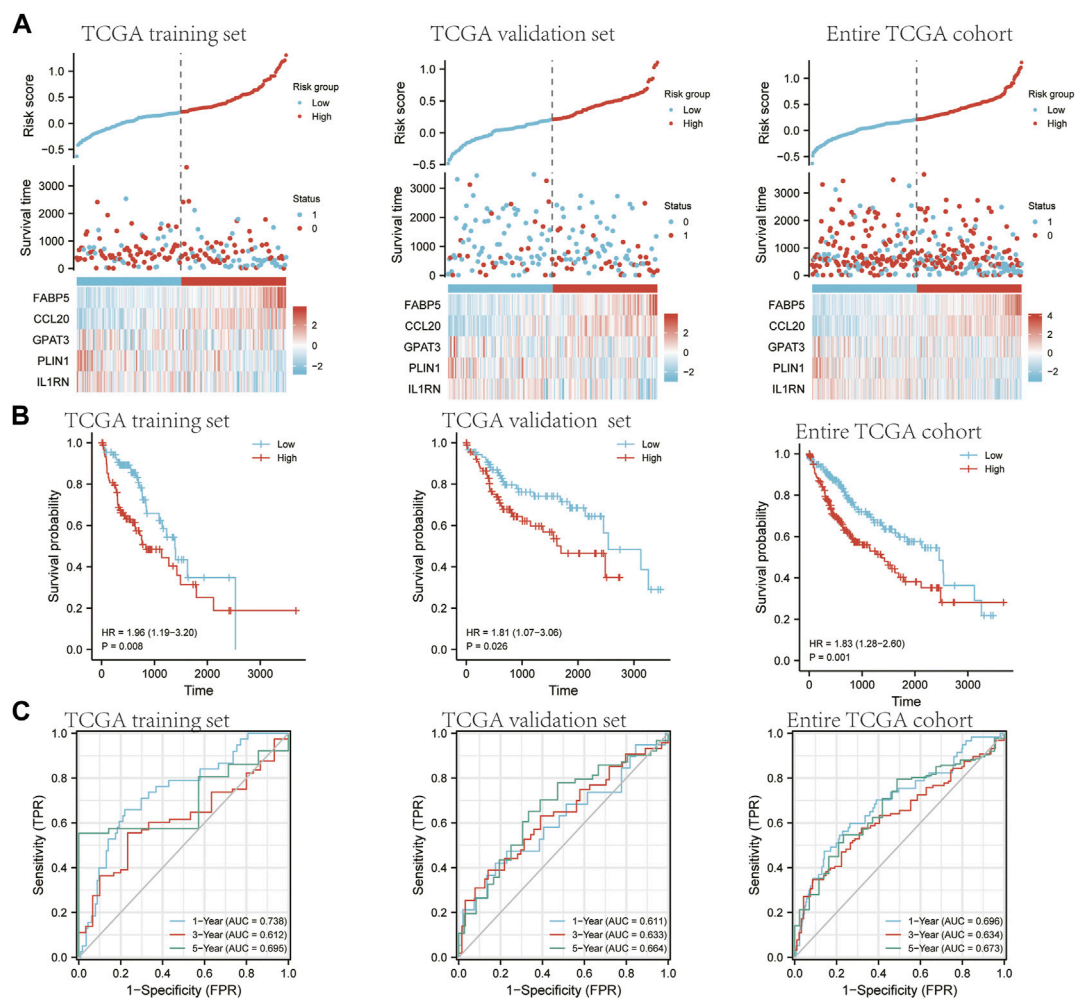


FIGURE 8

An analysis of the five-gene signature model in the TCGA cohort for prognosis. Half of TCGA is set as the training set. The other half of TCGA is designated as a validation set. The entire TCGA cohort is a verification set. (A) A comparison of risk score distribution, survival rates, and gene expression between patients in low- and high-risk groups in TCGA training set and TCGA validation set, entire TCGA cohort. (B) The Kaplan-Meier curves of OS for high-risk and low-risk groups in TCGA training set, TCGA validation set, and the entire TCGA cohort. (C) Time-dependent ROC curve AUCs from the TCGA training set, TCGA validation set, and entire TCGA cohort.

pathway enrichment analysis results, overlapping differential genes are mainly involved in the PPAR signalling pathway. As members of the nuclear receptor superfamily, PPARs can regulate multiple metabolic pathways and are effective targets in the treatment of many metabolic disorders, including NAFLD (Wu et al., 2021). The PPAR signalling pathway is critical to the progression of non-alcoholic steatohepatitis (Zhang Y. et al., 2021). It is possible to predict HCC prognosis using the PPAR signaling pathway effectively, independently, and usefully (Xu et al., 2021).

As a result of the PPI network and module analysis, we identified six key genes, including FABP5, SCD, CCL20, AGPAT9(GPAT3), PLIN1, and IL1RN. The six genes were all changed in both NAFLD patients and HCC patients with MRFs, suggesting that they may play an essential role in NAFLD and HCC with MRFs. An analysis of ROC curves was performed to validate the diagnostic value of NAFLD and HCC. This gene signature has shown good

diagnostic accuracy in both NAFLD and HCC. The expression of FABP5 in NAFLD correlates with histological progression and the loss of hepatic fat during cirrhosis progression in NASH (K. Enooku et al., 2020). Several studies have shown that (fatty acid binding protein 5, FABP5) is highly expressed in HCC. It has been shown that FABP5 promotes angiogenesis and activates the IL6/STAT3/VEGFA pathway in HCCs (F. Liu et al., 2020). Overall survival time for HCC patients was negatively correlated with FABP5 levels in monocytes. The FABP5 protein promotes immune tolerance in patients with HCC by regulating monocytes and tumor-associated monocytes' fatty acid oxidation process via suppressing the PPARα pathway (J. Liu et al., 2022). Our results indicated that FABP5 expression is significantly linked to the overall survival of HCC patients. FABP5 showed the most significant correlation with tumor-infiltrating immune subsets, such as B-cells, CD4⁺ T-cells, CD8⁺ T-cells, neutrophils, macrophages, and dendritic cells. The close association between certain genes,

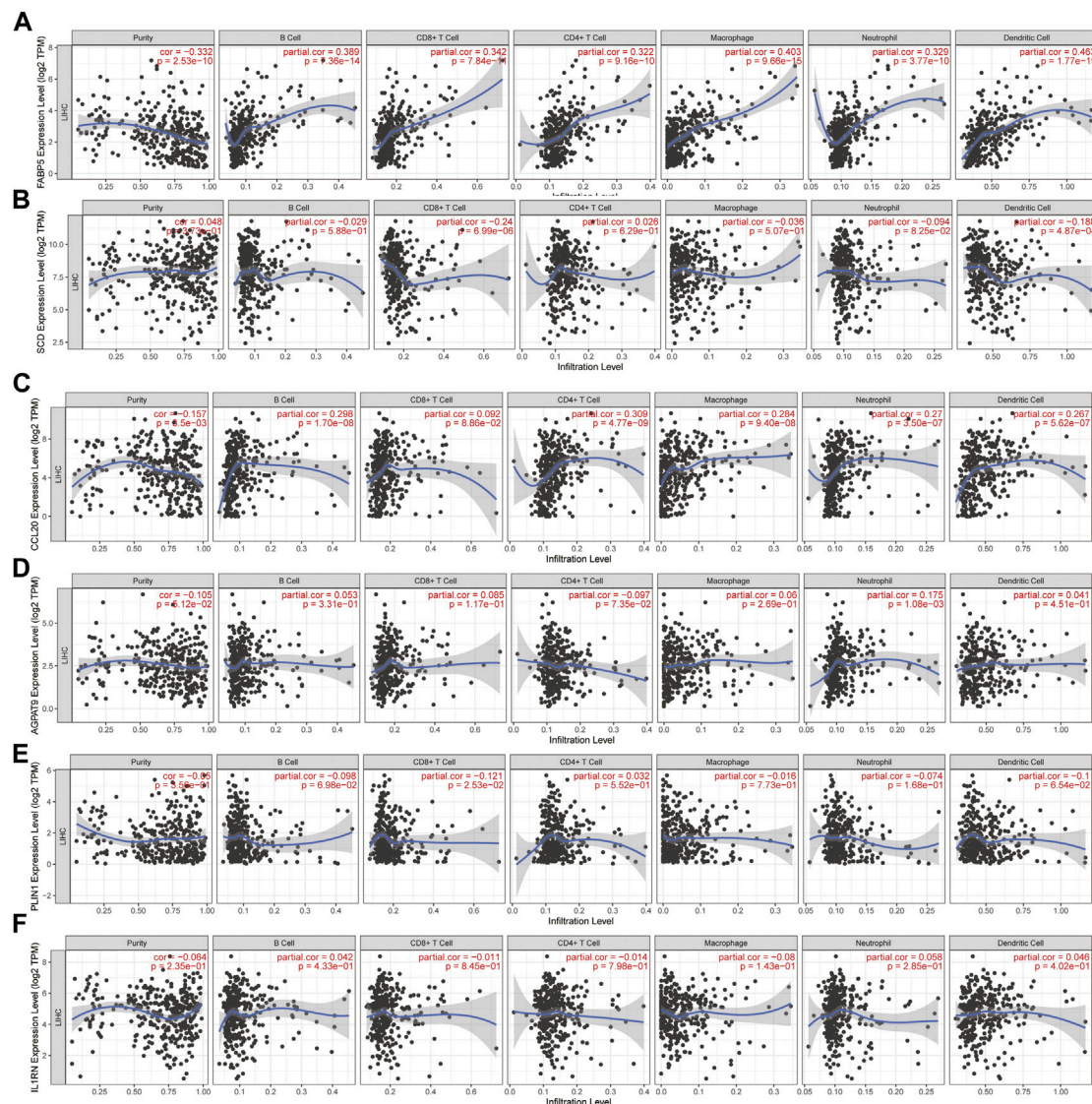


FIGURE 9 Correlation analysis of hub gene mRNA levels with tumor-infiltrating immune cells. (A–F) The correlation of FABP5, SCD, CCL20, AGPAT9, PLIN1, IL1RN mRNA with tumor-infiltrating immune cells. TIMER is the database used for the data (<https://cistrome.shinyapps.io/timer/>).

especially FATP5, and the presence of immune subsets that infiltrate tumor may indicate their importance in immune dysregulation in HCC. The SCD gene encodes an enzyme involved in the biosynthesis of fatty acids, primarily oleic acid. Cancer cells are resistant to chemotherapy-induced apoptosis partly because of the expression of SCD, which is mediated by phosphatidylinositol three kinase/c-Jun N-terminal kinases activation (Bansal et al., 2014). NAFLD fibrosis is known to be associated with an increase in CCL20, an essential inflammatory mediator (Chu et al., 2018). A poor prognosis is related to CCL20 expression in hepatocellular carcinomas after curative resection of cancer (X. Ding et al., 2012).

The GPAT3 (AGPAT9) gene encodes a lysophosphatidic acid acyltransferase family member. The protein encoded by this gene catalyses the conversion of glycerol-3-phosphate to lysophosphatidic acid in triacylglycerol synthesis (J. Cao et al., 2006). Mice with severe congenital generalised

lipodystrophies exhibit insulin resistance and hepatic steatosis when GPAT3 is deficient (Gao et al., 2020). It was found that knocking down GPAT3 effectively inhibited HCC cell growth, induced cell apoptosis, and blocked mTOR signalling in HCC cells.

IL1RN encodes an antagonist protein (IL1RA) that binds to IL-1 as a natural antagonist. IL1RN is involved in developing NAFLD features (M. G. Wolfs et al., 2015). A serum level of L-1RA is associated with inflammation of the liver and higher levels of ALT regardless of obesity, alcohol consumption, or insulin resistance. There is potential for IL-1RA to be used as a non-invasive indicator of NASH inflammatory responses (Pihlajamäki et al., 2012).

PLIN1, an adipocyte-specific protein encoded by this gene, coats lipid storage droplets to protect them until hormone-sensitive lipases can break them down. In adipocytes, PLIN1 is

the major cAMP-dependent protein kinase substrate, and it may inhibit lipolysis when unphosphorylated (J. H. Sohn et al., 2018). NAFLD (non-alcoholic steatohepatitis, NASH) leads to an upregulation of PLIN1. However, it impairs glucose homeostasis and may be protective against lipotoxicity³³ (Carr and Ahima, 2016). Our study indicated that PLIN1 mRNA expression is positively linked to overall survival.

5 Conclusion

Generally, by utilizing biological information research methods, we have identified six key genes for diagnosing NAFLD and HCC with MRFs. Moreover, five key genes were identified for the prognosis of HCC changes and the created gene marker composed of these genes was FABP5, CCL20, and GPAT3 may be the critical dangerous prognostic genes of HCC. PLIN1 and IL1RN are protective prognostic genes of HCC. Nevertheless, since our research is based on data analysis, further experiments would be required to confirm our findings. Nevertheless, we hope that our research findings will contribute to improving the diagnosis and treatment of liver disease associated with NAFLD and HCC progression.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding authors.

Author contributions

XZ, FH, and WC conceived and designed the study. BN, QL, XQ, and YH acquired and analysed the data. FY performed data analysis

and prepared the figures and tables. FY wrote the manuscript. QZ revised the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by the Natural Science Foundation of Xinjiang Uyghur Autonomous Region (2020D01C006). National Natural Science Foundation of China (82260110, 81870449, 82170674, 51933011), China Postdoctoral Science Foundation (2019M653904XB).

Acknowledgments

We acknowledge the GEO, TCGA, STRING, and TIMER databases for their open access.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Agosti, P., Carlo, S., and Mazzocca, A. (2018). Emerging metabolic risk factors in hepatocellular carcinoma and their influence on the liver microenvironment. *Biochim. Biophys. Acta. Mol. Basis Dis.* 1864 (2), 607–617. doi:10.1016/j.bbadis.2017.11.026
- Akahoshi, K., Tanaka, S., Mogushi, K., Shimada, S., Matsumura, S., Akiyama, Y., et al. (2016). Expression of connective tissue growth factor in the livers of non-viral hepatocellular carcinoma patients with metabolic risk factors. *J. Gastroenterol.* 51 (9), 910–922. doi:10.1007/s00535-015-1159-8
- Arendt, B. M., Comelli, E. M., Ma, D. W., Lou, W., Teterina, A., Kim, T., et al. (2015). Altered hepatic gene expression in nonalcoholic fatty liver disease is associated with lower hepatic n-3 and n-6 polyunsaturated fatty acids. *Hepatology* 61 (5), 1565–1578. doi:10.1002/hep.27695
- Awosika, Joy, and Sohal, Davendra (2022). A narrative review of systemic treatment options for hepatocellular carcinoma: State of the art review. *J. Gastrointest. Oncol.* 13 (1), 426–437. doi:10.21037/jgo-21-274
- Bansal, S., Berk, M., Alkhoury, N., DavidPartrick, A., JohnFungAriel Feldstein, J., and Feldstein, A. (2014). Stearoyl-CoA desaturase plays an important role in proliferation and chemoresistance in human hepatocellular carcinoma. *J. Surg. Res.* 186 (1), 29–38. doi:10.1016/j.jss.2013.07.001
- Bustoros, M., Sklaventis-Pistofidis, R., Park, J., Redd, R., Zhitomirsky, B., Dunford, A. J., et al. (2020). Genomic profiling of smoldering multiple myeloma identifies patients at a high risk of disease progression. *J. Clin. Oncol.* 38 (21), 2380–2389. doi:10.1200/JCO.20.00437
- Cao, J., Li, J. L., Li, D., Tobin, J. F., and Gimeno, R. E. (2006). Molecular identification of microsomal acyl-CoA:glycerol-3-phosphate acyltransferase, a key enzyme in de novo triacylglycerol synthesis. *Proc. Natl. Acad. Sci. U. S. A.* 103 (52), 19695–19700. doi:10.1073/pnas.0609140103
- Carr, R. M., and Ahima, R. S. (2016). Pathophysiology of lipid droplet proteins in liver diseases. *Exp. Cell Res.* 340 (2), 187–192. doi:10.1016/j.yexcr.2015.10.021
- Chen, C. L., Yang, H. I., Yang, W. S., Liu, C. J., Chen, P. J., You, S. L., et al. (2008). Metabolic factors and risk of hepatocellular carcinoma by chronic Hepatitis B/C infection: A follow-up study in taiwan. *Gastroenterology* 135 (1), 111–121. doi:10.1053/j.gastro.2008.03.073
- Chiyonobu, N., Shimada, S., Akiyama, Y., Mogushi, K., Itoh, M., Akahoshi, K., et al. (2018). "Fatty acid binding protein 4 (Fabp4) overexpression in intratumoral hepatic stellate cells within hepatocellular carcinoma with metabolic risk factors." *Am. J. Pathol.* 188 (5), 1213–1224. doi:10.1016/j.ajpath.2018.01.012
- Chu, X., Jin, Q., Chen, H., Craig Wood, G., Petrick, A., Strodel, W., et al. (2018). CCL20 is up-regulated in non-alcoholic fatty liver disease fibrosis and is produced by hepatic stellate cells in response to fatty acid loading. *J. Transl. Med.* 16 (1), 108. doi:10.1186/s12967-018-1490-y
- Ding, X., Wang, K., Wang, H., Zhang, G., Liu, Y., Yang, Q., et al. (2012). High expression of CCL20 is associated with poor prognosis in patients with hepatocellular carcinoma after curative resection. *J. Gastrointest. Surg.* 16 (4), 828–836. doi:10.1007/s11605-011-1775-4
- Enooku, K., Tsutsumi, T., Kondo, M., Fujiwara, N., Sasako, T., Shibahara, J., et al. (2020). Hepatic FATP5 expression is associated with histological progression and loss of hepatic fat in NAFLD patients. *J. Gastroenterol.* 55 (2), 227–243. doi:10.1007/s00535-019-01633-2
- Eriksen, R. E., Lim, S. L., McDonnell, E., Shuen, W. H., Vadiveloo, M., White, P. J., et al. (2019). Loss of BCAA catabolism during carcinogenesis enhances

- mTORC1 activity and promotes tumor development and progression. *Cell Metab.* 29 (5), 1151–1165. doi:10.1016/j.cmet.2018.12.020
- Frades, I., Andreasson, E., Mato, J. M., Alexandersson, E., Matthiesen, R., and Martínez-Chantar, M. L. (2015). "Integrative Genomic Signatures of Hepatocellular Carcinoma Derived from Nonalcoholic Fatty Liver Disease" *PLoS ONE* 10 (5), e0124544.
- Gao, M., Liu, Lin, Wang, X., Yin Mak, H., Liu, G., and Yang, H., (2020). GPAT3 deficiency alleviates insulin resistance and hepatic steatosis in a mouse model of severe congenital generalized lipodystrophy. *Hum. Mol. Genet.* 29 (3), 432–443. doi:10.1093/hmg/ddz300
- Gene Ontology Consortium (2006). The gene Ontology (go) project in 2006. *Nucleic Acids Res* 34, D322–D326. doi:10.1093/nar/gkj021
- George, B., Seals, S., and Aban, I. (2014). Survival analysis and regression models. *J. Nucl. Cardiol.* 21 (4), 686–694. doi:10.1007/s12350-014-9908-2
- Heagerty, P. J., Lumley, T., and Pepe, M. S. (2000). Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics* 56 (2), 337–344. doi:10.1111/j.0006-341x.2000.00337.x
- Huang, S. F., Chang, I. C., Hong, C. C., Yen, T. C., Chen, C. L., Wu, C. C., et al. (2018). Metabolic risk factors are associated with non-Hepatitis B non-hepatitis C hepatocellular carcinoma in Taiwan, an endemic area of chronic Hepatitis B. *Hepatol. Commun.* 2 (6), 747–759. doi:10.1002/hep4.1182
- Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2016). KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* 44 (D1), D457–D462. doi:10.1093/nar/gkv1070
- Kurebayashi, Y., Ojima, H., Tsujikawa, H., Kubota, N., Maehara, J., Abe, Y., et al. (2018). Landscape of immune microenvironment in hepatocellular carcinoma and its additional impact on histological and molecular classification. *Hepatology* 68 (3), 1025–1041. doi:10.1002/hep.29904
- Li, T., Fan, J., Wang, B., Traugh, N., Chen, Q., Liu, J. S., et al. (2017). TIMER: A web server for comprehensive analysis of tumor-infiltrating immune cells. *Cancer Res.* 77 (21), e108–e110. doi:10.1158/0008-5472.CAN-17-0307
- Lin, Y. P., Wang, P. M., Chuang, C. H., Yong, C. C., Liu, Y. W., Huang, P. Y., et al. (2022). Metabolic risks are increasing in non-B non-C early-stage hepatocellular carcinoma: A 10-year follow-up study. *Front. Oncol.* 12, 816472. doi:10.3389/fonc.2022.816472
- Liu, F., Liu, W., Zhou, S., Yang, C., Tian, M., Jia, G., et al. (2020). "Identification of Fabp5 as an immunometabolic marker in human hepatocellular carcinoma." *J immunother cancer* 8.
- Liu, J., Sun, B., Guo, K., Yang, Z., Zhao, Y., Gao, M., et al. (2022). "Lipid-Related Fabp5 activation of tumor-associated monocytes fosters immune privilege via Pd-L1 expression on treg cells in hepatocellular carcinoma." *cancer gene ther.*
- Llovet, J. M., Kate Kelley, R., Villanueva, A., AmitSingal, G., Pikarsky, E., Roayaie, S., et al. (2021). Hepatocellular carcinoma. *Nat. Rev. Dis. Prim.* 7 (1), 6. doi:10.1038/s41572-020-00240-3
- Lonardo, A., Byrne, C. D., Caldwell, S. H., Cortez-Pinto, H., and Targher, G. (2016). Global epidemiology of nonalcoholic fatty liver disease: Meta-analytic assessment of prevalence, incidence, and outcomes. *Hepatology* 64 (4), 1388–1389. doi:10.1002/hep.28584
- MancinaMargherita, R., Dongiovanni, P., Petta, S., Pingitore, P., Meroni, M., et al. (2016). The MBOAT7-TMC4 variant rs641738 increases risk of nonalcoholic fatty liver disease in individuals of European descent. *Gastroenterology*, 150 (5), 1219. doi:10.1053/j.gastro.2016.01.032
- Morse, M. A., Sun, W., Kim, R., He, A. R., Abada, P. B., Mynderse, M., et al. (2019). The role of angiogenesis in hepatocellular carcinoma. *Clin. Cancer Res.* 25 (3), 912–920. doi:10.1158/1078-0432.CCR-18-1254
- Pihlajamäki, J., Kuulasmaa, T., Kaminska, D., Simonen, M., Kärjä, V., Grönlund, S., et al. (2012). Serum interleukin 1 receptor antagonist as an independent marker of non-alcoholic steatohepatitis in humans. *J. Hepatol.* 56 (3), 663–670. doi:10.1016/j.jhep.2011.10.005
- Pouwels, S., Sakran, N., Graham, Y., Leal, A., Pintar, T., Yang, W., et al. (2022). Non-alcoholic fatty liver disease (NAFLD): A review of pathophysiology, clinical management and effects of weight loss. *BMC Endocr. Disord.* 22 (1), 63. doi:10.1186/s12902-022-00980-1
- Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., et al. (2015). Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 43 (7), e47. doi:10.1093/nar/gkv007
- Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J. C., et al. (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinforma.* 12, 77. doi:10.1186/1471-2105-12-77
- Sohn, J. H., Lee, Y. K., Han, J. S., Jeon, Y. G., Kim, J. I., Choe, S. S., et al. (2018). "Perilipin 1 (Plin1) deficiency promotes inflammatory responses in lean adipose tissue through lipid dysregulation." *J. Biol. Chem.* 293 (36), 13974–13988. doi:10.1074/jbc.RA118.003541
- Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., et al. (2015). STRING v10: Protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* 43, D447–D452. doi:10.1093/nar/gku1003
- Tan, Y., Zhang, X., Zhang, W., Tang, L., Yang, H., Yan, K., et al. (2019). The influence of metabolic syndrome on the risk of hepatocellular carcinoma in patients with chronic hepatitis B infection in mainland China. *Cancer Epidemiol. Biomarkers Prev.* 28 (12), 2038–2046. doi:10.1158/1055-9965.EPI-19-0303
- Tang, Z., Li, C., Kang, B., Gao, G., Li, C., and Zhang, Z. (2017). Gepia: A web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res.* 45 (W1), W98–W102. doi:10.1093/nar/gkx247
- Trierweiler, C., Hockenjos, B., Zatloukal, K., Thimme, R., Blum, H. E., Wagner, E. F., et al. (2016). The transcription factor c-JUN/AP-1 promotes HBV-related liver tumorigenesis in mice. *Cell Death Differ.* 23 (4), 576–582. doi:10.1038/cdd.2015.121
- Wattacheril, J. (2020). Extrahepatic manifestations of nonalcoholic fatty liver disease. *Gastroenterol. Clin. North Am.* 49 (1), 141–149. doi:10.1016/j.gtc.2019.10.002
- Wolfs, M. G., Gruben, N., Rensen, S. S., Verdam, F. J., Greve, J. W., Driessen, A., et al. (2015). Determining the association between adipokine expression in multiple tissues and phenotypic features of non-alcoholic fatty liver disease in obesity. *Nutr. Diabetes* 5 (2), e146. doi:10.1038/nutd.2014.43
- Wu, L., Li, J., Feng, J., Ji, J., Yu, Q., Li, Y., et al. (2021). Crosstalk between PPARs and gut microbiota in NAFLD. *136*, 111255. doi:10.1016/j.biopha.2021.111255
- Xu, W., Chen, Z., Liu, G., Dai, Y., Xu, X., Ma, D., et al. (2021). Identification of a potential PPAR-related multigene signature predicting prognosis of patients with hepatocellular carcinoma. *PPAR Res.* 2021, 6642939. doi:10.1155/2021/6642939
- Zhang, J., Liu, X., Zhou, W., Lu, S., Wu, C., Wu, Z., et al. (2021a). Identification of key genes associated with the process of hepatitis B inflammation and cancer transformation by integrated bioinformatics analysis. *Front. Genet.* 12, 654517. doi:10.3389/fgene.2021.654517
- Zhang, Y., Ge, X., Li, Y., Zhang, B., Wang, P., Hao, M., et al. (2021b). TWIST2 and the PPAR signaling pathway are important in the progression of nonalcoholic steatohepatitis. *Lipids Health Dis.* 20 (1), 39. doi:10.1186/s12944-021-01458-0