



Identification of Hub Genes in Liver Hepatocellular Carcinoma Based on Weighted Gene Co-expression Network Analysis

Jiawei Sun¹ · Zizhen Zhang² · Jiaru Cai¹ · Xiaoping Li¹ · Xiaoling Xu¹

Received: 9 January 2023 / Accepted: 5 April 2024
© The Author(s) 2024

Abstract

Liver hepatocellular carcinoma (LIHC) is a malignant cancer with high incidence and poor prognosis. To investigate the correlation between hub genes and progression of LIHC and to provide potential prognostic markers and therapy targets for LIHC. Our study mainly used The Cancer Genome Atlas (TCGA) LIHC database and the gene expression profiles of GSE54236 from the Gene Expression Omnibus (GEO) to explore the differential co-expression genes between LIHC and normal tissues. The differential co-expression genes were extracted by Weighted Gene Co-expression Network Analysis (WGCNA) and differential gene expression analysis methods. The Genetic Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) were carried out to annotate the function of differential genes. Then the hub genes were validated using protein-protein interaction (PPI) network. And the expression level and prognostic analysis were performed. The probable associations between the expression of hub genes and both tumor purity and infiltration of immune cells were explored by TIMER. A total of 68 differential co-expression genes were extracted. These genes were mainly enriched in complement activation (biological process), collagen trimer (cellular component), carbohydrate binding and receptor ligand activity (molecular function) and cytokine–cytokine receptor interaction. Then we demonstrated that the 10 hub genes (CFP, CLEC1B, CLEC4G, CLEC4M, FCN2, FCN3, PAMR1 and TIMD4) were weakly expressed in LIHC tissues, the qRT-PCR results of clinical samples showed that six genes were significantly downregulated in LIHC patients compared with adjacent tissues. Worse overall survival (OS) and disease-free survival (DFS) in LIHC patients were associated with the lower expression of CFP, CLEC1B, FCN3 and TIMD4. Ten hub genes had positive association with tumor purity. CFP, CLEC1B, FCN3 and TIMD4 could serve as novel potential molecular targets for prognosis prediction in LIHC.

Jiawei Sun, Zizhen Zhang contributed equally to this work and share first authorship. Xiaoping Li, Xiaoling Xu contributed equally to this work and share last authorship.

Extended author information available on the last page of the article

Keywords Weighted gene co-expression network analysis · Bioinformatics analysis · Liver hepatocellular carcinoma · Hub gene · Protein–protein interaction

Introduction

Liver hepatocellular carcinoma (LIHC) is one of the most common malignancies worldwide, ranking sixth for incidence and third for mortality (Kulik and El-Serag 2019; Sung et al. 2021). Based on the global cancer statistics report, primary liver cancer ranked as the sixth most frequently diagnosed cancer and the third leading cause of cancer-related mortality globally in 2020, LIHC represents 75–85% of all liver cancer cases. (Sung et al. 2021). Chronic infection with hepatitis B virus (HBV) or hepatitis C virus (HCV), aflatoxin-contaminated foods, heavy alcohol intake, excess body weight, type 2 diabetes, and smoking are main risk factors for LIHC (Kulik and El-Serag 2019; Yang et al. 2019). The commonly used therapies for LIHC include hepatectomy, liver transplantation and ablative therapy (Chakraborty and Sarkar 2022; Chen et al. 2020). Although there are many new methods for the diagnosis and treatment of LIHC, the prognosis is still poor due to high rates of recurrence and metastasis. An important reason for poor prognosis is the lack of an accurate prognosis grading system and an effective early screening method for LIHC. Therefore, it is reasonable to evaluate the value of potential biomarkers in the diagnosis and treatment of LIHC.

With the development of genome technology, a variety of tools (Langfelder and Horvath 2008; Newman et al. 2015; van Ijzendoorn et al. 2019) have been used for the study of the molecular mechanism of diseases and biomarker identification. Weighted gene co-expression network analysis (WGCNA) is a novel bioinformatics approach used to describe patterns of gene association between different samples. It can cluster genes and form modules by similar gene expression patterns and analyze the relationship between modules and specific features to identify candidate biomarker genes (Pei et al. 2017). WGCNA is helpful in understanding the molecular mechanisms of cancer and identifying reliable biomarkers for more effective diagnosis, prognosis and treatment.

In this investigation, we comprehensively screened pivotal central genes implicated in LIHC, thereby establishing a robust foundation for the analysis of differentially co-expressed genes. This framework facilitates a deeper understanding of the etiological factors and potential molecular mechanisms underlying LIHC pathogenesis.

Methods

Data Processing from TCGA and GEO Database

In order to screen the differential expression genes between LIHC and normal control, we obtained the data of gene expression profiles in TCGA (<https://portal.gdc.cancer.gov/projects/TCGA-LIHC>) (including 50 normal tissues and 374 tumor tissues) and GEO (<https://www.ncbi.nlm.nih.gov/gds>) GSE54236 (including 77 adjacent nontu-

morous samples and 78 LIHC samples; platform: GPL6480), After removing repetitive probes of the same gene, an interpreted gene expression profile of 18,076 genes was obtained for the following analysis.

Co-expression Network Construction by WGCNA

The *WGCNA* package in R was employed to construct a gene co-expression network using the gene expression data profiles of TCGA-LIHC and GSE54236 (Chen et al. 2017; Langfelder and Horvath 2008). The package includes functions for network construction, module detection, gene selection, calculations of topological properties, data simulation, visualization, and interfacing with external software (Langfelder and Horvath 2008). First, the expression levels of individual transcripts were transformed to generate a similarity matrix based on Pearson correlation values between paired genes. Then, the similarity matrix is transformed into the adjacency matrix. $\beta=3$ and 20 were selected as the soft threshold power. The adjacency matrix is made from the following formula: $a_{ij} = |c_{ij}|^\beta$. Next the adjacency matrix transformed to corresponding topological overlap matrix (TOM). The genes were divided into different modules by dynamic mixed cutting method, and the minimum module size cut-off value of 30 was simulated.

Identification of DEGs and Validation of Co-expression Modules

To explore differentially expressed genes (DEGs) in TCGA-LIHC and GSE54236 data sets, the R-package *limma*, an integrated solution of analysis for RNA-Sequencing and microarray data, was used to filter out DEGs in LIHC (Li et al. 2020). $|\log_{2}FC| \geq 1.0$ and $P < 0.05$ were defined for screening DEGs. A volcano plot using the R package *ggplot2* was constructed to reveal the differentially expressed genes. The overlapped genes between the DEGs extracted from the co-expression network were presented by Venn diagram.

Function Enrichment Analysis

In order to conduct the Genetic Ontology (GO) enrichment and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis, R Package *clusterProfiler* was used to explore the functions between the selected genes (“The Gene Ontology Resource: 20 years and still GOing strong,” 2019). Statistical significance of GO terms or KEGG pathways was considered as $P < 0.05$. The GO annotation contains biological process (BP), cellular component (CC) and molecular function (MF).

Construction of PPI Network and Screening of Hub Genes

To construct a PPI network of the identified differentially expressed genes (DEGs), the online database Search Tool for the Retrieval of Interacting Genes (STRING) (Szklarczyk et al. 2019) was used. Genes with confidence score ≥ 0.4 were chosen to build a network model by Cytoscape visualization software (v3.7.2) (Doncheva et al. 2019; Shannon et al. 2003). Maximal Clique Centrality (MCC), an effective way to

find a hub node, was calculated by CytoHubba (Chin et al. 2014), and the 10 genes with the highest MCC values were identified as the hub genes.

Human LIHC Samples

This study included 10 pairs of LIHC tissue and paired adjacent tissue samples, which were obtained from patients who underwent liver surgery at Shulan Hospital of Zhejiang Shuren University. Ethical approval for human subjects was obtained from the research ethics committee of Zhejiang Shuren University (approval number: 2024/041), and informed consent was obtained from each patient.

RNA Extraction and qRT-PCR

Total RNA from each sample was extracted applying TRIzol reagent (Thermo Fisher, USA). For qRT-PCR, a LightCycler 480 PCR System (Roche, USA) with FastStart Universal SYBR Green Master (Roche, Switzerland) was used. The PCR conditions were: 95 °C pre-denaturation for 30 s, followed by 39 cycles. Each cycle contained 95 °C denaturation for 5 s, 55 °C annealing for 30 s, and 72 °C extension for 30 s. Data were analyzed with the $2^{-\Delta\Delta C_t}$ method with GAPDH as an internal reference. Supplementary Table S1 listed the primer sequences.

Validation and Survival Analysis of Hub Genes

First, we used GEPIA2 (<http://gepia2.cancer-pku.cn/>) (Tang et al. 2019) to verify the expression pattern of hub genes in LIHC tissues and normal tissues from TCGA database. Gene expression data of liver cancer patients with clinical information was downloaded from the International Cancer Genome Consortium (ICGC) (<http://icgc.org/>), and analyzed to validate expression level of hub genes. Based on data from TCGA database, LIHC patients were divided into two groups according to the median expression level of hub gene. Then Kaplan-Meier survival analysis was performed using survival package in R software to detect the relationship between overall survival (OS) and hub genes. In addition, the relationship between disease-free survival (DFS) and hub gene expression in LIHC patients was determined with the online tool GEPIA2 (Tang et al. 2019). $P < 0.05$ was regarded as the judgment criterion of statistically significant survival-related hub genes.

Immune Infiltration Analysis TIMER

To find the correlation between the expression of the hub genes and tumor infiltrating immune cell subtypes including B cells, CD4⁺ T cells, CD8⁺ T cells, neutrophils, macrophages, and dendritic cells, we utilized the online comprehensive tool Tumor Immune Estimation Resource (TIMER, <http://timer.cistrome.org/>), an online systematic analysis database containing 10,897 samples across 32 cancer types from the TCGA database (Li et al. 2016, 2017, 2020a, b).

Results

Weighted Gene Co-expression Modules Construction

Using the WGCNA package, we constructed the gene expression data profiles of TCGA-LIHC and GSE54236 to gene co-expression networks. In the present study, each module was represented by a color, we emerged 9 modules in the TCGA-LIHC (Fig. 1A) and 27 modules in the GSE54236 (Fig. 2A) from the analysis. The results obtained from the preliminary analysis of the module-trait relationships showed that the blue module in the TCGA-LIHC (Fig. 1B) and brown module in the GSE54236 (Fig. 2B) were found to have the strongest association with normal tissues (blue module: $r=0.77$, $p=3e-85$; brown module: $r=0.57$, $p=2e-15$).

Identification of Genes between the DEGs and Co-expression Modules

We identified 2704 DEGs in TCGA-LIHC (Fig. 3A) and 691 DEGs in GSE54236 (Fig. 3B) using volcano plot. And 3103 and 1450 co-expression genes were separately selected in the blue module of TCGA-LIHC and brown module of GSE54236. A heatmap is a simple yet effective way to compare the content of multiple major gene lists. The heatmap showed the distribution of the top 100 DEGs in TCGA-LIHC and GSE54236 (Fig. 3C, D). To overlap the gene in two co-expression modules and DEG list, a total of 68 overlapping genes were performed by Venn diagram (Fig. 3E).

Functional Enrichment Analyses

To reveal the potential functions of the 68 genes, we divided genes into several enriched gene sets based on the Gene Ontology (GO) analysis. In the biological process (BP) analysis, the 68 genes mainly involved in complement activation. In the result of the cellular component (CC), these genes were mainly enriched in collagen trimer. The molecular function (MF) of 68 genes was mainly enriched in carbohydrate binding and receptor ligand activity (Fig. 4A). Meanwhile we conducted Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis, showing that cytokine–cytokine receptor interaction was widely related to the 68 genes (Fig. 4B).

Construction of the PPI Network

In order to find the hub genes, we performed the PPI analysis of overlapped genes the utilized the “String” website tool (Fig. 5A). Among the 11 topological analysis methods provided by CytoHubba, MCC performed better in predicting the accuracy of essential proteins from PPI networks. Using the MCC algorithm of CytoHubba plugin, the top 10 highest-scored genes, including CFP (Complement Factor Properdin), CLEC1B (C-type lectin domain family 1 member B), CLEC4G (C-type lectin domain family 4 member G), CLEC4M (C-type lectin domain family 4 member M), FCN2(Ficolin 2), FCN3(Ficolin 3), LYVE1(Lymphatic Vessel Endothelial Hyaluronan Receptor 1), MARCO(Macrophage Receptor With Collagenous Structure), PAMR1(Peptidase Domain Containing Associated With Muscle Regeneration 1),

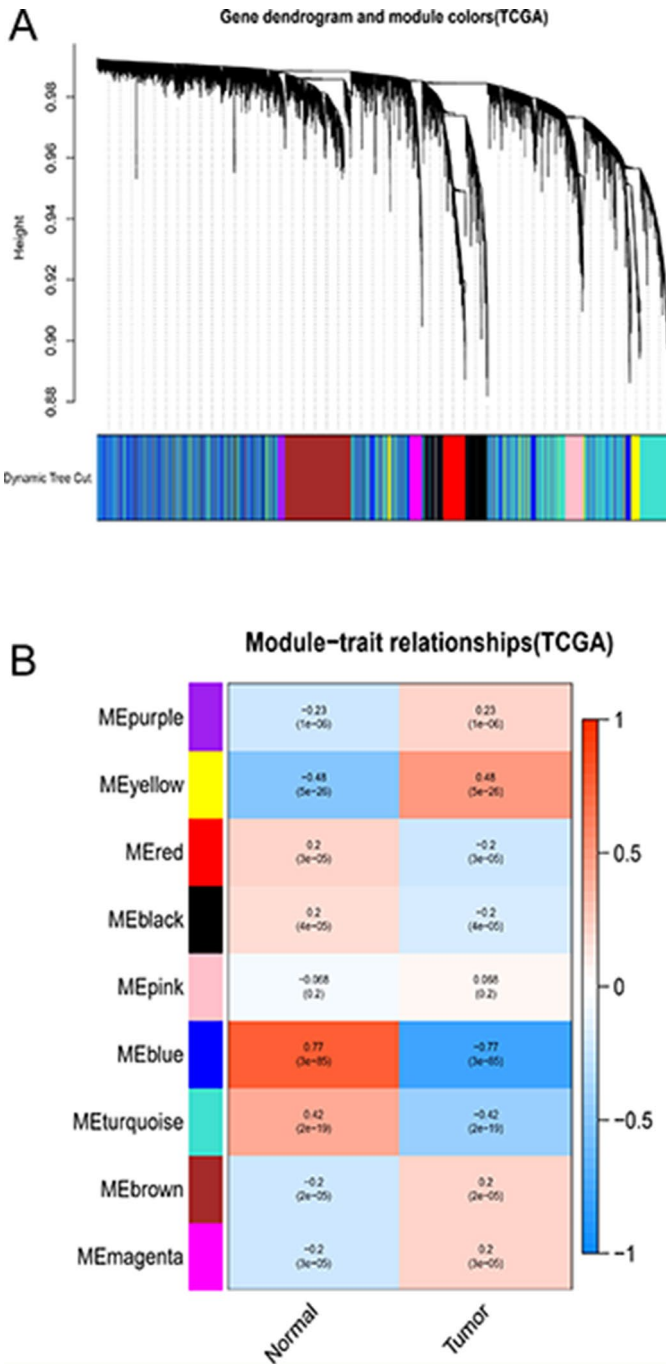


Fig. 1 Identification of key modules correlated with clinical information in the TCGA-LIHC dataset through WGCNA. **(A)** Cluster dendrogram of all co-expression network modules based on a dissimilarity measure (1-TOM). Each co-expression module was represented by a color. **(B)** Heatmap of module-trait relationships. Each cell contains the corresponding correlation and *P*-value

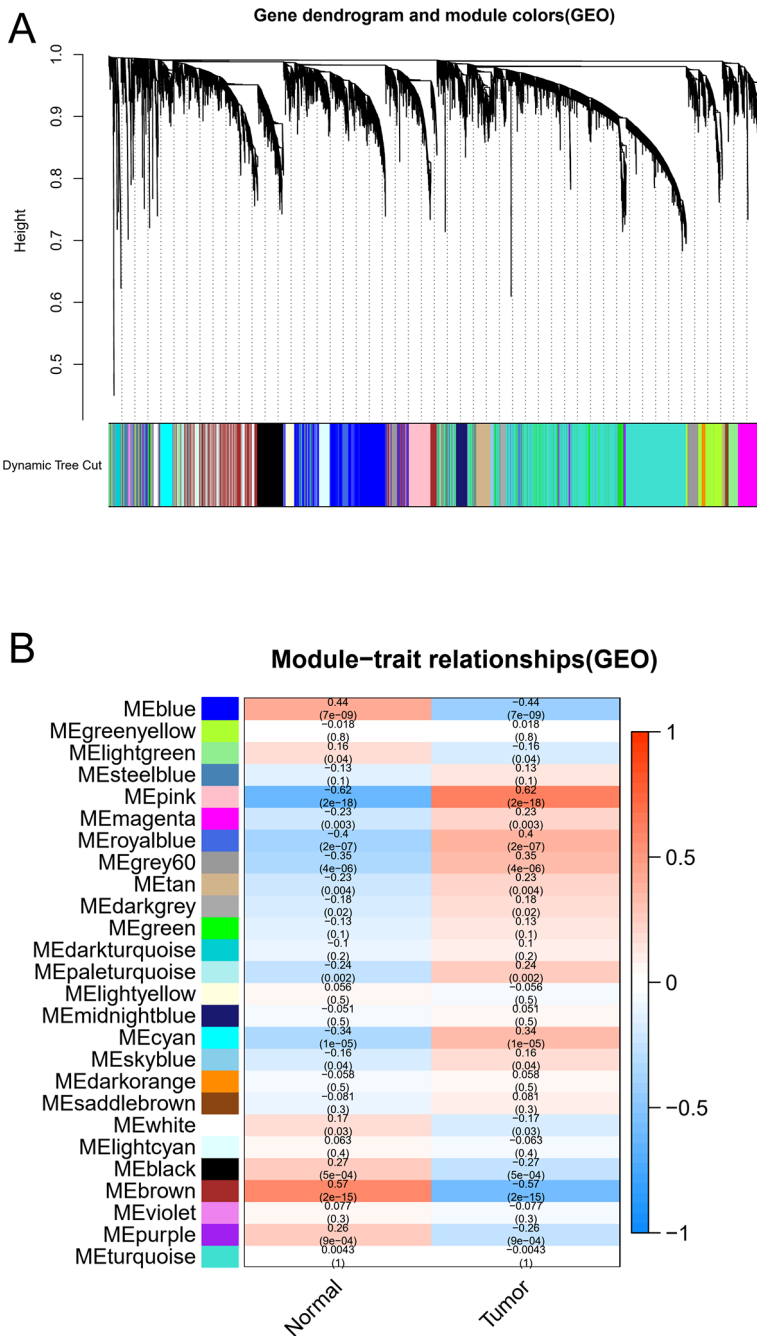


Fig. 2 Identification of key modules correlated with clinical information in the GSE54236 dataset through WGCNA. (A) Cluster dendrogram of all co-expression network modules based on a dissimilarity measure (1-TOM). Each co-expression module was represented by a color. (B) Heatmap of module-trait relationships. Each cell contains the corresponding correlation and P-value

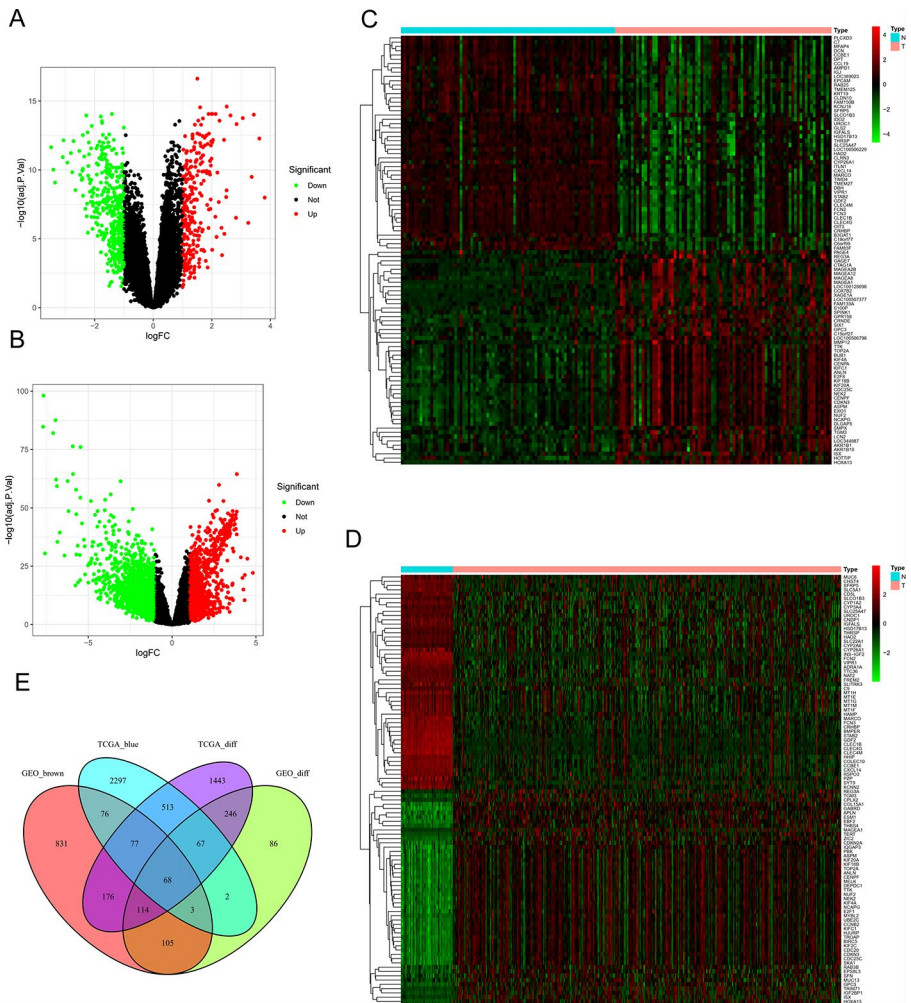


Fig. 3 Identification of differentially expressed genes (DEGs) among the TCGA and GSE54236 datasets of LIHC. **(A)** Volcano plot of DEGs in the TCGA dataset. **(B)** Volcano plot of DEGs in the GSE54236 dataset. $|\logFC| \geq 1.0$ and $adj.P < 0.05$ were defined as the cut-off criteria Green and red indicated low and high expression in LIHC, respectively. Black indicated that those genes no difference between LIHC and normal tissues. **(C)** Heatmap of DEGs in the TCGA dataset. **(D)** Heatmap of DEGs in the GSE54236 dataset. **(E)** The Venn diagram of 68 overlapping genes among two DEG lists and two co-expression modules

TIMD4(T cell immunoglobulin and mucin domain containing 4) were selected as the hub genes from the PPI network (Fig. 5B).

Validation and Survival Analysis of Hub Genes and Expression of Hub Genes

Then we verified the expression levels of the hub genes among patients from the TCGA and ICGC database. All 10 hub genes were found to be significantly down-

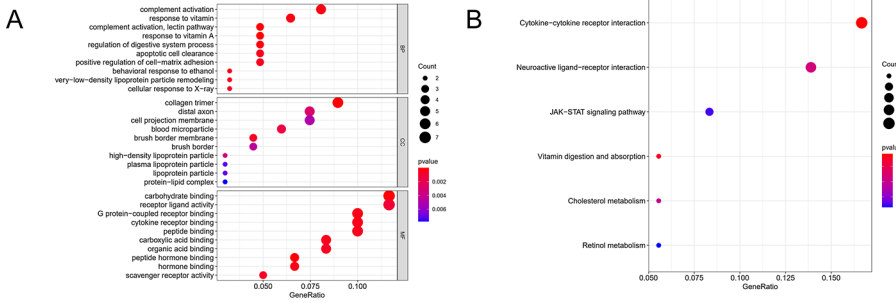


Fig. 4 GO functional and KEGG pathway enrichment analysis for the candidate genes. **(A)** Biological process (BP), cellular component (CC), and molecular function (MF) of GO enrichment. **(B)** KEGG pathway enrichment analysis. The color indicates the adjusted P-values, and the size of the spots indicates the gene number

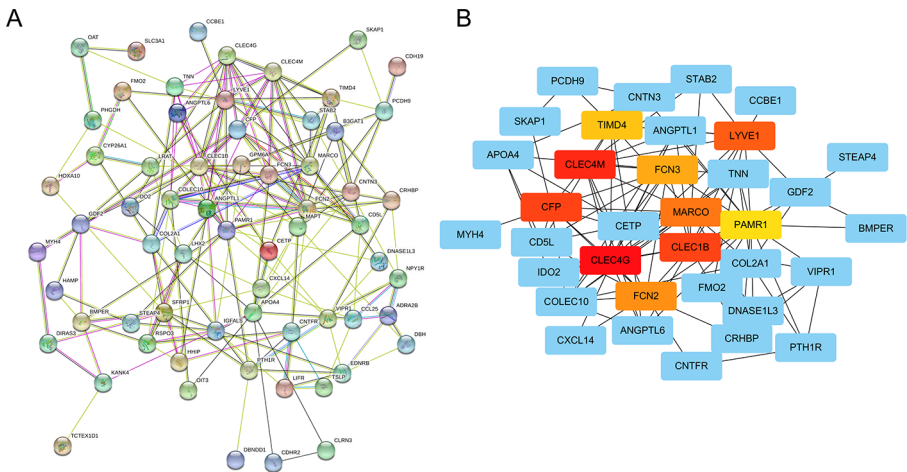


Fig. 5 The enrichment analysis of the candidate genes by protein interaction (PPI) network. **(A)** PPI network. **(B)** Identification of 10 hub genes from PPI network utilizing maximal clique centrality (MCC) algorithm. Red nodules represent MCC high genes and yellow nodules represent MCC low genes

regulated in LIHC compared to normal tissue (Fig. 6; Table 1). More convincingly, the qRT-PCR results of clinical samples showed that compared with adjacent tissues six genes (CFP, CLEC1B, CLEC4G, CLEC4M, FCN3, TIMD4) were significantly downregulated in LIHC patients ($P < 0.005$) (Fig. 7). However, the other for genes (LYVE1, MARCO, PAMR1, FCN2) did not show significant trends. The OS and DFS analyses were performed by Kaplan Meier plot to identify the effect of hub genes on the prognosis from PPI network complexes and WGCNA modules. The results showed that low expression of CFP (Fig. 8A), CLEC1B (Fig. 8B), CLEC4G (Fig. 8C), CLEC4M (Fig. 8D), FCN2 (Fig. 8E), FCN3 (Fig. 8F), PAMR1 (Fig. 8I) and TIMD4 (Fig. 8J) was associated with poor OS in LIHC patients ($P < 0.05$). Besides, among the hub genes, the lower expression level of CFP (Fig. 9A), CLEC1B

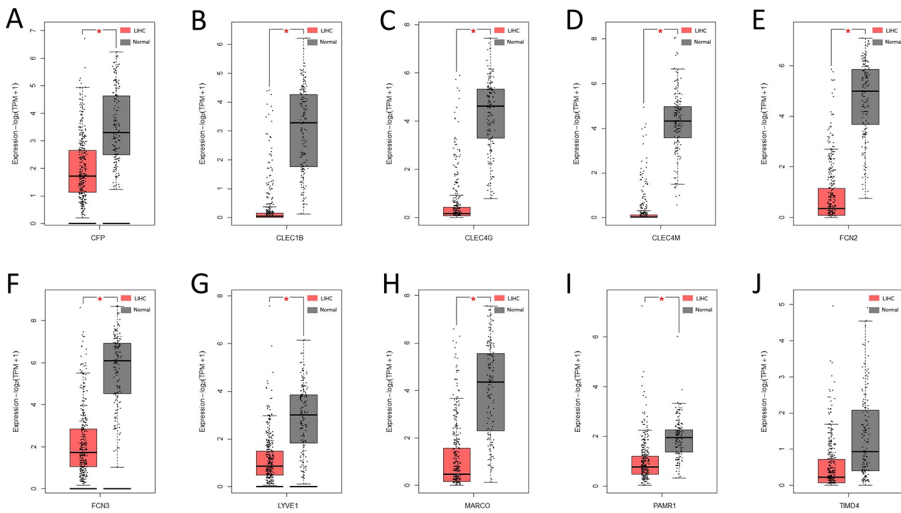


Fig. 6 Expression levels of the 10 hub genes among LIHC and normal tissues from the TCGA dataset. (A) CFP, (B) CLEC1B, (C) CLEC4G, (D) CLEC4M, (E) FCN2, (F) FCN3, (G) LYVE1, (H) MARCO, (I) PAMR1, (J) TIMD4. * $P < 0.05$. $P < 0.05$ denotes significance

Table 1 Expression of the Hub gene in LIHC and adjacent tissues from the ICGC database

Gene	Type	Expression $\log_2(\text{TPM} + 1)$	P value
CFP	LIHC	1.49 ± 0.89	8.95E-98
	Adjacent	3.89 ± 0.76	
CLEC1B	LIHC	0.59 ± 0.96	4.41E-107
	Adjacent	3.88 ± 1.05	
CLEC4G	LIHC	0.75 ± 1.18	1.20E-125
	Adjacent	4.71 ± 0.98	
CLEC4M	LIHC	0.52 ± 1.14	8.52E-110
	Adjacent	4.16 ± 1.10	
FCN2	LIHC	0.96 ± 1.30	4.77E-115
	Adjacent	4.90 ± 1.05	
FCN3	LIHC	2.091 ± 1.69	1.03E-101
	Adjacent	6.426 ± 1.13	
LYVE1	LIHC	1.26 ± 0.93	1.95E-62
	Adjacent	3.42 ± 1.10	
MARCO	LIHC	1.19 ± 1.42	7.06E-80
	Adjacent	4.88 ± 1.51	
PAMR1	LIHC	0.83 ± 0.63	2.70E-61
	Adjacent	2.10 ± 0.62	
TIMD4	LIHC	0.74 ± 0.84	9.22E-40
	Adjacent	2.23 ± 1.06	

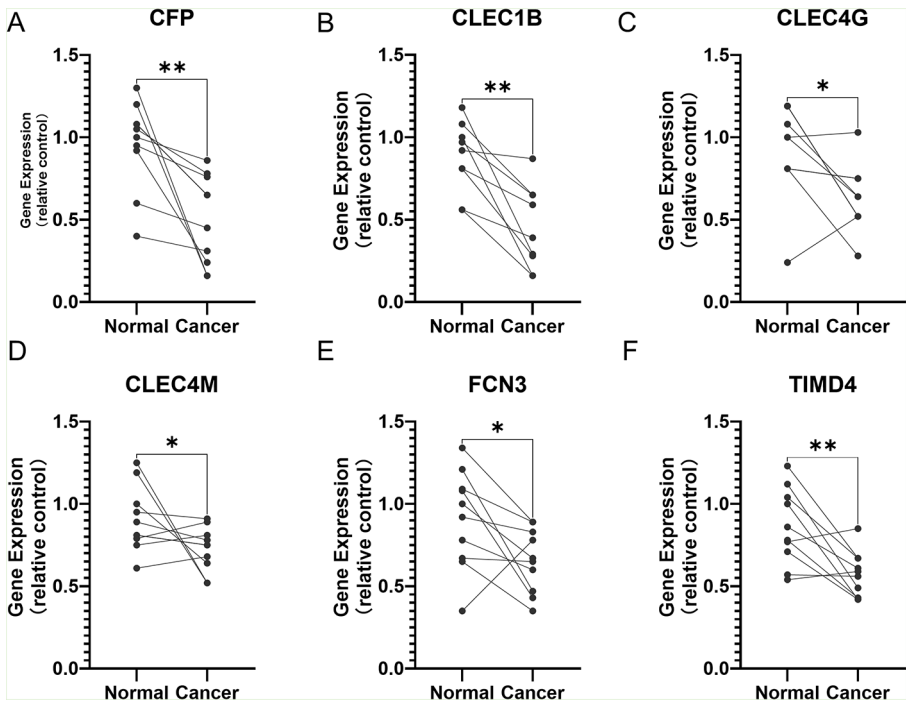


Fig. 7 The expression of hub genes in LIHC patients. **(A)** CFP, **(B)** CLEC1B, **(C)** CLEC4G, **(D)** CLEC4M, **(E)** FCN3, **(F)** TIMD4. (* $P < 0.05$, ** $P < 0.01$)

(Fig. 9B), FCN3 (Fig. 9F) and TIMD4 (Fig. 9J) was significantly associated with worse DFS of LIHC patients ($P < 0.05$).

Association between Hub Genes with Immune Infiltration Level

Infiltrating immune cells are an important part of tumor microenvironment besides tumor cells and stromal cells. The online search tool TIMER was utilized to explore probable associations between the expression of hub genes and both tumor purity and infiltration of immune cells. All those 10 hub genes had positive association with tumor purity (Fig. 10), while no or weak correlation with B cell infiltration, CD4+T cells, CD8+T cells, neutrophils, macrophages, and dendritic cells (Fig. 10).

Discussion

LIHC remains one of the most malignant cancers with poor prognosis. WGCNA has been used to explore biomarkers related to pathogenesis, diagnosis and prognosis of LIHC. Immune cell infiltration related biomarkers in LIHC including CCL5, CXCR6, CD3E and LCK were identified using WGCNA(Bai et al. 2020; Gao et al. 2023). Another study found that 13 hub genes (GTSE1, PLK1, NCAPH, SKA3, LMNB2, SPC25, HJURP, DEPDC1B, CDCA4, UBE2C, LMNB1, PRR11, and SNRPD2) had

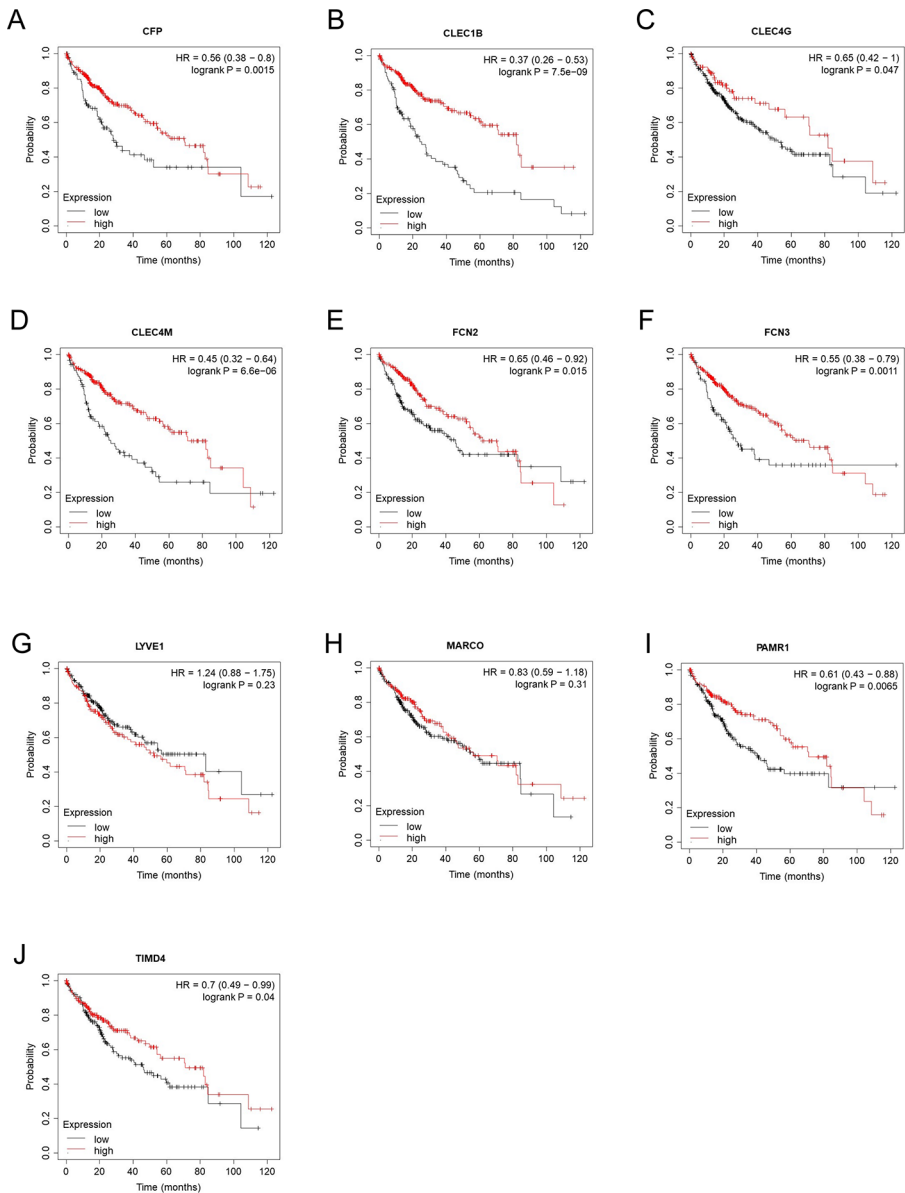


Fig. 8 Overall survival (OS) analysis of hub genes in LIHC patients. (A) CFP, (B) CLEC1B, (C) CLEC4G, (D) CLEC4M, (E) FCN2, (F) FCN3, (G) LYVE1, (H) MARCO, (I) PAMR1, (J) TIMD4. $P < 0.05$ denotes significance

high correlation with histologic grade in LIHC by analyzing TCGA LIHC dataset (Gu et al. 2020). While our study provides a more comprehensive and innovative idea from various sources of data.

Through a comprehensive bioinformatics analysis, a total of 68 significant DEGs were identified in the TCGA-LIHC and GSE54236 datasets. These genes were pri-

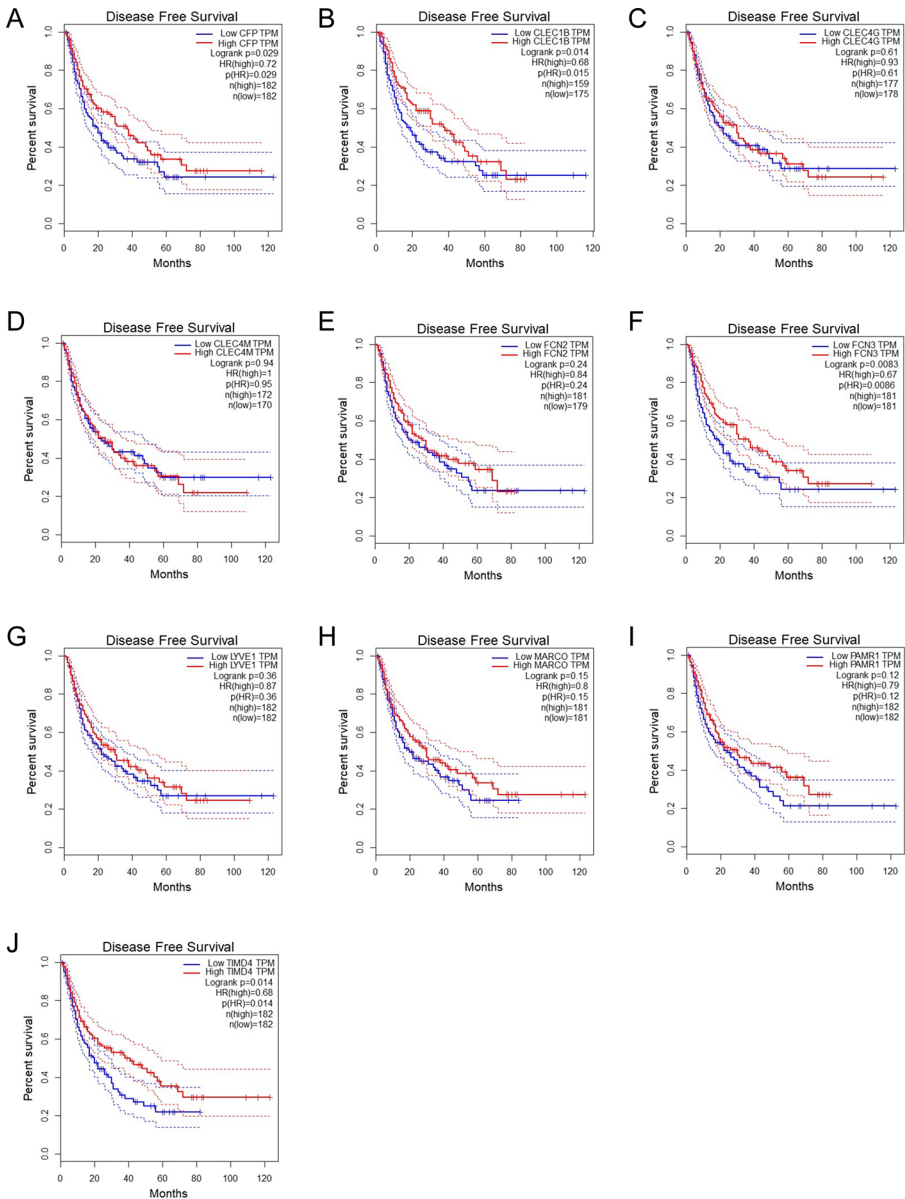


Fig. 9 Disease-free survival (DFS) analysis of hub genes in LIHC patients. (A) CFP, (B) CLEC1B, (C) CLEC4G, (D) CLEC4M, (E) FCN2, (F) FCN3, (G) LYVE1, (H) MARCO, (I) PAMR1, (J) TIMD4. $P < 0.05$ denotes significance

marily clustered in functional categories related to complement activation, collagen trimer, receptor-ligand activity, and cytokine-cytokine receptor interaction. The complement system, an ancient defense mechanism predating adaptive immunity, and its activation are known to play a crucial role in tumor-promoting inflammation

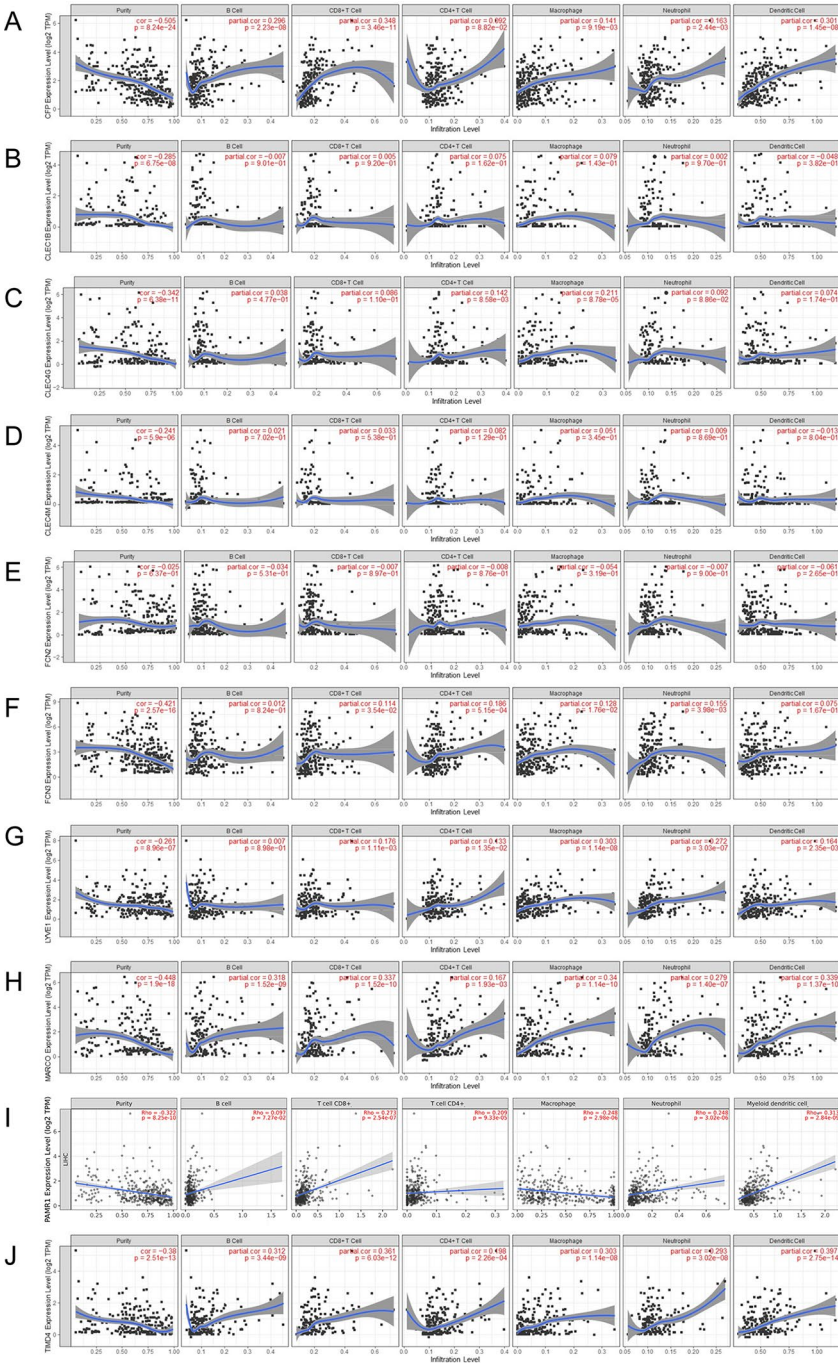


Fig. 10 Association between expression of hub genes and immune infiltration in LIHC. **(A)** CFP, **(B)** CLEC1B, **(C)** CLEC4G, **(D)** CLEC4M, **(E)** FCN2, **(F)** FCN3, **(G)** LYVE1, **(H)** MARCO, **(I)** PAMR1, **(J)** TIMD4. Each dot represents a sample from the TCGA-LIHC dataset. $P < 0.05$ denotes significance

(Afshar-Kharghan 2017; Senent et al. 2022). Chronic infection with HBV or HCV is a major risk factor for LIHC, with complement activation contributing to the progression from inflammation to tumor development. Pathways involving cytokines and ligand-receptor interactions are also implicated in the pathogenesis of early-stage LIHC (Ma et al. 2022; Molina et al. 2019). Collagen, a key component of the extracellular matrix, with collagen trimer serving as a biomarker for cancer metastasis (Xu et al. 2019). By constructing a PPI network, we identified 10 hub genes, including CFP, CLEC1B, CLEC4G, CLEC4M, FCN2, FCN3, LYVE1, MARCO, PAMR1, and TIMD4. The expression levels of these hub genes were down-regulated in LIHC compared to normal controls. Notably, the decreased expression of CFP, CLEC1B, FCN3, and TIMD4 was significantly associated with poor OS and DFS in LIHC patients. Analysis using the TIMER revealed that these hub genes were predominantly expressed in LIHC cells rather than immune cells, and did not play a role in immune regulation within the tumor microenvironment. These findings suggest that CFP, CLEC1B, FCN3, and TIMD4 may serve as potential therapeutic targets for LIHC and offer valuable insights for prognostic biomarker assessment.

In our study, CFP, CLEC1B, FCN3 and TIMD4 were down-regulated in LIHC tissues compared with normal tissues, and could be strong predictors of poor outcome in LIHC, which may indicate that these genes may function as tumor suppressor genes. Low expression levels of tumor suppressor genes are often linked to a poor prognosis. In addition to their potential involvement in carcinogenesis, there is also a possibility that their low expression contributes to the development of treatment resistance, thereby impacting patient prognosis. However, this is only a conjecture, and we plan to conduct *in vitro* experiments to further verify this next. The relationship between the four genes and cancer has been confirmed by several previous studies.

CFP (Complement Factor Properdin) is a protein coding gene, and can positively regulate the alternative complement pathway of the innate immune system yielding the elimination of pathogens, apoptotic and necrotic cells (Kemper et al. 2008, 2010). Several studies reported the tumor suppressive effect of CFP on melanoma, breast, stomach and lung cancer (Al-Rayahi et al. 2017; Block et al. 2019; Cui et al. 2021). In stomach and lung cancer, the expression level of CFP was lower than in normal tissues, and low expression level of CFP was associated with poor prognosis (Cui et al. 2021).

CLEC1B (C-type lectin domain family 1 member B) is a novel platelet-associated molecule secreted by activated platelets around tumors (Meng et al. 2021). CLECB1 has been confirmed its inhibitory effect on platelet aggregation and tumor metastasis in colon cancer (Suzuki-Inoue et al. 2007). CLEC1B has been reported to be significantly down-regulated in liver cancer (Critelli et al. 2017). Compared to paired normal tissues, the mRNA and protein levels of CLEC1B were significantly down-regulated in LIHC (Jing et al. 2023).

FCN3 (Ficolin 3) is a secreted lectin that activates the complement pathway (Endo et al. 2011). FCN3 expression was significantly down-regulated in lung cancer tissues compared with matched normal lung tissues, low expression levels of FCN3 have been described as prognostic biomarker for cancer (Jang et al. 2021). LIHC patients were divided into FCN3 high and low expression groups by immunohisto-

chemical staining. patients with high FCN3 expression had a higher overall survival rate than those with low FCN3 expression ($p=0.031$), which was consistent with our study. And high FCN3 expression in tumor tissue was independently associated with better overall survival ($p=0.042$)(Chen et al. 2023).

TIMD4 (T cell immunoglobulin and mucin domain containing 4) is a phosphatidyserine receptor that enhances the engulfment of apoptotic cells, and is involved in regulating T-cell proliferation and lymphotoxin signaling (Freeman et al. 2010). As opposed to LIHC, TIMD4 is overexpressed in lung cancer and colorectal cancer, which is correlated with poor prognosis (Dorfman et al. 2010; Tan et al. 2018).

We admit that this study has some limitations and deficiencies. First, this research mainly focuses on data mining and data analysis based on methodology, and the results have not been verified by cytology experiment. Further experiments are needed to better confirm the findings of this study. Our research also has certain limitations for the classification of different subtypes of tumors. Various subtypes of LIHC exhibit distinct clinical presentations, treatment options, and responses to treatment. LIHC can be categorized based on a range of criteria. For instance, common histological morphologies of LIHC include trabecular pattern, pseudoglandular pattern, macrotrabecular, and solid pattern types. The development of LIHC is linked to diverse chronic liver diseases or external exposures. Hepatocellular carcinoma staging can be determined by tumor characteristics or liver function impairment, such as American Joint Committee on Cancer-Tumor Node Metastasis staging (TNM), Cancer of the Liver Italian Program (CLIP) (Kulik and El-Serag 2019; Yang et al. 2019). The correlation between hub genes and LIHC of different etiologies, pathologic features, and clinical stages will be analyzed in greater depth if the analysis is based on a more comprehensive clinical profile of patients. Although our comprehensive bioinformatics analysis identifies potential diagnostic genes for LIHC, it may be biased for patients with different LIHC subtypes.

In conclusion, our employment of WGCNA has provided valuable insights into the progression from normal liver tissue to LIHC. Through this approach, we have identified two modules and ten pivotal hub genes that play critical roles in tumorigenesis. Notably, the decreased expression of CFP, CLEC1B, FCN3, and TIMD4 was significantly associated with poor OS and DFS outcomes in LIHC patients. Furthermore, TIMER for analysis revealed that these hub genes are predominantly expressed in LIHC cells rather than immune cells and do not participate in immune regulation within the tumor microenvironment. These collective findings not only highlight the potential of CFP, CLEC1B, FCN3, and TIMD4 as therapeutic targets for LIHC but also offer valuable insights for improving prognostic biomarker assessment.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10528-024-10803-8>.

Author Contributions (I) Conception and design: JWS, ZZZ, JRC (II) Collection and assembly of data: ZZZ (III) Data analysis and interpretation: JWS, JRC (IV) Manuscript writing: All authors (V) Final approval of manuscript: XPL, XLX.

Funding This work was supported in part by Natural Science Foundation of Zhejiang Province of China under Grant No. LQ23H030001 and Zhejiang Shuren University Science Foundation of China under Grant No. KXJ1720102.

Data Availability The datasets generated for this study can be found in the Publicly available datasets were analyzed in this study. This data can be found at: TCGA-LIHC: <https://portal.gdc.cancer.gov/projects/TCGA-LIHC> GSE54236: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE54236> ICGC-LIRI-JP: <https://dcc.icgc.org/projects/LIRI-JP>.

Declarations

Ethics Approval Not applicable.

Consent for Publication Not applicable.

Consent to Participate Not applicable.

Competing Interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Afshar-Kharghan V (2017) The role of the complement system in cancer. *J Clin Invest* 127(3):780–789. <https://doi.org/10.1172/jci90962>
- Al-Rayahi IA, Browning MJ, Stover C (2017) Tumour cell conditioned medium reveals greater M2 skewing of macrophages in the absence of properdin. *Immun Inflamm Dis* 5(1):68–77. <https://doi.org/10.1002/iid3.142>
- Bai KH, He SY, Shu LL, Wang WD, Lin SY, Zhang QY, Dai YJ (2020) Identification of cancer stem cell characteristics in liver hepatocellular carcinoma by WGCNA analysis of transcriptome stemness index. *Cancer Med* 9(12):4290–4298. <https://doi.org/10.1002/cam4.3047>
- Block I, Müller C, Sdogati D, Pedersen H, List M, Jaskot AM, Mollenhauer J (2019) CFP suppresses breast cancer cell growth by TES-mediated upregulation of the transcription factor DDIT3. *Oncogene* 38(23):4560–4573. <https://doi.org/10.1038/s41388-019-0739-0>
- Chakraborty E, Sarkar D (2022) Emerging therapies for Hepatocellular Carcinoma (HCC). *Cancers (Basel)* 14(11). <https://doi.org/10.3390/cancers14112798>
- Chen L, Yuan L, Wang Y, Wang G, Zhu Y, Cao R, Wang X (2017) Co-expression network analysis identified FCER1G in association with progression and prognosis in human clear cell renal cell carcinoma. *Int J Biol Sci* 13(11):1361–1372. <https://doi.org/10.7150/ijbs.21657>
- Chen Z, Xie H, Hu M, Huang T, Hu Y, Sang N, Zhao Y (2020) Recent progress in treatment of hepatocellular carcinoma. *Am J Cancer Res* 10(9):2993–3036
- Chen CC, Yu TH, Wu CC, Hung WC, Lee TL, Tang WH, Hsu CC (2023) Loss of ficolin-3 expression is associated with poor prognosis in patients with hepatocellular carcinoma. *Int J Med Sci* 20(8):1091–1096. <https://doi.org/10.7150/ijms.84729>
- Chin CH, Chen SH, Wu HH, Ho CW, Ko MT, Lin CY (2014) cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC Syst Biol* 8 Suppl 4(Suppl 4):S11. <https://doi.org/10.1186/1752-0509-8-s4-s11>

- Critelli R, Milosa F, Faillaci F, Condello R, Turola E, Marzi L, Villa E (2017) Microenvironment inflammatory infiltrate drives growth speed and outcome of hepatocellular carcinoma: a prospective clinical study. *Cell Death Dis* 8(8):e3017. <https://doi.org/10.1038/cddis.2017.395>
- Cui G, Geng L, Zhu L, Lin Z, Liu X, Miao Z, Wei F (2021) CFP is a prognostic biomarker and correlated with immune infiltrates in gastric Cancer and Lung Cancer. *J Cancer* 12(11):3378–3390. <https://doi.org/10.7150/jca.50832>
- Doncheva NT, Morris JH, Gorodkin J, Jensen LJ (2019) Cytoscape StringApp: Network Analysis and Visualization of Proteomics Data. *J Proteome Res* 18(2):623–632. <https://doi.org/10.1021/acs.jproteome.8b00702>
- Dorfman DM, Hornick JL, Shahsafaie A, Freeman GJ (2010) The phosphatidylserine receptors, T cell immunoglobulin mucin proteins 3 and 4, are markers of histiocytic sarcoma and other histiocytic and dendritic cell neoplasms. *Hum Pathol* 41(10):1486–1494. <https://doi.org/10.1016/j.humpath.2010.04.005>
- Endo Y, Matsushita M, Fujita T (2011) The role of ficolins in the lectin pathway of innate immunity. *Int J Biochem Cell Biol* 43(5):705–712. <https://doi.org/10.1016/j.biocel.2011.02.003>
- Freeman GJ, Casanovas JM, Umetsu DT, DeKruyff RH (2010) TIM genes: a family of cell surface phosphatidylserine receptors that regulate innate and adaptive immunity. *Immunol Rev* 235(1):172–189. <https://doi.org/10.1111/j.0105-2896.2010.00903.x>
- Gao XM, Zhou XH, Jia MW, Wang XZ, Liu D (2023) Identification of key genes in sepsis by WGCNA. *Prev Med* 172:107540. <https://doi.org/10.1016/j.ypmed.2023.107540>
- Gu Y, Li J, Guo D, Chen B, Liu P, Xiao Y, Liu Q (2020) Identification of 13 key genes correlated with progression and prognosis in Hepatocellular Carcinoma by Weighted Gene Co-expression Network Analysis. *Front Genet* 11:153. <https://doi.org/10.3389/fgene.2020.00153>
- Jang H, Jun Y, Kim S, Kim E, Jung Y, Park BJ, Kim J (2021) FCN3 functions as a tumor suppressor of lung adenocarcinoma through induction of endoplasmic reticulum stress. *Cell Death Dis* 12(4):407. <https://doi.org/10.1038/s41419-021-03675-y>
- Jing Q, Yuan C, Zhou C, Jin W, Wang A, Wu Y, Shao F (2023) Comprehensive analysis identifies CLEC1B as a potential prognostic biomarker in hepatocellular carcinoma. *Cancer Cell Int* 23(1):113. <https://doi.org/10.1186/s12935-023-02939-1>
- Kemper C, Mitchell LM, Zhang L, Hourcade DE (2008) The complement protein properdin binds apoptotic T cells and promotes complement activation and phagocytosis. *Proc Natl Acad Sci U S A* 105(26):9023–9028. <https://doi.org/10.1073/pnas.0801015105>
- Kemper C, Atkinson JP, Hourcade DE (2010) Properdin: emerging roles of a pattern-recognition molecule. *Annu Rev Immunol* 28:131–155. <https://doi.org/10.1146/annurev-immunol-030409-101250>
- Kulik L, El-Serag HB (2019) Epidemiology and management of Hepatocellular Carcinoma. *Gastroenterology* 156(2):477–491e471. <https://doi.org/10.1053/j.gastro.2018.08.065>
- Langfelder P, Horvath S (2008) WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9:559. <https://doi.org/10.1186/1471-2105-9-559>
- Li B, Severson E, Pignon JC, Zhao H, Li T, Novak J, Liu XS (2016) Comprehensive analyses of tumor immunity: implications for cancer immunotherapy. *Genome Biol* 17(1):174. <https://doi.org/10.1186/s13059-016-1028-7>
- Li T, Fan J, Wang B, Traugh N, Chen Q, Liu JS, Liu XS (2017) TIMER: a web server for Comprehensive Analysis of Tumor-infiltrating Immune cells. *Cancer Res* 77(21):e108–e110. <https://doi.org/10.1158/0008-5472.Can-17-0307>
- Li CY, Cai JH, Tsai JJP, Wang CCN (2020a) Identification of hub genes Associated with Development of Head and Neck Squamous Cell Carcinoma by Integrated Bioinformatics Analysis. *Front Oncol* 10:681. <https://doi.org/10.3389/fonc.2020.00681>
- Li T, Fu J, Zeng Z, Cohen D, Li J, Chen Q, Liu XS (2020b) TIMER2.0 for analysis of tumor-infiltrating immune cells. *Nucleic Acids Res* 48(W1):W509–W514. <https://doi.org/10.1093/nar/gkaa407>
- Ma L, Heinrich S, Wang L, Keggenhoff FL, Khatib S, Forgues M, Wang XW (2022) Multiregional single-cell dissection of tumor and immune cells reveals stable lock-and-key features in liver cancer. *Nat Commun* 13(1):7533. <https://doi.org/10.1038/s41467-022-35291-5>
- Meng D, Luo M, Liu B (2021) The role of CLEC-2 and its ligands in Thromboinflammation. *Front Immunol* 12:688643. <https://doi.org/10.3389/fimmu.2021.688643>
- Molina MF, Abdelnabi MN, Fabre T, Shoukry NH (2019) Type 3 cytokines in liver fibrosis and liver cancer. *Cytokine* 124:154497. <https://doi.org/10.1016/j.cyto.2018.07.028>

- Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Alizadeh AA (2015) Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods* 12(5):453–457. <https://doi.org/10.1038/nmeth.3337>
- Pei G, Chen L, Zhang W (2017) WGCNA Application to Proteomic and Metabolomic Data Analysis. *Methods Enzymol* 585:135–158. <https://doi.org/10.1016/bs.mie.2016.09.016>
- Senent Y, Tavira B, Pio R, Ajona D (2022) The complement system as a regulator of tumor-promoting activities mediated by myeloid-derived suppressor cells. *Cancer Lett* 549:215900. <https://doi.org/10.1016/j.canlet.2022.215900>
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Ideker T (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 13(11):2498–2504. <https://doi.org/10.1101/gr.1239303>
- Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F (2021) Global Cancer statistics 2020: GLOBOCAN estimates of incidence and Mortality Worldwide for 36 cancers in 185 countries. *Cancer J Clin* 71(3):209–249. <https://doi.org/10.3322/caac.21660>
- Suzuki-Inoue K, Kato Y, Inoue O, Kaneko MK, Mishima K, Yatomi Y, Ozaki Y (2007) Involvement of the snake toxin receptor CLEC-2, in podoplanin-mediated platelet activation, by cancer cells. *J Biol Chem* 282(36):25993–26001. <https://doi.org/10.1074/jbc.M702327200>
- Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Mering CV (2019) STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res* 47(D1):D607–d613. <https://doi.org/10.1093/nar/gky1131>
- Tan X, Zhang Z, Yao H, Shen L (2018) Tim-4 promotes the growth of colorectal cancer by activating angiogenesis and recruiting tumor-associated macrophages via the PI3K/AKT/mTOR signaling pathway. *Cancer Lett* 436:119–128. <https://doi.org/10.1016/j.canlet.2018.08.012>
- Tang Z, Kang B, Li C, Chen T, Zhang Z (2019) GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis. *Nucleic Acids Res* 47(W1):W556–w560. <https://doi.org/10.1093/nar/gkz430>
- The Gene Ontology Resource (2019) 20 years and still GOing strong. *Nucleic Acids Res* 47(D1):D330–d338. <https://doi.org/10.1093/nar/gky1055>
- van Ijzendoorn DGP, Szuhai K, Briaire-de Bruijn IH, Kostine M, Kuijjer ML, Bovée JVMG (2019) Machine learning analysis of gene expression data reveals novel diagnostic and prognostic biomarkers and identifies therapeutic targets for soft tissue sarcomas. *PLoS Comput Biol* 15(2):e1006826. <https://doi.org/10.1371/journal.pcbi.1006826>
- Xu S, Xu H, Wang W, Li S, Li H, Li T, Liu L (2019) The role of collagen in cancer: from bench to bedside. *J Transl Med* 17(1):309. <https://doi.org/10.1186/s12967-019-2058-1>
- Yang JD, Hainaut P, Gores GJ, Amadou A, Plymoth A, Roberts LR (2019) A global view of hepatocellular carcinoma: trends, risk, prevention and management. *Nat Rev Gastroenterol Hepatol* 16(10):589–604. <https://doi.org/10.1038/s41575-019-0186-y>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Jiawei Sun¹ · Zizhen Zhang² · Jiaru Cai¹ · Xiaoping Li¹ · Xiaoling Xu¹

✉ Xiaoping Li
li-xp@zjsru.edu.cn

✉ Xiaoling Xu
ziyao1988@zju.edu.cn

¹ Shulan International Medical College, Zhejiang Shuren University, Hangzhou 31005, China

² Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education/

Beijing), Department of Gastrointestinal Oncology, Peking University Cancer Hospital & Institute, Beijing 100142, China