

Article

Integrative Single-Cell and Bulk Transcriptomic Analysis Identifies Macrophage-Related Gene Signatures Predictive of Hepatocellular Carcinoma in Cirrhosis

Zhongyuan Zhang ¹ , Chuisheng Zeng ¹, Xuotong Yong ¹ , Wenping Zhou ¹, Yongfang Xie ^{1,*}  and Jianzhong Shu ^{2,3,*}

¹ Chongqing Key Laboratory of Big Data for Bio Intelligence, Chongqing University of Posts and Telecommunications, Chongqing 400065, China; 19823318477@163.com (Z.Z.)

² Chongqing University of Traditional Chinese Medicine, Chongqing 402760, China

³ Chongqing Traditional Chinese Medicine Hospital, Chongqing 400020, China

* Correspondence: xieyf@cqupt.edu.cn (Y.X.); shujz1688@126.com (J.S.)

Abstract

Background/Objectives: Liver cirrhosis is a major global health challenge and a key risk factor for hepatocellular carcinoma (HCC), a malignancy with high mortality due to late diagnosis. This study aimed to integrate single-cell RNA sequencing (scRNA-seq) and bulk RNA sequencing (bulk RNA-seq) data, using single-cell data to identify macrophage-associated transcriptomic changes during the progression from cirrhosis to HCC, and using bulk data to validate these findings in independent cohorts, while developing predictive models for early risk assessment. **Methods:** We integrated single-cell RNA sequencing (scRNA-seq) and bulk RNA sequencing datasets derived from liver tissues of cirrhosis and HCC patients. Single-cell data were used to identify macrophage subtypes and their dynamic transcriptional changes, while bulk data provided validation in independent cohorts. Gene expression and network analyses were performed, and candidate genes were used to construct diagnostic models with Lasso regression, Random Forest, and Extreme Gradient Boosting (XGBoost). Model performance was evaluated using receiver operating characteristic curves. **Results:** We identified eleven macrophage-associated genes, among which KLK11, MARCO, CFP, KRT19, GAS1, SOD3, and CYP2C8 were downregulated in HCC, indicating loss of tumor-suppressive and pro-apoptotic functions, while TOP2A, CENPF, MKI67, and NUPR1 were upregulated, reflecting enhanced cell cycle progression, proliferation, and M2 polarization. These are all associated with the progression from liver cirrhosis to HCC. Based on these findings, we established predictive models using Lasso, Random Forest, and XGBoost, which stratified cirrhotic patients into high- and low-risk groups according to cutoff values using liver tissue transcriptomic data. All three models demonstrated high diagnostic performance. **Conclusions:** This study highlights the critical role of macrophage-associated transcriptomic remodeling in liver disease progression. The machine learning-based predictive models offer a promising approach for early diagnosis and clinical decision-making in patients with cirrhosis.

Keywords: single-cell RNA sequencing; liver cirrhosis; hdWGCNA; machine learning; predictive model



Academic Editor: Piero Fariselli

Received: 5 August 2025

Revised: 10 October 2025

Accepted: 14 October 2025

Published: 15 October 2025

Citation: Zhang, Z.; Zeng, C.; Yong, X.; Zhou, W.; Xie, Y.; Shu, J. Integrative Single-Cell and Bulk Transcriptomic Analysis Identifies Macrophage-Related Gene Signatures Predictive of Hepatocellular Carcinoma in Cirrhosis. *Genes* **2025**, *16*, 1213. <https://doi.org/10.3390/genes16101213>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

HCC is a classic example of an inflammation-associated cancer, with over 90% of cases developing in the context of chronic liver injury [1]. Risk factors, such as cirrhosis, trigger persistent inflammatory responses characterized by infiltration of macrophages and immature myeloid cells, as well as dysregulated cytokine production [2–5]. These unresolved responses drive progressive fibrosis, cirrhosis, and ultimately, HCC [6–8].

Hepatic macrophages play a central role in liver homeostasis and disease, participating in diverse processes including injury exacerbation, anti-inflammation, tissue repair, regulation of fibrosis (both pro- and anti-fibrotic roles), and tumor progression (both pro- and anti-tumorigenic functions) [9]. Tumor-associated macrophages (TAMs), one of the most abundant immune cell populations within the tumor microenvironment, significantly contribute to HCC progression through inflammation-mediated mechanisms. Based on their origin, TAMs can be classified as either tissue-resident macrophages, such as Kupffer cells (KCs), or monocyte-derived macrophages (MoMφs) [10,11]. Under physiological conditions, KCs are the most common. However, during HCC development, both KCs and MoMφs undergo phenotypic changes driven by tumor-promoting signals, ultimately contributing to immune remodeling and tumor growth [12].

With the rapid advancement of high-throughput sequencing technologies and artificial intelligence, computational methods have emerged as powerful tools for dissecting the molecular mechanisms driving the progression from cirrhosis to HCC [13,14]. In recent years, machine learning (ML) has gained momentum in oncology, offering solutions to challenges posed by cancer's complexity and data dimensionality [15]. These techniques facilitate early diagnosis, risk stratification, and personalized treatment planning, thereby improving clinical outcomes [16,17].

Among commonly used ML algorithms, random forest and extreme gradient boosting (XGBoost) are ensemble-based approaches that perform well on structured biological data [18–20]. Lasso regression is another effective method for performing variable selection and regularization and is especially suited for high-dimensional transcriptomic datasets.

In this study, we aim to explore the dynamic changes that occur in hepatic macrophages during the progression from cirrhosis to HCC by integrating single-cell transcriptomic and bulk RNA sequencing data. We further utilize ML algorithms—Lasso regression, random forest, and XGBoost—to construct a predictive model based on macrophage-associated gene expression features. Our objective is to develop a robust diagnostic tool with which to identify cirrhotic patients who are at high risk of developing HCC, thereby facilitating early intervention and improving clinical outcomes.

2. Materials and Methods

2.1. Data Collection

scRNA-seq datasets were obtained from GEO, including GSE136103, GSE186343, GSE242889, and GSE202642. These datasets comprise healthy, cirrhotic, and HCC liver tissues, all generated using the 10× Genomics Chromium platform with sequencing on Illumina HiSeq or NovaSeq instruments. For downstream analysis, we used the processed gene–cell count matrices provided by GEO, while raw FASTQ files are available in the SRA.

Bulk RNA-seq datasets, GSE25097 and GSE14323, were likewise retrieved from GEO. GSE25097 (Affymetrix Rosetta/Merck Human RSTA 1.0 array, GPL10687) contains 268 HCC, 243 adjacent non-tumor, 40 cirrhosis, and 6 normal samples, while GSE14323 (Affymetrix U133A/U133A 2.0 arrays, GPL96/GPL571) comprises 124 samples including HCV-related HCC, cirrhosis, and normal liver tissues. In both datasets, normalized expression matrices provided by GEO were used for analysis, with raw CEL files also accessible.

2.2. Data Preprocessing and Analysis for Single-Cell RNA Sequencing

In this study, scRNA-seq data were processed using the R package Seurat (version 5.1.0) (<https://www.github.com/satijalab/seurat>) (accessed on 30 September 2024) [21]. Seurat objects were initially created in bulk using the CreateSeuratObject function. Quality control procedures were then applied to filter out low-quality cells based on metrics such as gene count, unique molecular identifiers (UMIs), and mitochondrial gene expression.

Following quality control, data normalization and identification of highly variable genes were performed using the SCTransform method according to Seurat's official guidelines [22]. To further mitigate batch effects across multiple datasets, we applied Reciprocal PCA (RPCA)-based integration implemented in Seurat's IntegrateLayers function, combined with SCT normalization. This approach identifies shared anchors across datasets, aligns them into a common expression space, and effectively corrects technical variability while preserving true biological signals, thereby enhancing the consistency and reproducibility of downstream analyses.

Next, principal component analysis (PCA) was conducted using the RunPCA function for linear dimensionality reduction. Based on the inflection point observed in the ElbowPlot, the top 20 principal components were selected for subsequent analyses. Dataset integration was performed following Seurat's standard workflow to correct for inter-sample batch effects.

After integration, cell clustering was carried out using the FindClusters function, with a resolution parameter set to 0.5. Clustering was based on the construction of a shared nearest neighbor (SNN) graph. To visualize the spatial distribution of cell clusters in two dimensions, Uniform Manifold Approximation and Projection (UMAP), a nonlinear dimensionality reduction algorithm, was applied [23].

Differentially expressed genes (DEGs) for each cluster were identified using the FindAllMarkers function. Cell type annotation was performed by referencing the CellMarker 2.0 database (<http://117.50.127.228/CellMarker>) (accessed on 30 October 2024) [24] and the Cell Taxonomy database (<https://ngdc.cncb.ac.cn/celltaxonomy/>) (accessed on 30 October 2024) [25], and by examining the expression profiles of canonical marker genes for specific cell populations.

2.3. Differential Expression Analysis and Gene Set Enrichment Analysis

Differential expression analysis was conducted to identify genes with significantly altered expression between different disease stages. Specifically, macrophage-labeled cells were extracted, and the FindMarkers function in Seurat was applied to identify differentially expressed genes (DEGs) between cirrhosis and HCC.

To explore the biological roles of macrophage-related DEGs, GSEA was performed. Functional enrichment was carried out using the clusterProfiler package in R [26,27]. Gene Ontology (GO) annotations were used to categorize enriched functions into three domains: biological process (BP), molecular function (MF), and cellular component (CC).

2.4. Cell Interaction Analysis

To elucidate the complex landscape of cell–cell communication within the tumor microenvironment, we utilized the CellChat package, which enables quantitative analysis of interaction networks based on ligand–receptor–cofactor databases derived from scRNA-seq data [28]. Following the standard CellChat workflow, we systematically quantified and characterized the dynamics, strength, and signaling pathways of intercellular communication. Particular emphasis was placed on comparing macrophage-mediated interactions between cirrhosis and HCC to better understand their functional heterogeneity under distinct pathological conditions [29].

2.5. High-Dimensional Weighted Gene Co-Expression Network Analysis (hdWGCNA)

To identify genes closely associated with disease-specific macrophage subclusters, we applied high-dimensional weighted gene co-expression network analysis (hdWGCNA) to the macrophage scRNA-seq data. Initially, macrophage-labeled cells were extracted, re-normalized, and highly variable genes were re-identified. Dimensionality reduction and nonlinear dimensionality reduction were then performed to characterize cell subclusters.

Next, hdWGCNA was conducted on the expression profiles of these subclusters. For metacell construction, the parameter k was set to 12, and a soft threshold power of 8 was chosen based on data characteristics to build the co-expression network. This analysis identified 13 highly correlated gene modules. For each module, the top 20 genes with the highest module membership scores were defined as hub genes.

To pinpoint modules related to macrophage-specific transcriptional changes, differentially expressed genes (DEGs) identified in macrophages were intersected with genes in each module. Modules in which more than 10% of their hub genes overlapped with macrophage DEGs were considered differentially regulated modules.

2.6. Transcriptomic Data Analysis and Gene Selection

In this study, transcriptomic data from the GSE25097 dataset were processed using the R package *affy* to import robust multi-array average (RMA) expression data and convert them into a standardized gene expression matrix. Differential expression analysis was then performed using the *limma* package to identify significantly dysregulated genes. To enhance biological relevance, the differentially expressed genes identified from the bulk RNA-seq data were intersected with both the macrophage-specific differentially expressed genes and the module-related genes derived from the single-cell transcriptomic analysis. The overlapping genes were selected as candidate biomarkers for downstream model construction. To further refine the gene set and develop a robust diagnostic model, least absolute shrinkage and selection operator (LASSO) regression analysis was applied to the GSE25097 dataset using the R package *glmnet*. The penalty parameter was optimized by cross-validation, and a lambda value of 0.0004814197 was selected to construct the final LASSO regression model comprising the most informative genes.

2.7. Model Construction and Performance Evaluation

To develop diagnostic models for cirrhosis and HCC, three machine learning approaches were employed: random forest, extreme gradient boosting (XGBoost), and least absolute shrinkage and selection operator (LASSO) regression. The random forest model was constructed using gene expression profiles from the GSE25097 dataset as the training set and those from the GSE14323 dataset as the validation set. A total of 500 decision trees were used during model training, and feature importance and proximity metrics were enabled. The receiver operating characteristic (ROC) curve was plotted, and the area under the curve (AUC) was calculated to assess model performance. Subsequently, the training and validation sets were combined to form an integrated dataset, during which platform-related batch effects were adjusted using the *removeBatchEffect* function in the *limma* package, thereby ensuring data comparability before further evaluation of model robustness and generalizability.

For the XGBoost model, a binary logistic regression objective was used. Key parameters included a learning rate of 0.3, a maximum tree depth of 6, a log loss evaluation metric, and 100 training rounds. Following prediction on the validation dataset, the ROC curve and corresponding AUC were computed to determine diagnostic accuracy. The LASSO regression model, trained on the same datasets, was evaluated using an identical strategy. All three models were implemented for the classification of cirrhosis and HCC. For

each model, the optimal cutoff value was determined by selecting the threshold that yielded the best diagnostic performance based on ROC curve analysis, considering the highest sensitivity and specificity. Patients were subsequently stratified into high-risk and low-risk groups according to this cutoff. The three models were then jointly applied to establish predictive frameworks for early diagnosis.

3. Results

3.1. Single-Cell Transcriptomic Landscape of Hepatitis, Cirrhosis, Hepatocellular Carcinoma, and Adjacent Non-Tumorous Liver Tissue

We processed and integrated single-cell RNA sequencing (scRNA-seq) data from four publicly available datasets (GSE136103, GSE186343, GSE242889, and GSE202642), yielding a total of 124,148 cells. Among these, 25,337 cells were derived from liver cirrhosis samples, 20,928 from hepatitis samples, and 77,883 from HCC samples. After standard normalization, clustering, integration, and nonlinear dimensionality reduction, we identified 29 distinct cellular clusters (Figure 1A). By referencing public databases and evaluating the expression of canonical marker genes, these clusters were preliminarily annotated into 12 major cell types: T cells (CD3D, IL7R, CD69, NKG7, GZMA, CD2, CD8A, CD3G, CD4, TRAC, FOXP3, TRDC), natural killer (NK) cells (NKG7, GZMB, GNLY, PRF1, KLRF1), macrophages (CD14, CD68, CD163, CTSD, ADAP2, C1QC), dendritic cells (CD1C, CSTA, FCN1, LYZ), endothelial cells (VWF, EMCN, CD34, VWA1, PODXL), B cells (CD79A, MS4A1, MZB1, IGKC, IGHM), neutrophils (CSF3R, S100A8, S100A9, MND1), hepatocytes (ALB, APOC2, APOE, ALDOB, APOC1), hepatic stellate cells (MYH11, NOTCH3, RGS5), cholangiocytes (CLDN4, CLDN10, CXCL6), plasma cells (MKI67, UBE2C), and monocytes (CD14) [24,25]. The final cell count for each type was as follows: T cells (38,460), dendritic cells (10,464), NK cells (18,000), endothelial cells (9756), macrophages (14,745), hepatocytes (14,901), B cells (6497), neutrophils (3786), hepatic stellate cells (2461), cholangiocytes (1976), plasma cells (1677), and monocytes (1425) (Figure 1B,C). These quantifications provided a foundational dataset for subsequent analyses, ensuring the accuracy and reliability of downstream investigations (Figure S1, Table S1). We further compared the relative proportions of these cell types across different disease stages. (Figure 1D). The results revealed significant compositional changes in many cell populations during disease progression. Notably, the proportion of macrophages increased progressively from hepatitis to cirrhosis and ultimately to HCC, with statistically significant differences ($p < 0.05$) (Figure 1E). Considering that cirrhosis is a primary precursor of HCC and that changes in cellular and microenvironmental composition between these two stages hold critical clinical and biological implications, we focused subsequent analyses specifically on cirrhosis and HCC. In particular, we aimed to elucidate the dynamic states and functional diversity of macrophages during the transition from cirrhosis to HCC.

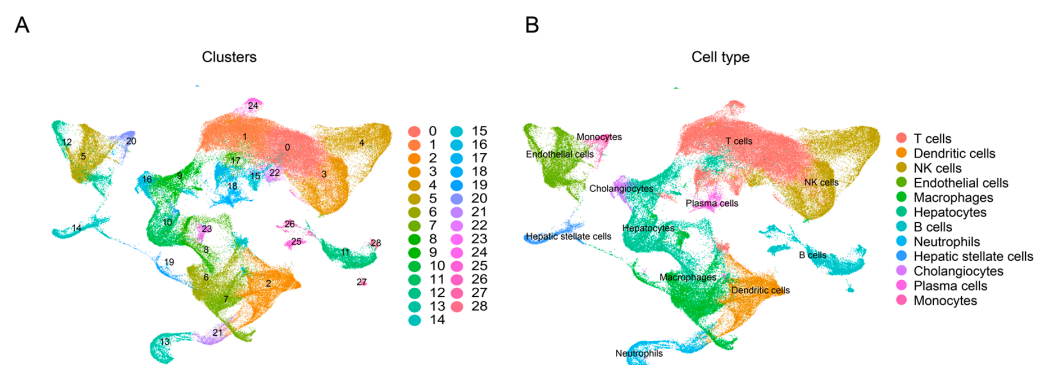


Figure 1. Cont.

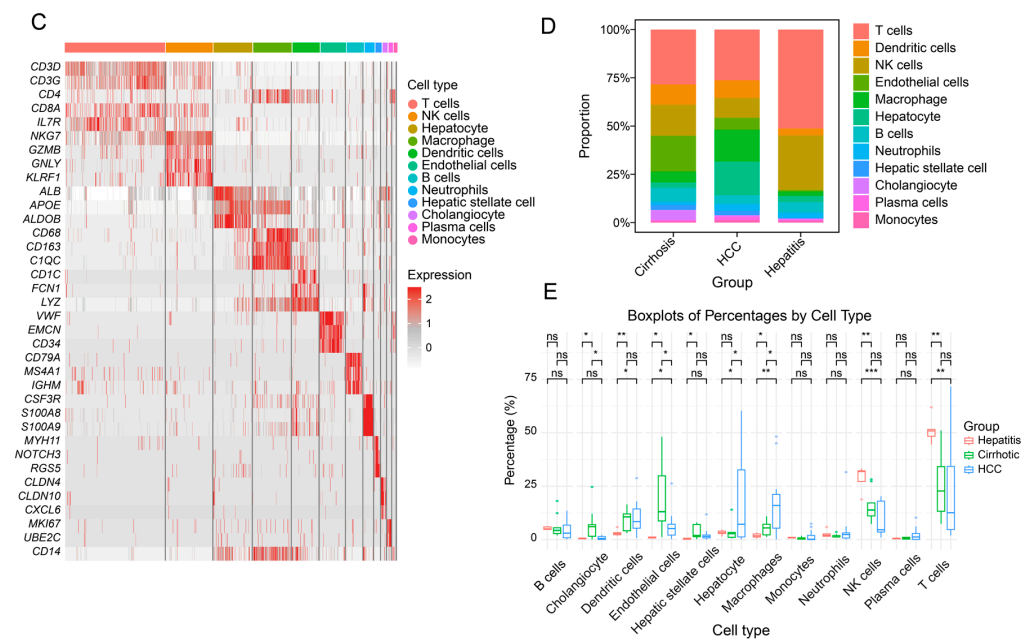


Figure 1. Single-cell transcriptomic landscape and cellular composition of liver disease samples. (A) UMAP plot showing the distribution of samples following the removal of batch effects. (B) UMAP plot displaying different cell types in liver disease. (C) Heatmap showing the expression of representative marker genes across 12 major cell types. (D) Stacked bar plot comparing the relative proportions of each cell type across different disease stages (hepatitis, cirrhosis, and HCC). (E) Boxplot showing the proportion differences in multiple cell types across different liver disease states, with macrophages exhibiting a progressive. (Statistical significance is indicated as follows: ns, not significant; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$).

3.2. Macrophage Profiling in Hepatic Diseases: Gene Expression, Pathways, and Polarization Dynamics

First, we extracted all cells annotated as macrophages from the integrated single-cell dataset (Figure 2A). Subsequently, we identified a total of 6408 DEGs in macrophages between the stages of cirrhosis and HCC. After applying stringent filtering criteria ($p_{\text{val_adj}} < 0.01$; $\text{pct.1} > 0.25$), 4144 high-confidence DEGs were retained for downstream analyses (Figure 2B). These were categorized into upregulated and downregulated genes in cirrhosis, and each group was subjected to gene enrichment analysis to elucidate biological pathways associated with their differential expression. The pathways enriched in cirrhosis-associated macrophages revealed prominent immune activation and inflammatory responses, including leukocyte-mediated immunity, lymphocyte activation, adaptive immune pathways, chemotaxis, and immunoregulatory signaling. These findings highlight the pivotal role of macrophages in promoting liver fibrosis and modulating the immune response: by enhancing antigen presentation, migration, and chemotactic activity, macrophages facilitate the recruitment and activation of inflammatory cells to damaged hepatic tissues, thereby sustaining and exacerbating chronic inflammation. Furthermore, their active involvement in organization and remodeling of the extracellular matrix contributes to the progression of fibrosis. The upregulation of protein synthesis-related pathways reflects the highly activated state of these macrophages, marked by sustained secretion of pro-inflammatory cytokines and immune modulators (Figure 2C).

In contrast, tumor-associated macrophages in HCC exhibited notable metabolic reprogramming, characterized by increased expression in pathways related to xenobiotic metabolism, lipid metabolism, and amino acid metabolism. This metabolic shift reflects the role of macrophages in adapting to the tumor microenvironment, regulating immune

responses, and supporting tumor progression. Such differences underscore the heterogeneity of macrophages between the stages of cirrhosis and HCC (Figure 2D).

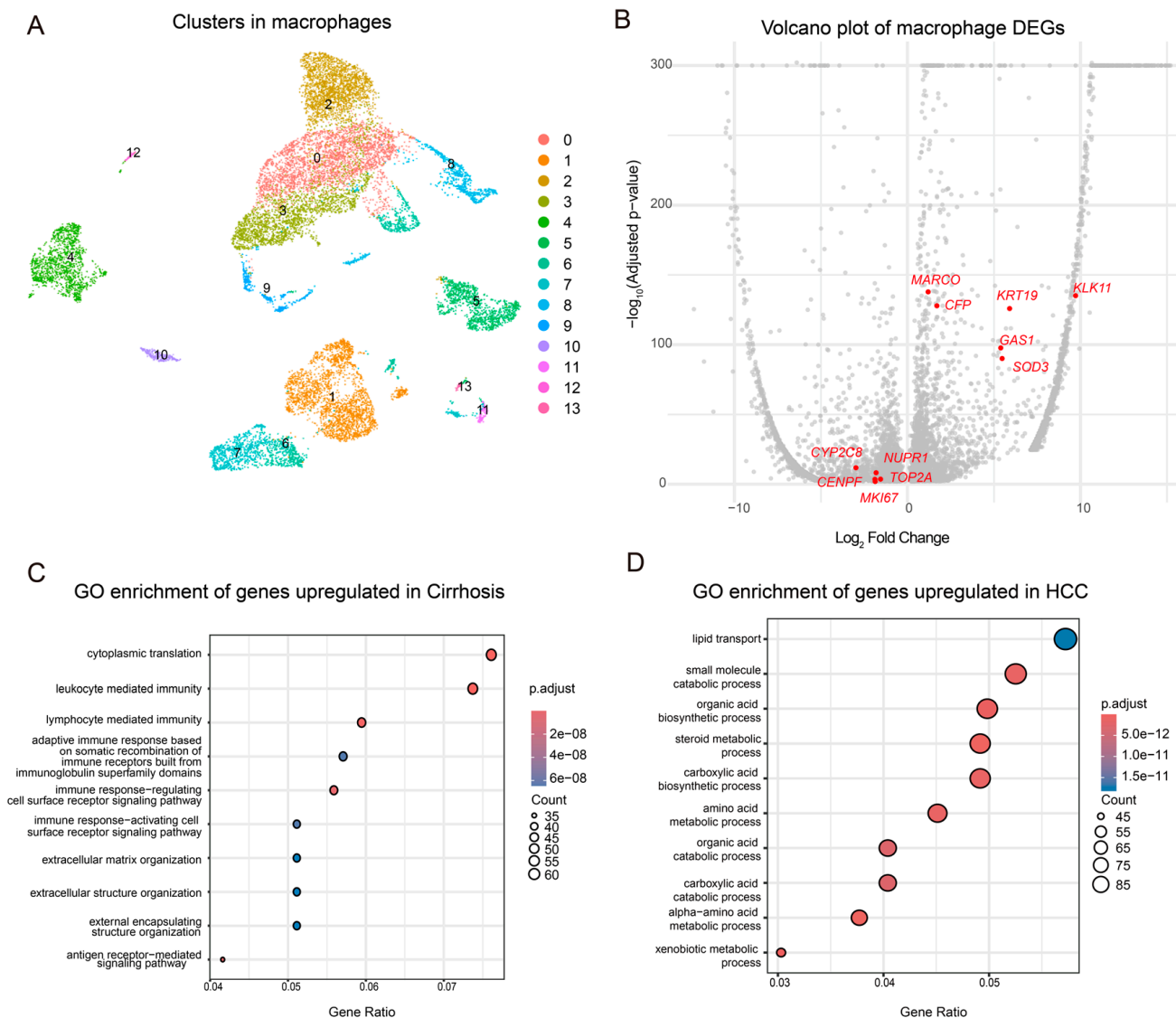


Figure 2. Differential gene enrichment and macrophage polarization dynamics during the transition from cirrhosis to HCC. **(A)** UMAP plot showing reclustered macrophages extracted from all cell populations. Each color represents a distinct macrophage subcluster identified by unsupervised clustering. **(B)** Volcano plot of DEGs in macrophages between cirrhosis and HCC, Eleven key DEGs with potential diagnostic or functional relevance (MARCO, KLK11, CFP, KRT19, GAS1, SOD3, CYP2C8, NUPR1, TOP2A, CENPF, MKI67) are highlighted. **(C)** Gene ontology and pathway enrichment analysis of DEGs upregulated in cirrhosis-associated macrophages. **(D)** Gene ontology and pathway enrichment analysis of DEGs upregulated in tumor-associated macrophages.

Additionally, we performed a polarization analysis of hepatic macrophages at the single-cell level. By scoring the expression of gene signatures associated with M1 (pro-inflammatory) and M2 (anti-inflammatory) phenotypes, we observed a gradual decline in M1 scores and a significant increase in M2 scores from cirrhosis to HCC. This dynamic shift suggests that, during the progression of chronic liver disease, macrophages transition from an immune-activated state focused on pathogen defense and inflammation toward an immunosuppressive phenotype involved in tissue remodeling and repair. Notably, macrophages in cirrhosis still maintain relatively high

M1 activity, which is consistent with their role in sustaining inflammation and promoting fibrosis. However, in the HCC stage, M2-polarized macrophages become dominant, potentially facilitating angiogenesis, enhancing immune evasion, and promoting tumor-supportive remodeling of the microenvironment (Figure S2).

3.3. Communication Network Differences Between Cirrhosis and HCC

To investigate the differences in intercellular communication between cirrhosis and HCC, we compared the overall cell–cell interaction networks in both conditions. We observed that despite the total number of interactions in HCC samples being significantly higher than in cirrhosis, the average strength of communication was relatively weaker. Notably, both the number and strength of communications involving macrophages and neutrophils were markedly higher in cirrhosis compared to HCC (Figure 3A,B). These results suggest that while HCC tissues harbor a more diverse array of cell types and broader intercellular connectivity, the signaling strength is weaker, reflecting a more immunosuppressive and functionally inert tumor microenvironment. In contrast, macrophages and neutrophils in cirrhotic tissue exhibit a more pro-inflammatory and active state, contributing robustly to immune surveillance and tissue repair processes.

Further analysis of macrophage-specific signaling revealed a consistently high incoming signaling strength in both cirrhosis and HCC, underscoring the central regulatory role of macrophages, which continuously receive external signals to modulate their functional states. Interestingly, the communication between macrophages and neutrophils was notably stronger during cirrhosis, indicating close interactions between these cells during active inflammation (Figure 3C,D).

We then analyzed ligand–receptor interactions to identify communication pathways that differ between cirrhosis and HCC. In cirrhotic tissues, several inflammatory and chemotactic ligand–receptor pairs, such as TNF–TNFRSF1A/B and MIF–CD74+CXCR4, were significantly more active than in HCC. Similarly, chemokine pathways including CCL5–CCR1 and CXCL12–CXCR4 were highly enriched. Multiple MHC class II–CD4 pairs (e.g., HLA-DRA–CD4, HLA-DRB1–CD4) also showed elevated expressions in cirrhosis. These findings indicate that cirrhosis is characterized by heightened inflammation-associated communication, with elevated expressions of pro-inflammatory ligands, chemokines, and antigen presentation molecules compared to HCC.

Additionally, ligand–receptor pairs such as SPP1–ITGAV, VTN–ITGAV, and HLA-related interactions were more active in cirrhosis, suggesting their potential roles in macrophage polarization, immune evasion, and tumor cell adhesion or migration. These interactions highlight the distinct functional states of macrophages in cirrhosis versus HCC (Figure 3E).

A focused analysis on the communication between macrophages and neutrophils revealed that macrophage-derived signaling to neutrophils was significantly enhanced in cirrhosis. This included pro-inflammatory pathways such as TNF–TNFRSF1A/B, MIF–CD74+CD44, and several MHC class II–CD4 interactions, suggesting that macrophages play an active antigen-presenting and inflammatory role during cirrhosis. Conversely, in HCC, these macrophage-to-neutrophil interactions were markedly reduced, and a subset of immunoregulatory signals—such as TGFB1–TGFB1/2 and PTPRC–MRC1—mediated from neutrophils to macrophages became further restricted. This shift indicates the functional remodeling of macrophage–neutrophil communication during the transition from cirrhosis to HCC, reflecting the progression from an immune-activated state to one of immune suppression and evasion (Figure 3F).

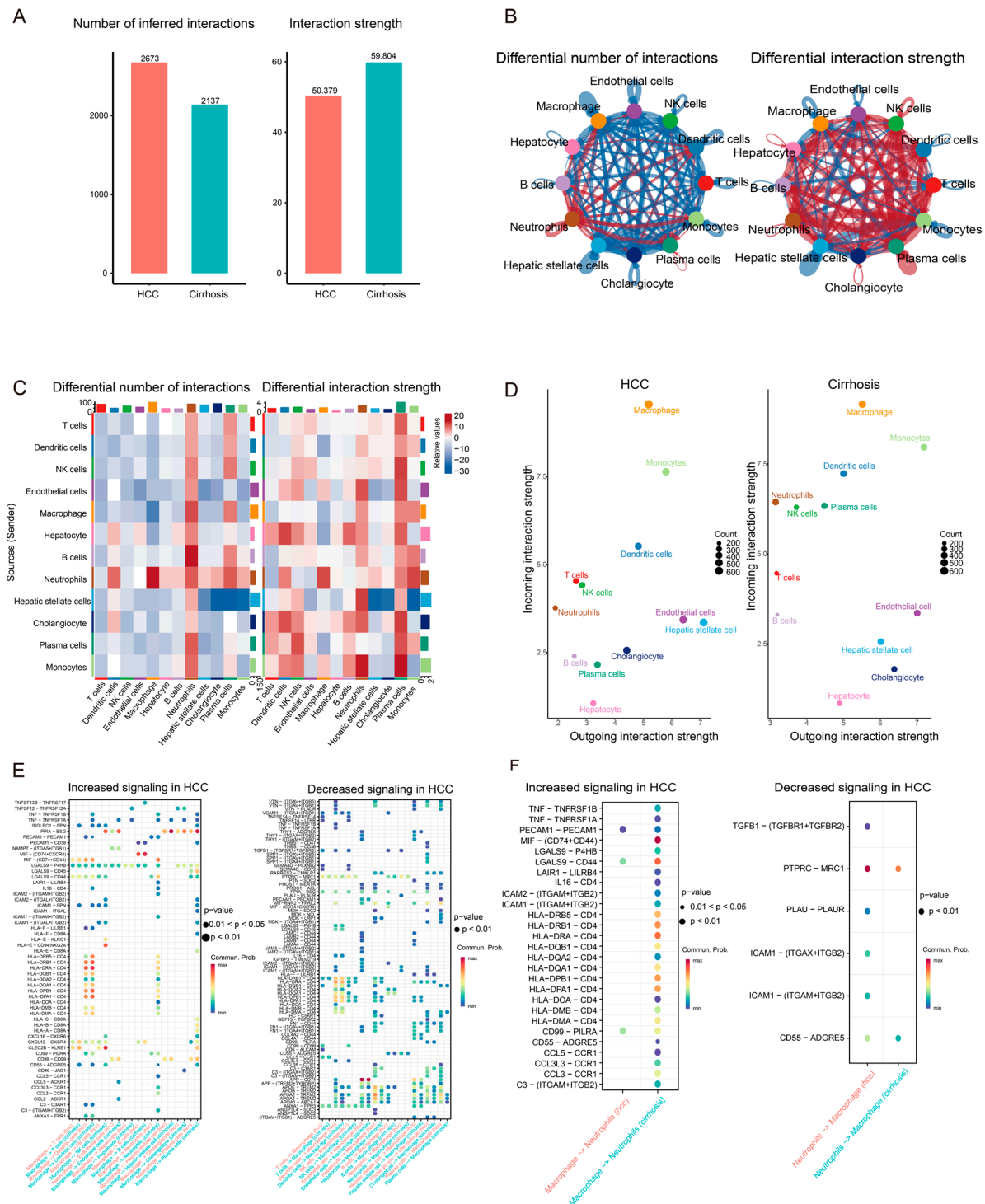


Figure 3. Intercellular communication landscape in cirrhosis and hepatocellular carcinoma (A) Overall comparison of the number and strength of cell–cell interactions between cirrhosis and HCC. (B) Differential interaction networks highlighting macrophage- and neutrophil-related communications; red edges indicate stronger interactions in cirrhosis, whereas blue edges indicate stronger interactions in HCC. (C) Cell type–specific differences in interaction number and strength; red represents higher values in cirrhosis, blue represents higher values in HCC. (D) Comparison of signaling strengths of major cell types as signal senders (outgoing) and receivers (incoming) in cirrhosis and HCC. (E) Ligand–receptor signaling pathways, with inflammation- and chemotaxis-related pathways enriched in cirrhosis, and immunosuppressive or metabolic pathways predominant in HCC. (F) Differential communication between macrophages and neutrophils in cirrhosis and HCC.

3.4. Gene Selection by High-Dimensional Weighted Gene Co-Expression Network Analysis (hdWGCNA) and Gene Selection by Bulk RNA Sequencing and Lasso Regression

To identify gene modules associated with cirrhosis and HCC, we extracted all cells annotated as macrophages from the single-cell RNA sequencing data and performed hdWGCNA following appropriate preprocessing. After constructing meta cells based on the macrophage expression profiles from cirrhotic and HCC tissues, we scanned across a range of soft-thresholding powers, selecting a power of eight as optimal (Figure 4A). Using this threshold, a gene co-expression network was constructed, and hierarchical clustering analysis was performed, resulting in the identification of 13 distinct gene modules (Figure 4B).

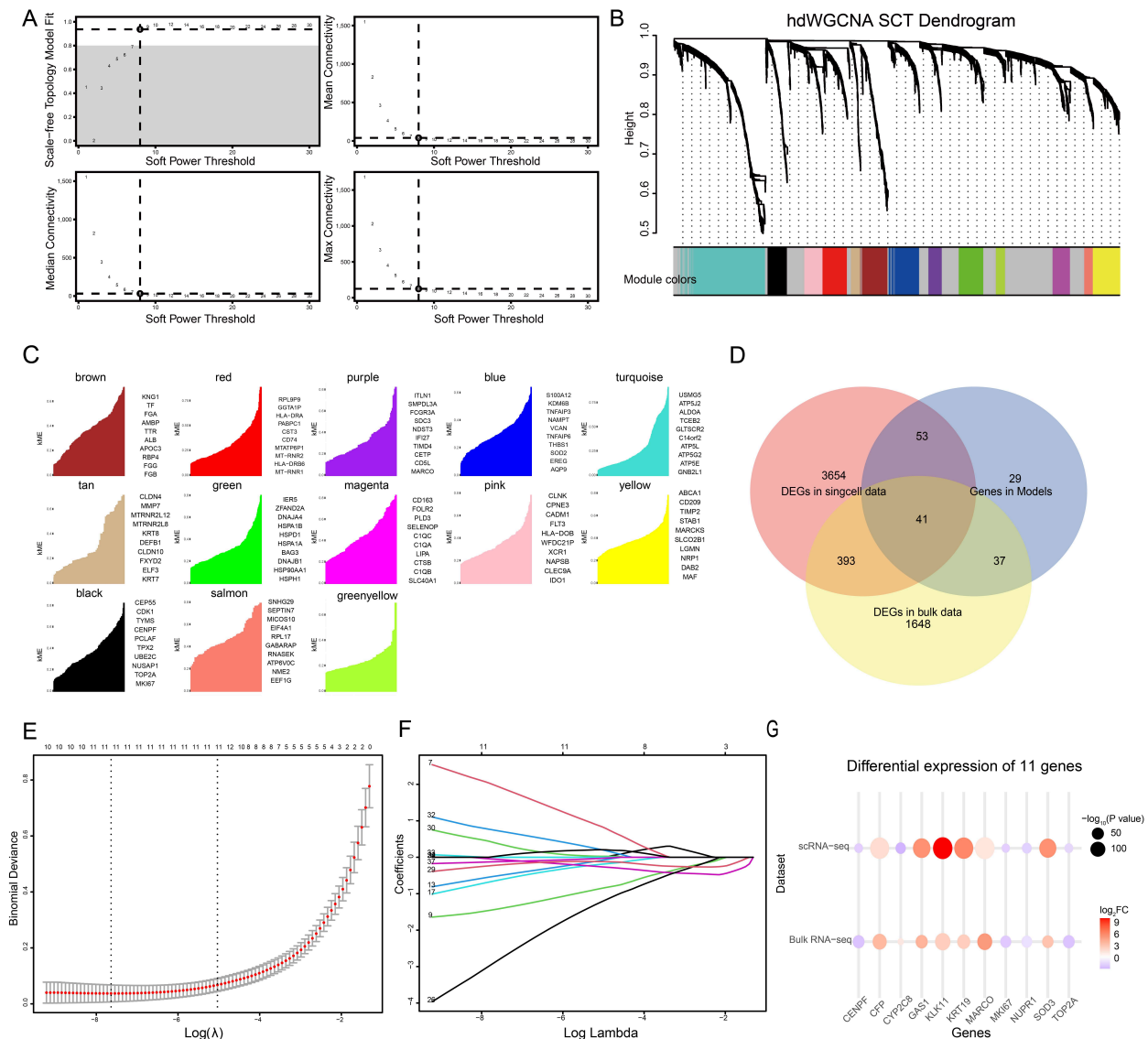


Figure 4. Identification of macrophage-associated gene modules and key diagnostic genes using hdWGCNA and Lasso regression. (A) Soft-thresholding power analysis for network topology, with power 8 selected as optimal. (B) Hierarchical clustering dendrogram of macrophage gene co-expression network, identifying 13 distinct gene modules. (C) Module membership plots showing top hub genes with highest intramodular connectivity. (D) Venn diagram showing the overlap among bulk RNA-seq DEGs, macrophage scRNA-seq DEGs, and hdWGCNA module genes, yielding 41 potential key genes. (E) Ten-fold cross-validation plot for Lasso regression to determine the optimal lambda value. (F) Lasso coefficient profiles and 11 genes selected for the final diagnostic model. (G) Comparative differential expression of 11 key genes between cirrhosis and HCC in bulk RNA-seq and single-cell RNA-seq datasets.

We quantified gene connectivity within each non-gray module to evaluate the centrality of genes and identified the top 20 hub genes per module. These hub genes represent the most interconnected members of their respective modules and are likely to play key roles in relevant biological processes (Figure 4C). To further narrow down functional modules associated with disease progression, we compared the differentially expressed genes (DEGs) between cirrhotic and HCC macrophages with the identified module-specific hub genes. Modules with a notable overlap were selected, and their corresponding genes were used for downstream analysis.

To identify candidate genes with high diagnostic value from the module-associated genes, we first analyzed the bulk RNA sequencing dataset GSE25097. Differential expression analysis was performed, and 2119 significantly differentially expressed genes (DEGs) were identified based on an adjusted p -value < 0.01 and an absolute $\log_2$ fold change ($|\log_2\text{FC}| > 1$). We then intersected these DEGs with both the macrophage-related DEGs identified from single-cell RNA sequencing and the module genes derived from hdWGCNA. This resulted in a set of 41 overlapping and potentially key genes (Figure 4D).

Subsequently, we used the GSE25097 dataset as the training cohort to construct a least absolute shrinkage and selection operator (Lasso) regression model. Ten-fold cross-validation was performed using the R package *glmnet* to determine the optimal lambda value, which was identified as 0.0004814197 (Figure 4E). The final model included 11 selected genes: CENPF, CFP, CYP2C8, GAS1, KLK11, KRT19, MARCO, MKI67, NUPR1, SOD3, and TOP2A (Figure 4F). The differential expression patterns of these genes in the training cohort were largely consistent with those observed in the single-cell RNA sequencing dataset (Figure 4G), with the exception of CYP2C8, which exhibited an opposite expression trend between the two datasets. Specifically, KLK11, MARCO, CFP, KRT19, GAS1, SOD3, and CYP2C8 were downregulated in HCC. These genes are functionally associated with tumor suppression, the induction of apoptosis, and the classification of HCC molecular subtypes; their reduced expression may contribute to unfavorable prognosis in HCC patients [30–34]. In contrast, TOP2A, CENPF, MKI67, and NUPR1 were upregulated in HCC. These genes are primarily involved in cell cycle regulation, particularly G2/M phase transition during mitosis, as well as cell proliferation and macrophage M2 polarization; thus, their overexpression is also strongly associated with poor clinical outcomes [35–38]. Therefore, it is biologically plausible and clinically reasonable to consider these genes as potential diagnostic biomarkers for HCC.

3.5. Model Construction and Performance Evaluation

In this study, we selected 11 key genes to construct three diagnostic models: a least absolute shrinkage and selection operator (Lasso) regression model, a random forest model, and an extreme gradient boosting (XGBoost) model. The training dataset was derived from GSE25097, which included 308 samples (40 cirrhosis and 268 hepatocellular carcinoma samples). The independent validation dataset was obtained from GSE14323, containing 96 samples (41 cirrhosis and 55 HCC samples). These two datasets were subsequently merged to form a unified cohort for further evaluation.

Model performance was assessed using the area under the receiver operating characteristic curve (AUC). The Lasso regression model demonstrated good discriminatory power on the independent test set, with an AUC of 0.8461 (Figure 5A). After integrating the training and test datasets, the AUC increased to 0.9634 (Figure 5B), indicating excellent predictive performance.

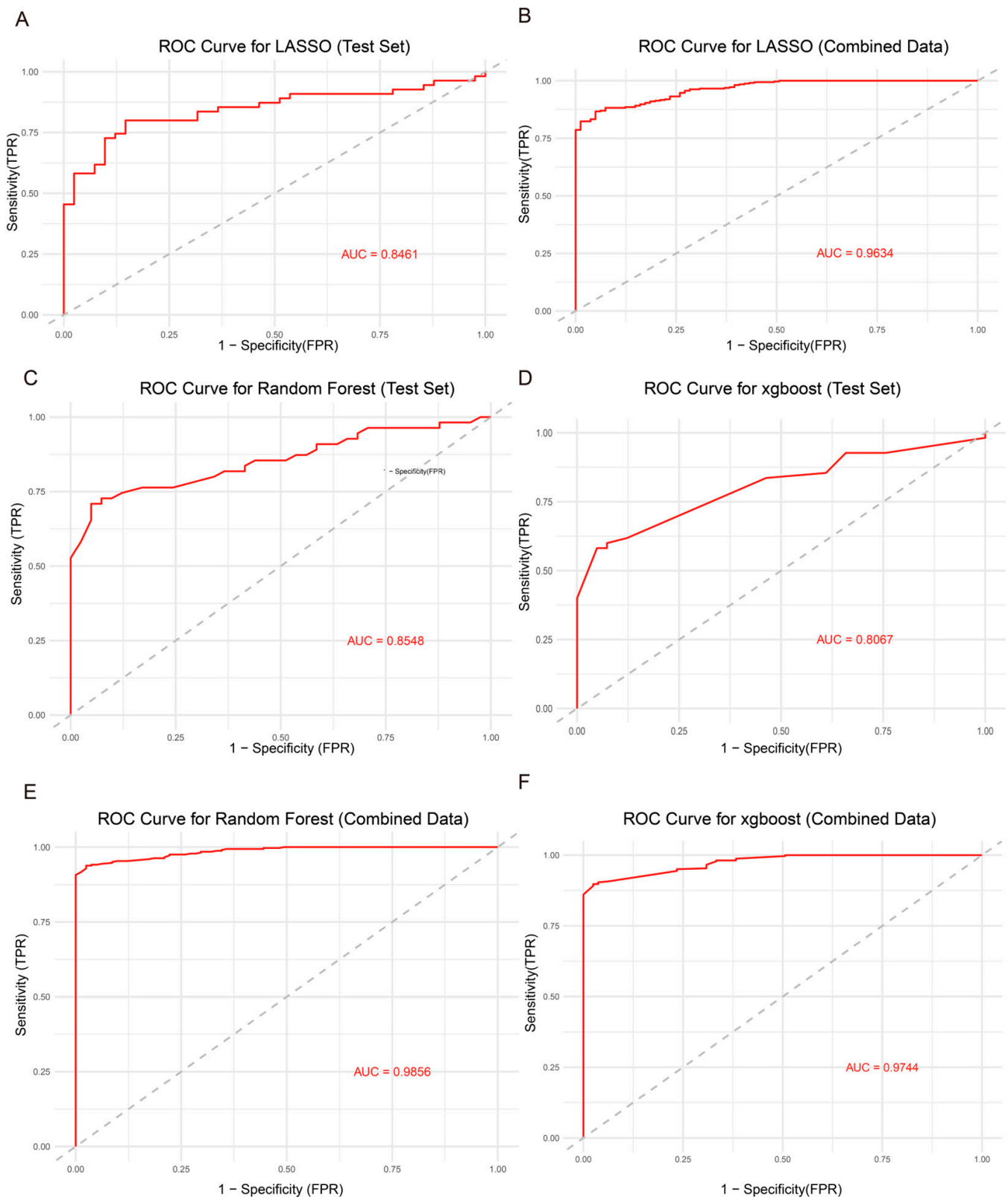


Figure 5. Diagnostic performance of Lasso, random forest, and XGBoost models in distinguishing cirrhosis from HCC. (A) ROC curve of the Lasso regression model on the independent validation dataset. (B) ROC curve of the Lasso model on the combined dataset of training and validation samples. (C) ROC curve of the random forest model on the independent validation dataset. (D) ROC curve of the XGBoost model on the independent validation dataset. (E) ROC curve of the random forest model on the combined dataset. (F) ROC curve of the XGBoost model on the combined dataset.

The random forest and XGBoost models also performed well on the independent validation dataset, producing AUCs of 0.8548 and 0.8067 (Figure 5C,D), respectively. When evaluated on the combined dataset, the AUCs increased to 0.9856 and 0.9744 (Figure 5E,F), further confirming the robustness and diagnostic potential of these models in distinguishing between cirrhosis and HCC.

To further evaluate the clinical applicability of the 11-gene signature, optimal cutoff values were determined for each model using ROC curve analysis. These cutoffs allowed patient stratification into high-risk (positive) and low-risk (negative) groups. The LASSO, XGBoost, and Random Forest models achieved robust diagnostic performance, with sensitivity and specificity values ranging from 0.78 to 0.95 across independent test datasets (Table S2). In addition, the importance of each gene in the different models was assessed and visualized (Figure S3), highlighting their relative contributions to the predictive performance. This demonstrates that the expression patterns of the selected genes can be directly applied in clinical settings to distinguish cirrhosis patients at high risk of HCC from those at lower risk. In clinical practice, the expression levels of these genes can be measured in patient samples and input into the predictive models to determine the risk category based on the established cutoff, thereby guiding early monitoring and intervention.

4. Discussion

Hepatocellular carcinoma is a globally prevalent malignancy and ranks as the third leading cause of cancer-related mortality worldwide [39,40]. Liver cirrhosis, frequently resulting from chronic liver diseases such as hepatitis B virus infection, is a major risk factor for HCC and represents a common pathological transition toward tumorigenesis [41,42]. Elucidating the cellular and molecular alterations that occur during the progression from cirrhosis to HCC is crucial for improving early diagnostic strategies and patient outcomes. In this study, we systematically analyzed single-cell transcriptomic data from patients with HBV infection, cirrhosis, and HCC, focusing on disease-associated changes in the hepatic microenvironment. Cell annotation and compositional analysis revealed a marked increase in macrophage abundance across disease stages, suggesting the potential key role of macrophages in the transition from cirrhosis to HCC. Unlike previous hypothesis-driven studies, we employed a data-driven approach to identify significantly altered cell types and subsequently focused on macrophages for in-depth characterization of their transcriptional and functional states in cirrhosis and HCC. Through differential gene expression analysis, M1/M2 polarization scoring [43], and intercellular communication profiling, we comprehensively identified the phenotypic differences in macrophages between these two disease states. Furthermore, we applied high-dimensional weighted gene co-expression network analysis to construct macrophage-specific co-expression modules, identifying functionally relevant gene clusters and hub genes. These findings were cross-validated with bulk RNA-seq data to enhance the robustness and clinical relevance of our results. Finally, using LASSO regression, we identified 11 key genes and developed predictive models based on multiple machine learning algorithms. These models demonstrated excellent diagnostic performance with area under the curve values exceeding 0.9 across independent datasets, underscoring their generalizability and translational potential. Collectively, our findings reveal dynamic transcriptional changes in macrophages during the cirrhosis-to-HCC transition, identifying novel biomarker candidates and analytic frameworks for early HCC diagnosis. Importantly, the final 11 genes were prioritized because they consistently showed stable predictive performance across single-cell and bulk datasets, were repeatedly validated by multiple machine learning approaches, and exhibited clear biological relevance to macrophage function and hepatocarcinogenesis, whereas other candidates lacked such robustness.

It is noteworthy that although approximately 80% of HCC arises in the context of cirrhosis, a subset of patients develop HCC without cirrhosis through distinct mechanisms. Metabolism-related single nucleotide polymorphisms (such as PNPLA3, TM6SF2, and HSD17B13) increase susceptibility to HCC in non-cirrhotic individuals, while environmental carcinogens, including aflatoxin B1 and aristolochic acid, can directly induce somatic mutations and synergize with HBV infection to promote tumorigenesis. At the molecular level, TERT promoter mutations represent the most frequent early “gatekeeper” event, with additional alterations in CTNNB1, TP53, and epigenetic regulators (BAP1, ARID1A/B, ARID2) further driving tumor progression. Importantly, these alterations are not dependent on a fibrotic background, providing a plausible explanation for HCC development in non-cirrhotic livers. Moreover, in NASH-related HCC, immunological mechanisms play a critical role, as autoreactive CD8+PD1+ T cells induce hepatocyte death and impair immune surveillance, thereby facilitating tumor formation.

Liver fibrosis is characterized by the excessive accumulation of extracellular matrix (ECM) components in the liver due to chronic liver injury from various etiologies. By contrast, cirrhosis represents an advanced stage of fibrosis [44]. In our study, the proportion of macrophages increased progressively with disease progression, reflecting their dynamic functional states in different pathological contexts. Notably, fibrosis-associated pathways were highly enriched in macrophages from cirrhotic tissues, including extracellular matrix organization, extracellular structure organization, and pathways related to chemotaxis and leukocyte migration, which are directly involved in ECM deposition and macrophage recruitment. This phenomenon is consistent with the pathological mechanisms of liver fibrosis, in which damaged or apoptotic hepatocytes release damage-associated molecular patterns (DAMPs) that activate hepatic stellate cells (HSCs), promoting their transdifferentiation into collagen-producing myofibroblasts [41,45]. During this process, lymphocytes and macrophages are recruited and activated, secreting pro-inflammatory and pro-fibrotic cytokines that further stimulate activation of HSC and progression of fibrosis [46,47]. The activated HSCs, in turn, contribute to ECM production as part of tissue repair responses [48]; however, persistent ECM accumulation eventually leads to pathological fibrosis. In our study, hepatic stellate cells were also identified, supporting the presence of fibrogenic activity and further validating the functional phenotype of macrophages observed in cirrhosis.

In contrast, macrophages in hepatocellular carcinoma (HCC) exhibited reduced enrichment of immune-related pathways, such as leukocyte-mediated immunity, lymphocyte-mediated immunity, and antigen receptor-mediated signaling; however, they showed increased activity in pathways associated with metabolic reprogramming. Tumor cells in HCC frequently face hypoxic and nutrient-deprived conditions and undergo adaptive metabolic shifts—referred to as metabolic reprogramming—to meet high energy demands or reshape the tumor microenvironment in favor of growth [49]. Pathways such as amino acid metabolism, fatty acid metabolism, and organic acid catabolism observed in this study are strongly associated with the Warburg effect and alter amino acid metabolism, both of which are hallmarks of tumor metabolic reprogramming [50,51]. The weakened immune response to HCC may be largely attributed to the immunosuppressive roles of tumor-associated macrophages (TAMs) [52]. M2-polarized TAMs can directly interact with myeloid-derived suppressor cells (MDSCs) and actively suppress T-cell-mediated anti-tumor immunity [53]. Moreover, they express immune-inhibitory surface molecules, such as PD-L1 and B7-H4 [54], which contribute to immune evasion. Collectively, these findings explain why immune activation in HCC is significantly lower compared to cirrhosis.

In recent years, a growing body of research in the literature has highlighted the pivotal role of macrophage polarization in numerous pathophysiological processes, including inflammation, tumor progression, tissue repair, and metabolic regulation [55,56]. Interestingly, these pathological mechanisms are also prominent features of liver diseases, suggesting a close association between macrophage polarization and the progression or resolution of hepatic disorders such as hepatitis, fibrosis, and hepatocellular carcinoma (HCC) [57,58]. In general, macrophage polarization is broadly classified into two phenotypes: classically activated M1 and alternatively activated M2 macrophages [43]. M1 macrophages are primarily involved in antigen presentation and exhibit pro-inflammatory, antimicrobial, and anti-tumor functions. In contrast, M2 macrophages are associated with anti-inflammatory responses, tissue remodeling, parasite resistance, and the promotion of angiogenesis, immune regulation, and tumor progression [59]. In this study, we quantified macrophage polarization states across hepatitis, cirrhosis, and HCC samples. We observed a progressive decline in M1 polarization scores and a corresponding increase in M2 polarization scores as disease advanced. Chronic hepatitis induced by viral infections, including hepatitis B virus (HBV) or hepatitis C virus (HCV), shows enhanced M1 polarization, while M2 polarization is suppressed [60–62]. In contrast, the M2 polarization of macrophages in HCC plays a key role in promoting immune suppression, angiogenesis, and cancer cell metastasis, resulting in the highest M2 polarization scores observed in HCC tissues. These findings are highly consistent with previous studies and further reinforce the role of M2-type macrophages in shaping an immunosuppressive tumor microenvironment that favors HCC progression [48].

In the cell–cell communication analysis of this study, we observed that although the overall number of intercellular interactions in cirrhotic tissues was lower than that in hepatocellular carcinoma (HCC), the communication frequency of T cells and neutrophils was notably higher in cirrhosis. This may be attributed to the immunosuppressive microenvironment and metabolic reprogramming characteristic of HCC, consistent with our pathway enrichment analysis findings. Furthermore, macrophage–neutrophil interactions were significantly enhanced in cirrhotic tissues, particularly in signaling pathways such as TNF–TNFRSF1A/B, MIF–(CD74+CD44), and multiple MHC class II ligand–receptor pairs. These results suggest that macrophages may actively participate in pro-inflammatory signaling and antigen presentation during the early stages of liver disease. In contrast, these pathways were generally weakened in HCC, and several immunoregulatory signals, including TGFB1–TGFB1R and PTPRC–MRC1, were also significantly downregulated, reflecting the formation of an immunosuppressive or immune-evasive environment. This shift in communication patterns may contribute to the transition from chronic inflammation to tumorigenesis, supporting the concept of the “inflammation-to-cancer continuum”.

Traditional weighted gene co-expression network analysis (WGCNA) is primarily designed for bulk transcriptomic data and may perform suboptimally in high-dimensional scRNA-seq datasets. To address this, we applied hdWGCNA to identify co-expression modules from macrophage-specific scRNA-seq data. Modules with a significant overlap of hub genes and macrophage-specific differentially expressed genes were selected for further analysis. Considering that macrophage-derived feature genes might not show significant differential expression at the bulk tissue level, we performed differential expression analysis using bulk transcriptomic datasets of liver cirrhosis and HCC. We then filtered genes that were significantly altered and exhibited consistent expression trends with macrophage-specific genes. Eleven genes were identified, many of which have been previously reported as biomarkers or functional regulators in HCC or other cancers. CENPF serves as both an important prognostic biomarker and a key oncogenic driver in HCC; its role has been independently confirmed through both

bioinformatics and experimental validation, and it can cooperate with FOXM1 to co-regulate downstream targets such as POLD1, thereby promoting hepatocarcinogenesis and tumor progression [36,63]. CFP is significantly downregulated in HCC, and its low expression is associated with poorer overall survival (OS) and disease-free survival (DFS), suggesting that it may serve as a potential prognostic biomarker and immune-related therapeutic target [64]. CYP2C8 is also downregulated in HCC tissues, and its reduced expression correlates with unfavorable OS and DFS. Moreover, CYP2C8, together with FCN3 and other prognosis-related genes, was incorporated into an 11-gene risk prediction model that accurately predicted overall survival [64–66]. KRT19 interacts with long non-coding RNAs to activate signaling pathways related to tumor proliferation and metastasis [67,68]. MARCO is downregulated in HCC tissues, and its low expression is associated with higher alpha-fetoprotein (AFP) levels, microvascular invasion, and poorer OS and DFS following liver transplantation. MARCO has been identified as an independent prognostic factor and is mainly expressed in macrophages, where it is linked to M2-like polarization [69]. MKI67 is an important prognostic gene in HCC, showing high expression in HBV/HCV-associated cases and being incorporated into hypoxia-related prognostic models, indicating its involvement in virus-induced tumor progression and hypoxia-driven cellular proliferation [37,70]. NUPR1 is highly expressed in tumor-associated macrophages within HCC and is related to immunosuppressive polarization and poor immunotherapeutic response, suggesting its potential as a biomarker and therapeutic target in HCC [71]. SOD3 is downregulated in liver fibrosis, and its deficiency promotes hepatic stellate cell activation and epithelial–mesenchymal transition (EMT), thereby aggravating liver injury and collagen deposition [72]. Finally, TOP2A is significantly upregulated in HCC tissues, and its high expression is associated with advanced tumor stage, poor differentiation, and unfavorable prognosis, particularly in Asian populations, suggesting that TOP2A may serve as a molecular marker for HCC progression and prognosis [35,73]. These genes are closely linked to pathways involved in cirrhosis-to-HCC progression and patient prognosis, making them suitable for diagnostic modeling. Using these genes, we constructed machine learning models—including LASSO regression, random forest, and XGBoost—trained on bulk transcriptomic data and validated on independent cohorts. All models demonstrated robust performance in distinguishing cirrhosis from HCC, underscoring their potential clinical utility. In this study, we observed that some genes showed opposite expression trends between single-cell RNA sequencing (scRNA-seq) and bulk RNA sequencing (bulk RNA-seq). This discrepancy largely reflects the distinct nature of the two technologies: scRNA-seq highlights features of minor cell subpopulations, whereas bulk data represent averaged tissue signals and are strongly influenced by changes in cell composition. For instance, CYP2C8 appeared downregulated in bulk HCC data but retained expression in specific subsets at the single-cell level, suggesting regulation shaped by both cellular composition and functional states. Such differences are not contradictory but rather underscore the complementary value of integrating scRNA-seq and bulk RNA-seq for a more comprehensive understanding of disease progression.

Despite the novel insights provided by this study, several limitations should be acknowledged. First, our analyses were primarily based on publicly available single-cell RNA sequencing (scRNA-seq) and bulk RNA sequencing datasets derived from different cohorts. The inherent batch effects and population heterogeneity may limit the generalizability of our findings. Moreover, the current conclusions are mainly supported by bioinformatic analyses without experimental validation, which remains a key limitation. In future work, we plan to perform quantitative real-time PCR (qRT-PCR) and immunohistochemical (IHC) assays using cirrhosis and HCC samples to validate

the expression and biological relevance of the 11 macrophage-related genes identified in this study.

Second, although the machine learning-based diagnostic models demonstrated promising performance, their robustness requires further evaluation in larger, multi-center independent cohorts. Future studies should validate the 11 macrophage-related genes and predictive models in larger, multi-center cohorts and incorporate longitudinal data to confirm their predictive value. Mechanistic experiments are also required to clarify how these genes regulate macrophage remodeling and the tumor microenvironment. Advances in spatial transcriptomics and multi-omics integration will further provide a comprehensive biological framework and accelerate their translation into personalized clinical applications.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/genes16101213/s1>, Figure S1: UMAP plots showing cell annotations for integrated and disease-specific datasets. Figure S2: Polarization states of macrophages in hepatitis, cirrhosis, and HCC. Figure S3: Feature importance of LASSO, XGBoost, and Random Forest models; Table S1: Summary of cell numbers for each sample and total cells in each disease group; Table S2: Summary of performance metrics (AUC, sensitivity, specificity) for predictive models.

Author Contributions: Conceptualization, Y.X. and Z.Z.; methodology, Y.X. and Z.Z.; validation, Z.Z. and W.Z.; formal analysis, Z.Z.; resources, Y.X. and J.S.; data curation, Z.Z. and C.Z.; writing—original draft preparation, Z.Z.; writing—review and editing, Y.X. and Z.Z.; visualization, Z.Z. and X.Y.; project administration, Y.X. and J.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Natural Science Foundation of Chongqing, China, grant number cstc2021jcyj-msxmX0834, National Undergraduate Innovation and Entrepreneurship Training Program, grant number 202310617015 and this work was also supported by the Natural Science Foundation of Chongqing, China, grant number cstc2021jcyj-msxmX0848.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available in the Gene Expression Omnibus (GEO) database at <https://www.ncbi.nlm.nih.gov/geo/>, accessed on 30 September 2024, reference numbers GSE136103, GSE186343, GSE242889, GSE202642, GSE25097, and GSE14323. These data were derived from the following resources available in the public domain: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE136103>, <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE186343>, <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE242889>, <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE202642>, <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE25097>, and <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE14323> (accessed on 30 September 2024).

Acknowledgments: The authors would like to express their sincere gratitude to Yongfang Xie for her invaluable guidance on the conceptual development of this manuscript and her continuous support beyond the academic scope. We also thank Pu Liu for his assistance during the research process. Furthermore, we are grateful to Yu Zhong, and Wending Pang and Yuhang Wang for their technical support in the revision and generation of the figures presented in this study.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Yang, J.D.; Hainaut, P.; Gores, G.J.; Amadou, A.; Plymoth, A.; Roberts, L.R. A Global View of Hepatocellular Carcinoma: Trends, Risk, Prevention and Management. *Nat. Rev. Gastroenterol. Hepatol.* **2019**, *16*, 589–604. [\[CrossRef\]](#)
2. European Association for the Study of the Liver. EASL Clinical Practice Guidelines: Management of Hepatocellular Carcinoma. *J. Hepatol.* **2018**, *69*, 182–236. [\[CrossRef\]](#)
3. Forner, A.; Reig, M.; Bruix, J. Hepatocellular Carcinoma. *Lancet* **2018**, *391*, 1301–1314. [\[CrossRef\]](#)

4. Llovet, J.M.; Zucman-Rossi, J.; Pikarsky, E.; Sangro, B.; Schwartz, M.; Sherman, M.; Gores, G. Hepatocellular Carcinoma. *Nat. Rev. Dis. Primer* **2016**, *2*, 16018. [\[CrossRef\]](#)
5. Jia, W.; Liang, S.; Cheng, B.; Ling, C. The Role of Cancer-Associated Fibroblasts in Hepatocellular Carcinoma and the Value of Traditional Chinese Medicine Treatment. *Front. Oncol.* **2021**, *11*, 763519. [\[CrossRef\]](#)
6. Yoshiji, H.; Nagoshi, S.; Akahane, T.; Asaoka, Y.; Ueno, Y.; Ogawa, K.; Kawaguchi, T.; Kurosaki, M.; Sakaida, I.; Shimizu, M.; et al. Evidence-Based Clinical Practice Guidelines for Liver Cirrhosis 2020. *J. Gastroenterol.* **2021**, *56*, 593–619. [\[CrossRef\]](#)
7. Morgan, R.L.; Baack, B.; Smith, B.D.; Yartel, A.; Pitasi, M.; Falck-Ytter, Y. Eradication of Hepatitis C Virus Infection and the Development of Hepatocellular Carcinoma: A Meta-Analysis of Observational Studies. *Ann. Intern. Med.* **2013**, *158*, 329–337. [\[CrossRef\]](#) [\[PubMed\]](#)
8. van der Meer, A.J.; Veldt, B.J.; Feld, J.J.; Wedemeyer, H.; Dufour, J.-F.; Lammert, F.; Duarte-Rojo, A.; Heathcote, E.J.; Manns, M.P.; Kuske, L.; et al. Association between Sustained Virological Response and All-Cause Mortality among Patients with Chronic Hepatitis C and Advanced Hepatic Fibrosis. *JAMA* **2012**, *308*, 2584–2593. [\[CrossRef\]](#) [\[PubMed\]](#)
9. Wen, Y.; Lambrecht, J.; Ju, C.; Tacke, F. Hepatic Macrophages in Liver Homeostasis and Diseases-Diversity, Plasticity and Therapeutic Opportunities. *Cell. Mol. Immunol.* **2021**, *18*, 45–56. [\[CrossRef\]](#)
10. Casanova-Acebes, M.; Dalla, E.; Leader, A.M.; LeBerichel, J.; Nikolic, J.; Morales, B.M.; Brown, M.; Chang, C.; Troncoso, L.; Chen, S.T.; et al. Tissue-Resident Macrophages Provide a pro-Tumorigenic Niche to Early NSCLC Cells. *Nature* **2021**, *595*, 578–584. [\[CrossRef\]](#) [\[PubMed\]](#)
11. Laviron, M.; Boissonnas, A. Ontogeny of Tumor-Associated Macrophages. *Front. Immunol.* **2019**, *10*, 1799. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Mass, E.; Ballesteros, I.; Farlik, M.; Halbritter, F.; Günther, P.; Crozet, L.; Jacome-Galarza, C.E.; Händler, K.; Klughammer, J.; Kobayashi, Y.; et al. Specification of Tissue-Resident Macrophages during Organogenesis. *Science* **2016**, *353*, aaf4238. [\[CrossRef\]](#) [\[PubMed\]](#)
13. MacParland, S.A.; Liu, J.C.; Ma, X.-Z.; Innes, B.T.; Bartczak, A.M.; Gage, B.K.; Manuel, J.; Khoo, N.; Echeverri, J.; Linares, I.; et al. Single Cell RNA Sequencing of Human Liver Reveals Distinct Intrahepatic Macrophage Populations. *Nat. Commun.* **2018**, *9*, 4383. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Zhao, J.; Zhang, S.; Liu, Y.; He, X.; Qu, M.; Xu, G.; Wang, H.; Huang, M.; Pan, J.; Liu, Z.; et al. Single-Cell RNA Sequencing Reveals the Heterogeneity of Liver-Resident Immune Cells in Human. *Cell Discov.* **2020**, *6*, 22. [\[CrossRef\]](#)
15. Kourou, K.; Exarchos, T.P.; Exarchos, K.P.; Karamouzis, M.V.; Fotiadis, D.I. Machine Learning Applications in Cancer Prognosis and Prediction. *Comput. Struct. Biotechnol. J.* **2015**, *13*, 8–17. [\[CrossRef\]](#)
16. Swanson, K.; Wu, E.; Zhang, A.; Alizadeh, A.A.; Zou, J. From Patterns to Patients: Advances in Clinical Machine Learning for Cancer Diagnosis, Prognosis, and Treatment. *Cell* **2023**, *186*, 1772–1791. [\[CrossRef\]](#)
17. Boehm, K.M.; Khosravi, P.; Vanguri, R.; Gao, J.; Shah, S.P. Harnessing Multimodal Data Integration to Advance Precision Oncology. *Nat. Rev. Cancer* **2022**, *22*, 114–126. [\[CrossRef\]](#)
18. Hosny, A.; Parmar, C.; Quackenbush, J.; Schwartz, L.H.; Aerts, H.J.W.L. Artificial Intelligence in Radiology. *Nat. Rev. Cancer* **2018**, *18*, 500–510. [\[CrossRef\]](#)
19. López-Bueno, R.; Andersen, L.L.; Koyanagi, A.; Núñez-Cortés, R.; Calatayud, J.; Casaña, J.; del Pozo Cruz, B. Thresholds of Handgrip Strength for All-Cause, Cancer, and Cardiovascular Mortality: A Systematic Review with Dose-Response Meta-Analysis. *Ageing Res. Rev.* **2022**, *82*, 101778. [\[CrossRef\]](#)
20. Mokoatle, M.; Marivate, V.; Mapiye, D.; Bornman, R.; Hayes, V.M. A Review and Comparative Study of Cancer Detection Using Machine Learning: SBERT and SimCSE Application. *BMC Bioinform.* **2023**, *24*, 112. [\[CrossRef\]](#)
21. Hao, Y.; Stuart, T.; Kowalski, M.H.; Choudhary, S.; Hoffman, P.; Hartman, A.; Srivastava, A.; Molla, G.; Madad, S.; Fernandez-Granda, C.; et al. Dictionary Learning for Integrative, Multimodal and Scalable Single-Cell Analysis. *Nat. Biotechnol.* **2024**, *42*, 293–304. [\[CrossRef\]](#) [\[PubMed\]](#)
22. Hafemeister, C.; Satija, R. Normalization and Variance Stabilization of Single-Cell RNA-Seq Data Using Regularized Negative Binomial Regression. *Genome Biol.* **2019**, *20*, 296. [\[CrossRef\]](#)
23. Becht, E.; McInnes, L.; Healy, J.; Dutertre, C.-A.; Kwok, I.W.H.; Ng, L.G.; Ginhoux, F.; Newell, E.W. Dimensionality Reduction for Visualizing Single-Cell Data Using UMAP. *Nat. Biotechnol.* **2018**, *37*, 38–44. [\[CrossRef\]](#)
24. Hu, C.; Li, T.; Xu, Y.; Zhang, X.; Li, F.; Bai, J.; Chen, J.; Jiang, W.; Yang, K.; Ou, Q.; et al. CellMarker 2.0: An Updated Database of Manually Curated Cell Markers in Human/Mouse and Web Tools Based on scRNA-Seq Data. *Nucleic Acids Res.* **2023**, *51*, D870–D876. [\[CrossRef\]](#)
25. Jiang, S.; Qian, Q.; Zhu, T.; Zong, W.; Shang, Y.; Jin, T.; Zhang, Y.; Chen, M.; Wu, Z.; Chu, Y.; et al. Cell Taxonomy: A Curated Repository of Cell Types with Multifaceted Characterization. *Nucleic Acids Res.* **2023**, *51*, D853–D860. [\[CrossRef\]](#) [\[PubMed\]](#)
26. Subramanian, A.; Tamayo, P.; Mootha, V.K.; Mukherjee, S.; Ebert, B.L.; Gillette, M.A.; Paulovich, A.; Pomeroy, S.L.; Golub, T.R.; Lander, E.S.; et al. Gene Set Enrichment Analysis: A Knowledge-Based Approach for Interpreting Genome-Wide Expression Profiles. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 15545. [\[CrossRef\]](#)

27. Yu, G.; Wang, L.-G.; Han, Y.; He, Q.-Y. clusterProfiler: An R Package for Comparing Biological Themes Among Gene Clusters. *OMICS J. Integr. Biol.* **2012**, *16*, 284. [\[CrossRef\]](#)
28. Jin, S.; Guerrero-Juarez, C.F.; Zhang, L.; Chang, I.; Ramos, R.; Kuan, C.-H.; Myung, P.; Plikus, M.V.; Nie, Q. Inference and Analysis of Cell-Cell Communication Using CellChat. *Nat. Commun.* **2021**, *12*, 1088. [\[CrossRef\]](#)
29. Xu, K.; Dong, M.; Wu, Z.; Luo, L.; Xie, F.; Li, F.; Huang, H.; Wang, F.; Xiong, X.; Wen, Z. Single-Cell RNA Sequencing Identifies Crucial Genes Influencing the Polarization of Tumor-Associated Macrophages in Liver Cancer. *Int. J. Genomics* **2024**, *2024*, 7263358. [\[CrossRef\]](#) [\[PubMed\]](#)
30. Ke, H.; Yuan, R.; Liu, H.; Luo, M.; Hu, H.; Zhang, E.; Zhuang, K.; Yang, Y.; Yang, R. Serum Protein Biomarkers for HCC Risk Prediction in HIV/HBV Co-Infected People: A Clinical Proteomic Study Using Mass Spectrometry. *Front. Immunol.* **2023**, *14*, 1282469. [\[CrossRef\]](#)
31. Crouchet, E.; Bandiera, S.; Fujiwara, N.; Li, S.; El Saghiere, H.; Fernández-Vaquero, M.; Riedl, T.; Sun, X.; Hirschfield, H.; Jühling, F.; et al. A Human Liver Cell-Based System Modeling a Clinical Prognostic Liver Signature for Therapeutic Discovery. *Nat. Commun.* **2021**, *12*, 5525. [\[CrossRef\]](#)
32. Chaudhary, K.; Poirion, O.B.; Lu, L.; Garmire, L.X. Deep Learning-Based Multi-Omics Integration Robustly Predicts Survival in Liver Cancer. *Clin. Cancer Res.* **2018**, *24*, 1248–1259. [\[CrossRef\]](#)
33. Sun, J.; Mao, F.; Liu, C.; Zhang, F.; Jiang, D.; Guo, W.; Huo, L.; Zhou, L.; Lau, W.Y.; Shi, J.; et al. Combined FOLFOX4 with All-Trans Retinoic Acid versus FOLFOX4 with Placebo in Treatment of Advanced Hepatocellular Carcinoma with Extrahepatic Metastasis: A Randomized, Double-Blind Comparative Study. *Signal Transduct. Target. Ther.* **2023**, *8*, 368. [\[CrossRef\]](#) [\[PubMed\]](#)
34. Zhou, X.; Li, T.-M.; Luo, J.-Z.; Lan, C.-L.; Wei, Z.-L.; Fu, T.-H.; Liao, X.-W.; Zhu, G.-Z.; Ye, X.-P.; Peng, T. CYP2C8 Suppress Proliferation, Migration, Invasion and Sorafenib Resistance of Hepatocellular Carcinoma via PI3K/Akt/P27kip1 Axis. *J. Hepatocell. Carcinoma* **2021**, *8*, 1323–1338. [\[CrossRef\]](#)
35. Wang, T.; Lu, J.; Wang, R.; Cao, W.; Xu, J. TOP2A Promotes Proliferation and Metastasis of Hepatocellular Carcinoma Regulated by miR-144-3p. *J. Cancer* **2022**, *13*, 589–601. [\[CrossRef\]](#)
36. Huang, Y.; Chen, X.; Wang, L.; Wang, T.; Tang, X.; Su, X. Centromere Protein F (CENPF) Serves as a Potential Prognostic Biomarker and Target for Human Hepatocellular Carcinoma. *J. Cancer* **2021**, *12*, 2933–2951. [\[CrossRef\]](#)
37. Saeed, M.M.; Ma, X.; Fu, X.; Ullah, I.; Ali, T.; Bai, C.; Liu, Y.; Dong, C.; Cui, X. RACGAP1 and MKI67 Are Potential Prognostic Biomarker in Hepatocellular Carcinoma Caused by HBV/HCV via Lactylation. *Front. Oncol.* **2025**, *15*, 1537084. [\[CrossRef\]](#) [\[PubMed\]](#)
38. Zhan, Y.; Zhang, Z.; Liu, Y.; Fang, Y.; Xie, Y.; Zheng, Y.; Li, G.; Liang, L.; Ding, Y. NUPR1 Contributes to Radiation Resistance by Maintaining ROS Homeostasis via AhR/CYP Signal Axis in Hepatocellular Carcinoma. *BMC Med.* **2022**, *20*, 365. [\[CrossRef\]](#) [\[PubMed\]](#)
39. Toh, M.R.; Wong, E.Y.T.; Wong, S.H.; Ng, A.W.T.; Loo, L.-H.; Chow, P.K.-H.; Ngeow, J. Global Epidemiology and Genetics of Hepatocellular Carcinoma. *Gastroenterology* **2023**, *164*, 766–782. [\[CrossRef\]](#)
40. Brar, G.; Greten, T.F.; Graubard, B.I.; McNeel, T.S.; Petrick, J.L.; McGlynn, K.A.; Altekruse, S.F. Hepatocellular Carcinoma Survival by Etiology: A SEER-Medicare Database Analysis. *Hepatol. Commun.* **2020**, *4*, 1541–1551. [\[CrossRef\]](#)
41. Zhou, W.-C.; Zhang, Q.-B.; Qiao, L. Pathogenesis of Liver Cirrhosis. *World J. Gastroenterol.* **2014**, *20*, 7312–7324. [\[CrossRef\]](#)
42. Smith, A.; Baumgartner, K.; Bositis, C. Cirrhosis: Diagnosis and Management. *Am. Fam. Physician* **2019**, *100*, 759–770.
43. Yunna, C.; Mengru, H.; Lei, W.; Weidong, C. Macrophage M1/M2 Polarization. *Eur. J. Pharmacol.* **2020**, *877*, 173090. [\[CrossRef\]](#)
44. Aydın, M.M.; Akçali, K.C. Liver Fibrosis. *Turk. J. Gastroenterol.* **2018**, *29*, 14–21. [\[CrossRef\]](#)
45. Elpek, G.Ö. Cellular and Molecular Mechanisms in the Pathogenesis of Liver Fibrosis: An Update. *World J. Gastroenterol.* **2014**, *20*, 7260–7276. [\[CrossRef\]](#)
46. Krenkel, O.; Tacke, F. Liver Macrophages in Tissue Homeostasis and Disease. *Nat. Rev. Immunol.* **2017**, *17*, 306–321. [\[CrossRef\]](#)
47. Barron, L.; Wynn, T.A. Fibrosis Is Regulated by Th2 and Th17 Responses and by Dynamic Interactions between Fibroblasts and Macrophages. *Am. J. Physiol. Gastrointest. Liver Physiol.* **2011**, *300*, G723–G728. [\[CrossRef\]](#)
48. Kisseleva, T.; Cong, M.; Paik, Y.; Scholten, D.; Jiang, C.; Benner, C.; Iwaisako, K.; Moore-Morris, T.; Scott, B.; Tsukamoto, H.; et al. Myofibroblasts Revert to an Inactive Phenotype during Regression of Liver Fibrosis. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 9448–9453. [\[CrossRef\]](#) [\[PubMed\]](#)
49. Yang, F.; Hilakivi-Clarke, L.; Shaha, A.; Wang, Y.; Wang, X.; Deng, Y.; Lai, J.; Kang, N. Metabolic Reprogramming and Its Clinical Implication for Liver Cancer. *Hepatology* **2023**, *78*, 1602–1624. [\[CrossRef\]](#) [\[PubMed\]](#)
50. Liberti, M.V.; Locasale, J.W. The Warburg Effect: How Does It Benefit Cancer Cells? *Trends Biochem. Sci.* **2016**, *41*, 211–218. [\[CrossRef\]](#) [\[PubMed\]](#)
51. Huang, Q.; Tan, Y.; Yin, P.; Ye, G.; Gao, P.; Lu, X.; Wang, H.; Xu, G. Metabolic Characterization of Hepatocellular Carcinoma Using Nontargeted Tissue Metabolomics. *Cancer Res.* **2013**, *73*, 4992–5002. [\[CrossRef\]](#) [\[PubMed\]](#)
52. Komohara, Y.; Fujiwara, Y.; Ohnishi, K.; Takeya, M. Tumor-Associated Macrophages: Potential Therapeutic Targets for Anti-Cancer Therapy. *Adv. Drug Deliv. Rev.* **2016**, *99*, 180–185. [\[CrossRef\]](#)

53. Parker, K.H.; Sinha, P.; Horn, L.A.; Clements, V.K.; Yang, H.; Li, J.; Tracey, K.J.; Ostrand-Rosenberg, S. HMGB1 Enhances Immune Suppression by Facilitating the Differentiation and Suppressive Activity of Myeloid-Derived Suppressor Cells. *Cancer Res.* **2014**, *74*, 5723–5733. [[CrossRef](#)] [[PubMed](#)]
54. Matsunaga, T.; Saito, H.; Ikeguchi, M. Increased B7-H1 and B7-H4 Expressions on Circulating Monocytes and Tumor-Associated Macrophages Are Involved in Immune Evasion in Patients with Gastric Cancer. *Yonago Acta Med.* **2011**, *54*, 1–10. [[PubMed](#)]
55. Wang, L.; Zhang, H.; Sun, L.; Gao, W.; Xiong, Y.; Ma, A.; Liu, X.; Shen, L.; Li, Q.; Yang, H. Manipulation of Macrophage Polarization by Peptide-Coated Gold Nanoparticles and Its Protective Effects on Acute Lung Injury. *J. Nanobiotechnol.* **2020**, *18*, 38. [[CrossRef](#)]
56. Xu, F.; Guo, M.; Huang, W.; Feng, L.; Zhu, J.; Luo, K.; Gao, J.; Zheng, B.; Kong, L.-D.; Pang, T.; et al. Annexin A5 Regulates Hepatic Macrophage Polarization via Directly Targeting PKM2 and Ameliorates NASH. *Redox Biol.* **2020**, *36*, 101634. [[CrossRef](#)]
57. Ma, P.-F.; Gao, C.-C.; Yi, J.; Zhao, J.-L.; Liang, S.-Q.; Zhao, Y.; Ye, Y.-C.; Bai, J.; Zheng, Q.-J.; Dou, K.-F.; et al. Cytotherapy with M1-Polarized Macrophages Ameliorates Liver Fibrosis by Modulating Immune Microenvironment in Mice. *J. Hepatol.* **2017**, *67*, 770–779. [[CrossRef](#)]
58. Yang, Y.; Ye, Y.-C.; Chen, Y.; Zhao, J.-L.; Gao, C.-C.; Han, H.; Liu, W.-C.; Qin, H.-Y. Crosstalk between Hepatic Tumor Cells and Macrophages via Wnt/ β -Catenin Signaling Promotes M2-Like Macrophage Polarization and Reinforces Tumor Malignant Behaviors. *Cell Death Dis.* **2018**, *9*, 793. [[CrossRef](#)]
59. Mantovani, A.; Biswas, S.K.; Galdiero, M.R.; Sica, A.; Locati, M. Macrophage Plasticity and Polarization in Tissue Repair and Remodelling. *J. Pathol.* **2013**, *229*, 176–185. [[CrossRef](#)]
60. Zhao, X.; Sun, L.; Mu, T.; Yi, J.; Ma, C.; Xie, H.; Liu, M.; Tang, H. An HBV-Encoded miRNA Activates Innate Immunity to Restrict HBV Replication. *J. Mol. Cell Biol.* **2020**, *12*, 263–276. [[CrossRef](#)]
61. Yi, H.; Zhang, Y.; Yang, X.; Li, M.; Hu, H.; Xiong, J.; Wang, N.; Jin, J.; Zhang, Y.; Song, Y.; et al. Hepatitis B Core Antigen Impairs the Polarization While Promoting the Production of Inflammatory Cytokines of M2 Macrophages via the TLR2 Pathway. *Front. Immunol.* **2020**, *11*, 535. [[CrossRef](#)]
62. Zhang, Q.; Wang, Y.; Zhai, N.; Song, H.; Li, H.; Yang, Y.; Li, T.; Guo, X.; Chi, B.; Niu, J.; et al. HCV Core Protein Inhibits Polarization and Activity of Both M1 and M2 Macrophages through the TLR2 Signaling Pathway. *Sci. Rep.* **2016**, *6*, 36160. [[CrossRef](#)]
63. Ho, D.W.-H.; Lam, W.-L.M.; Chan, L.-K.; Ng, I.O.-L. Investigation of Functional Synergism of CENPF and FOXM1 Identifies POLD1 as Downstream Target in Hepatocellular Carcinoma. *Front. Med.* **2022**, *9*, 860395. [[CrossRef](#)] [[PubMed](#)]
64. Wang, S.; Song, Z.; Tan, B.; Zhang, J.; Zhang, J.; Liu, S. Identification and Validation of Hub Genes Associated with Hepatocellular Carcinoma Via Integrated Bioinformatics Analysis. *Front. Oncol.* **2021**, *11*, 614531. [[CrossRef](#)] [[PubMed](#)]
65. Ren, X.; Ji, Y.; Jiang, X.; Qi, X. Downregulation of CYP2A6 and CYP2C8 in Tumor Tissues Is Linked to Worse Overall Survival and Recurrence-Free Survival from Hepatocellular Carcinoma. *BioMed Res. Int.* **2018**, *2018*, 5859415. [[CrossRef](#)]
66. Gao, X.; Ren, X.; Wang, F.; Ren, X.; Liu, M.; Cui, G.; Liu, X. Immunotherapy and Drug Sensitivity Predictive Roles of a Novel Prognostic Model in Hepatocellular Carcinoma. *Sci. Rep.* **2024**, *14*, 9509. [[CrossRef](#)] [[PubMed](#)]
67. Zhang, X.; Xu, X.; Zhang, Z.; Xue, C.; Kong, Z.; Wu, S.; Yun, X.; Fu, Y.; Zhu, C.; Qin, X. Linc-KILH Potentiates Notch1 Signaling through Inhibiting KRT19 Phosphorylation and Promotes the Malignancy of Hepatocellular Carcinoma. *Int. J. Biol. Sci.* **2021**, *17*, 768–780. [[CrossRef](#)]
68. Tang, J.; Zhuo, H.; Zhang, X.; Jiang, R.; Ji, J.; Deng, L.; Qian, X.; Zhang, F.; Sun, B. A Novel Biomarker Linc00974 Interacting with KRT19 Promotes Proliferation and Metastasis in Hepatocellular Carcinoma. *Cell Death Dis.* **2014**, *5*, e1549. [[CrossRef](#)]
69. Zhang, Q.; Wei, Y.; Li, Y.; Jiao, X. Low MARCO Expression Is Associated with Poor Survival in Patients with Hepatocellular Carcinoma Following Liver Transplantation. *Cancer Manag. Res.* **2022**, *14*, 1935–1944. [[CrossRef](#)]
70. Li, K.; Yang, Y.; Ma, M.; Lu, S.; Li, J. Hypoxia-Based Classification and Prognostic Signature for Clinical Management of Hepatocellular Carcinoma. *World J. Surg. Oncol.* **2023**, *21*, 216. [[CrossRef](#)] [[PubMed](#)]
71. Cai, J.; Zhang, P.; Cai, Y.; Zhu, G.; Chen, S.; Song, L.; Du, J.; Wang, B.; Dai, W.; Zhou, J.; et al. Lactylation-Driven NUPR1 Promotes Immunosuppression of Tumor-Infiltrating Macrophages in Hepatocellular Carcinoma. *Adv. Sci.* **2025**, *12*, e2413095. [[CrossRef](#)] [[PubMed](#)]
72. Sun, Y.-L.; Bai, T.; Zhou, L.; Zhu, R.-T.; Wang, W.-J.; Liang, R.-P.; Li, J.; Zhang, C.-X.; Gou, J.-J. SOD3 Deficiency Induces Liver Fibrosis by Promoting Hepatic Stellate Cell Activation and Epithelial-Mesenchymal Transition. *J. Cell. Physiol.* **2021**, *236*, 4313–4329. [[CrossRef](#)] [[PubMed](#)]
73. Cai, H.; Shao, B.; Zhou, Y.; Chen, Z. High Expression of TOP2A in Hepatocellular Carcinoma Is Associated with Disease Progression and Poor Prognosis. *Oncol. Lett.* **2020**, *20*, 232. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.