

The integration of machine learning and multi-omics analysis provides a powerful approach to screen aging-related genes and predict prognosis and immunotherapy efficacy in hepatocellular carcinoma

Jiahui Shen^{1,*}, Han Gao^{2,*}, Bowen Li^{3,*}, Yan Huang³, Yinfang Shi²

¹Department of Pharmacy, Huzhou Maternity and Child Health Care Hospital, Huzhou, China

²Department of Stomatology, First Affiliated Hospital of Huzhou University, Huzhou, China

³School of Pharmacy, Anhui Medical University, Hefei, China

*Equal contribution

Correspondence to: Yinfang Shi, Yan Huang; email: 52136@zjhu.edu.cn, huangyan_@ahmu.edu.cn

Keywords: machine learning, aging, hepatocellular carcinoma, prognostic model, immunotherapy

Received: March 31, 2023

Accepted: June 15, 2023

Published: July 28, 2023

Copyright: © 2023 Shen et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/3.0/) (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ABSTRACT

Background: Hepatocellular carcinoma (HCC) is a highly malignant tumor with high incidence and mortality rates. Aging-related genes are closely related to the occurrence and development of cancer. Therefore, it is of great significance to evaluate the prognosis of HCC patients by constructing a model based on aging-related genes.

Method: Non-negative matrix factorization (NMF) clustering analysis was used to cluster the samples. The correlation between the risk score and immune cells, immune checkpoints, and Mismatch Repair (MMR) was evaluated through Spearman correlation test. Real Time Quantitative PCR (RT-qPCR) and immunohistochemistry were used to validate the expression levels of key genes in tissue and cells for the constructed model.

Result: By performing NMF clustering, we were able to effectively group the liver cancer samples into two distinct clusters. Considering the potential correlation between aging-related genes and the prognosis of liver cancer patients, we used aging-related genes to construct a prognostic model. Spearman correlation analysis showed that the model risk score was closely related to MMR and immune checkpoint expression. Drug sensitivity analysis also provided guidance for the clinical use of chemotherapy drugs. RT-qPCR showed that TFDP1, NDRG1, and FXR1 were expressed at higher levels in different liver cancer cell lines compared to normal liver cells.

Conclusion: In summary, we have developed an aging-related model to predict the prognosis of hepatocellular carcinoma and guide clinical drug treatment for different patients.

INTRODUCTION

Hepatocellular carcinoma is a highly malignant tumor and is one of the leading causes of cancer-related deaths worldwide [1]. Despite significant progress in the diagnosis and treatment of HCC, its incidence and mortality rates remain high [2]. In clinical practice, early diagnosis and treatment of HCC are of paramount

importance. Numerous studies have shown that early detection and intervention can greatly improve patient prognosis [3, 4]. Therefore, finding new methods for the diagnosis and treatment of HCC is currently a hot topic in HCC research. In recent years, extensive genomic and transcriptomic studies have shown that aberrant changes in many genes and signaling pathways may contribute to HCC formation and progression [5–7].

These research findings provide new ideas and methods for the diagnosis and treatment of HCC. Traditional treatment methods for HCC include surgical resection, radiation therapy and chemotherapy, but these treatments still have many problems and limitations [8, 9]. With the emergence of new technologies and drugs, personalized therapy is gradually becoming a new trend in HCC treatment [10]. For example, targeted therapy against tumor-related signaling pathways has become a hot topic in HCC treatment [11]. In addition, immunotherapy as a novel approach to HCC treatment has also received widespread attention [12, 13]. In summary, HCC is a serious disease. Due to the complex etiology and mechanisms of HCC, the therapeutic effects of HCC vary greatly. Therefore, new biomarkers and prognostic models are needed to achieve precision management for individuals.

Aging with the significant feature of permanent growth arrest is often a response to endogenous and exogenous stresses, including telomere dysfunction, oncogene activation, and persistent DNA damage [14]. The generation of senescent cells occurs throughout a person's life and plays a functional role in various physiological and pathological processes, including embryonic development, wound healing, host defense, and tumor suppression [15]. Studies have shown that aging is an effective barrier to prevent tumor development [16]. Cell senescence is associated with the decline of hematopoietic stem cell (HSC) function and an increased risk of malignancies in the hematopoietic system, especially leukemias, multiple myeloma, myelodysplastic syndromes, and lymphomas, which are more common in the elderly [14]. According to the literature, cell senescence can promote skin carcinogenesis through the p38MAPK and p44/42MAPK signaling pathways [17]. Additionally, research has found that the aging-related SIN3B can promote inflammation and pancreatic cancer progression [18]. Liu et al. research has shown that dysbiosis of the liver microbiota can cause activation and aging of hepatic stellate cells, thereby driving the progression of liver cirrhosis to hepatocellular carcinoma [19]. Previous research has revealed a close relationship between aging and cancer. However, current studies on the link between liver cancer and aging are often limited to individual molecules, and research on multiple key aging genes and liver cancer is still lacking.

The aim of this study is to construct a risk prognosis model by integrating multiple key genes related to aging and to explore the relationship between risk score and immune cell and tumor microenvironment by combining bulk and single-cell sequencing. Furthermore, we investigate their correlation with MMR, immune checkpoints, and IC50 to determine the effectiveness of

immune therapy and different chemotherapy drugs. In addition, we validated these results by multi-omics analysis and basic experiments.

MATERIALS AND METHODS

Data source

Transcriptomic and clinical data for hepatocellular carcinoma were downloaded from The Cancer Genome Atlas (TCGA) (<https://portal.gdc.cancer.gov/>) database, including 374 cancer samples and 50 normal samples. Clinical data included survival status, survival time, gender, grade, and TNM stage. The liver cancer-related dataset (GSE14520) was downloaded from the GEO database (<https://www.ncbi.nlm.nih.gov/>), and after collating the data, a total of 221 liver cancer samples were used for subsequent model verification analysis. Dataset GSE39791 was used for differential analysis of cancerous and paraneoplastic tissues.

NMF clustering analysis

Non-Negative Matrix Factorization (NMF) is a commonly used clustering method for subgroup identification. Prior to the clustering analysis, the data were filtered and sorted. Subsequently, the differential analysis and prognosis of 278 aging-related genes were performed, followed by the application of the NMF algorithm to classify the samples into two clusters, namely C1 and C2. The C1 cluster consisted of 68 samples, while the C2 cluster consisted of 153 samples.

Functional enrichment analysis

To explore the underlying biological processes and signaling pathways associated with the acquisition of differential genes, gene ontology (GO) and KEGG enrichment analyses were performed using the “clusterProfiler” R package. GO analysis included BP, CC and MF. The annotated gene set file is “c2.cp.kegg.v7.4.symbols.gmt” and “c5.go.v7.4.symbols.gmt”. A threshold value of P value < 0.05 was determined.

Single-cell data analysis

Single cell sequencing data were obtained from the GEO database (GSE146115). The Seurat package analyzes the dataset and clusters the samples after PCA dimensionality reduction and t-SNE dimensionality reduction. The SingleR package was used for cell type annotation of single-cell data.

Weighted gene co-expression network analysis

WGCNA is an analytical method for analyzing gene expression patterns of multiple samples. Genes with similar expression patterns can be clustered and the association between modules and specific traits or phenotypes can be analyzed. Aging-associated genes were used to construct the weighted correlation network, where prognosis-related modules were selected for further analysis.

Modeling analysis

A risk score model was constructed using the screened key genes, and a risk score was also calculated for each patient. A comprehensive evaluation of the role of these key molecules in the prognosis of patients with hepatocellular carcinoma. Lasso regression analysis was used to construct a prognostic model. The risk score of each HCC patient was calculated by the formula: risk score = (Expression of FXR1 \times coefficient) + (Expression of NDRG1 \times coefficient) + (Expression of TFDPI \times coefficient). The TCGA dataset is divided into training and test sets, while we use the GSE14520 dataset for further model validation. Survival analysis was performed using KM curves to assess whether there was a difference in survival between high and low risk groups of patients with liver cancer in the training and test sets. Subsequently, risk-survival curves were used to assess patient survival and death in the high and low risk groups and how the key genes that were modeled differed between the two groups. And the ROC curve is used to determine the effectiveness of this prediction model.

Immunoassay and drug sensitivity analysis

The Cibersort algorithm was used to quantify the abundance of various immune cells in each sample. In total, we evaluated 22 human immune cells. The immune score, stromal score and total score are derived by the ESTIMATE algorithm to assess the immune microenvironment in the tumor tissue. The tumor stemness index is an important metric used to assess the similarity between tumor cells and stem cells. Analyzing the correlation between the stemness index and the risk score can help predict whether the model can serve as an indicator of stemness. Correlations with immune checkpoints and MMR were calculated using Spearman correlation analysis to determine the suitability of hub genes for predicting the efficacy of immunotherapy. Immunotherapy was validated between different risk groups using the IMvigor210 dataset. The “oncoPredict” package was used to assess drug sensitivity in different groups, and we analyzed the current chemotherapeutic agents that have some relevance to liver cancer.

Real-Time quantitative PCR (RT-qPCR)

Total RNA was extracted using TRIZOL (#15596026, Invitrogen, USA), and the concentration was determined and placed on ice for use. Add the calculated amount of RNA to PCR tubes, generally reverse transcribe 500 ng of total RNA per tube, add PrimeScript™ RT Master Mix (Takara Bio, Japan) about 2 μ l, and add RNase free water (Takara Bio, Japan) to fix the volume to 10 μ l. Set up the reaction program: 37°C for 15 min, 85°C for 5 s. Reverse transcription was completed to obtain cDNA. Dilute the qPCR primers to 10 μ M and mix the Forward and Reverse primers in equal volumes. Prepare the qPCR system in the following proportions: 2xTB Green (#RR820L, Takara, Japan) 10 μ l, ddH₂O 8 μ l, template cDNA 1 μ l, primers (Forward primer + Reverse primer) 1 μ l. Run the qPCR according to the following procedure: 94°C for 2 min, 94°C for 30 s, 60°C for 32 s, 60–94°C for 45 cycles, and collect the solubility curve. Data were collected and qPCR results were analyzed [20]. The primer sequences are as follows: FXR1-F: 5'-GAGAAGACGGTATGGTTCCATTT-3', FXR1-R: 5'-AGGCGTTCATTCTTAGCTGT-3'; NDRG1-F: 5'-CTCCTGCAAGAGTTTGATGTCC-3', NDRG1-R: 5'-TCATGCCGATGTCATGGTAGG-3'; TFDPI-F: 5'-AATTGAAGCCAACGGAGAACTC-3', TFDPI-R: 5'-CGGTCTCTGAGGCGTACCA-3'; GAPDH-F: 5'-GGAGCGAGATCCCTCCAAAAT-3', GAPDH-R: 5'-GGCTGTTGTCATACTTCTCATGG-3'.

Immunohistochemistry analysis

The Human Protein Atlas (HPA) (<https://www.proteinatlas.org/>) database is based on proteomic, transcriptomic and systems biology data for statistical analysis, and covers protein expression in normal and tumor tissues. Among them, the expression of FXR1 and NDRG1 in the liver, which we used to construct aging-related prognostic models, was also included.

Data statistics

The Wilcoxon test was used for the analysis of differences between the two groups, while the correlation analysis was based on the Spearman correlation test. Kaplan-Meier analysis and log-rank test were used to compare the survival analysis between the two groups. *P* values are bilateral and *P* < 0.05 is considered statistically significant. R software (version 4.2.2) was used to perform the statistical analysis.

Data availability statement

The article/supplementary material contains the original contributions presented in this study. For further information, please contact the corresponding author.

RESULTS

NMF clustering analysis

First, we draw a flowchart illustrating the whole analysis process in detail (Figure 1). As mentioned previously, we obtained RNA-Seq data and relevant clinical information for hepatocellular carcinoma through the TCGA database. A total of 278 aging-

associated genes were obtained from the CellAge database (<https://genomics.senescence.info/cells/>), and we performed NMF Clustering based on the expression matrix of these aging-associated genes. It can be seen that the samples are better divided into two clusters (Figure 2A). KM analysis was used to analyze the prognostic differences between the two groups of patients, and we saw a poorer prognosis for patients in the C1 cluster (Figure 2B). Immediately after, we

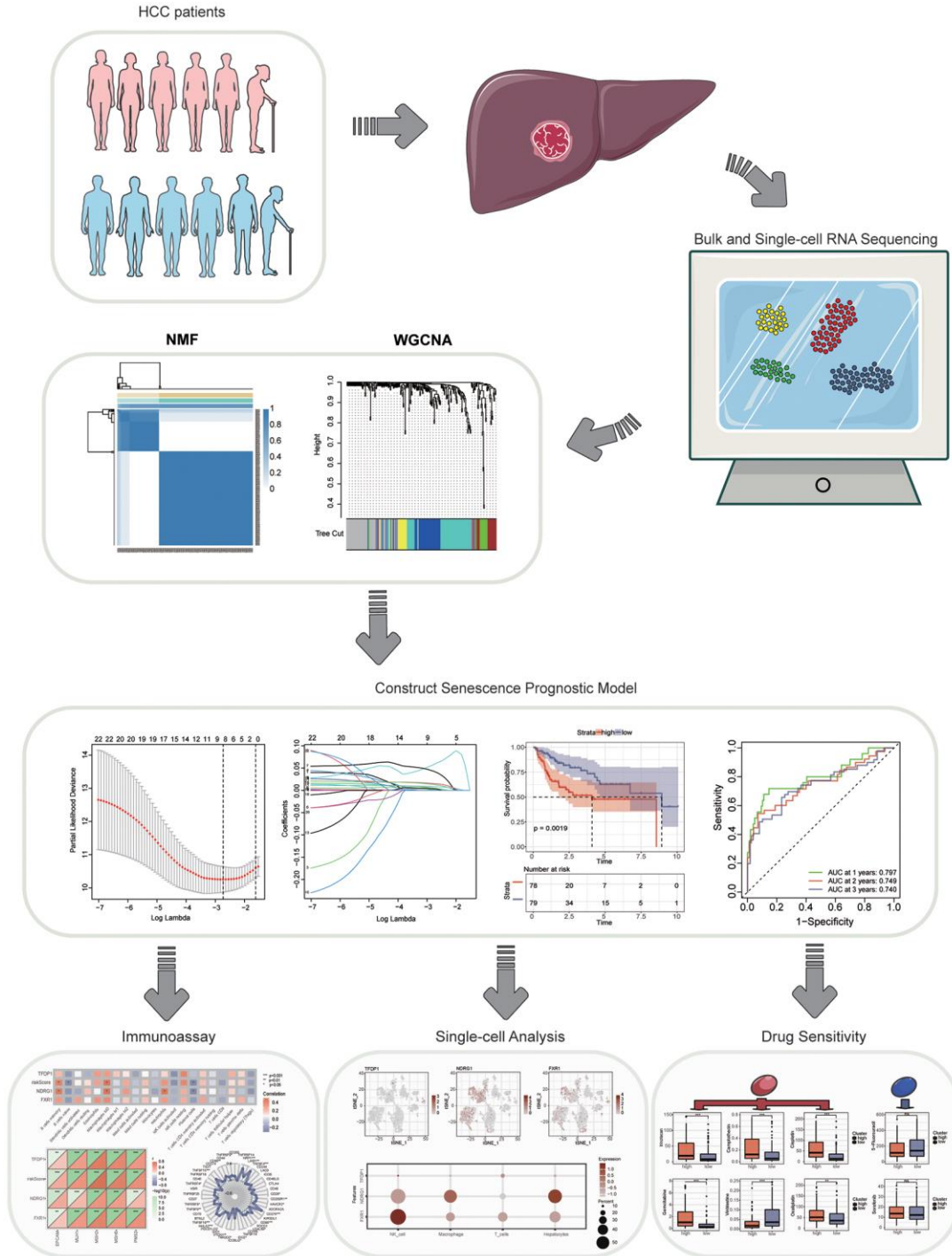


Figure 1. A flow chart of the manuscript.

performed a differential analysis of the two clusters of samples and Hierarchical clustering clearly showed a total of 13 aging-related differential genes ($p < 0.05$). Also, we found significant differences in T and M staging and Stage between the two clusters (Figure 2C). Subsequently, we performed GO analysis. BP was mainly enriched in regulation of epithelial cell

differentiation, CC was mainly enriched in vesicle lumen and secretory granule lumen, while MF was mainly enriched in protein serine/threonine kinase inhibitor activity (Figure 2D). And the main pathways enriched by KEGG analysis are cell cycle, cellular senescence, central carbon metabolism in cancer, and HIF-1 signaling pathway (Figure 2E).

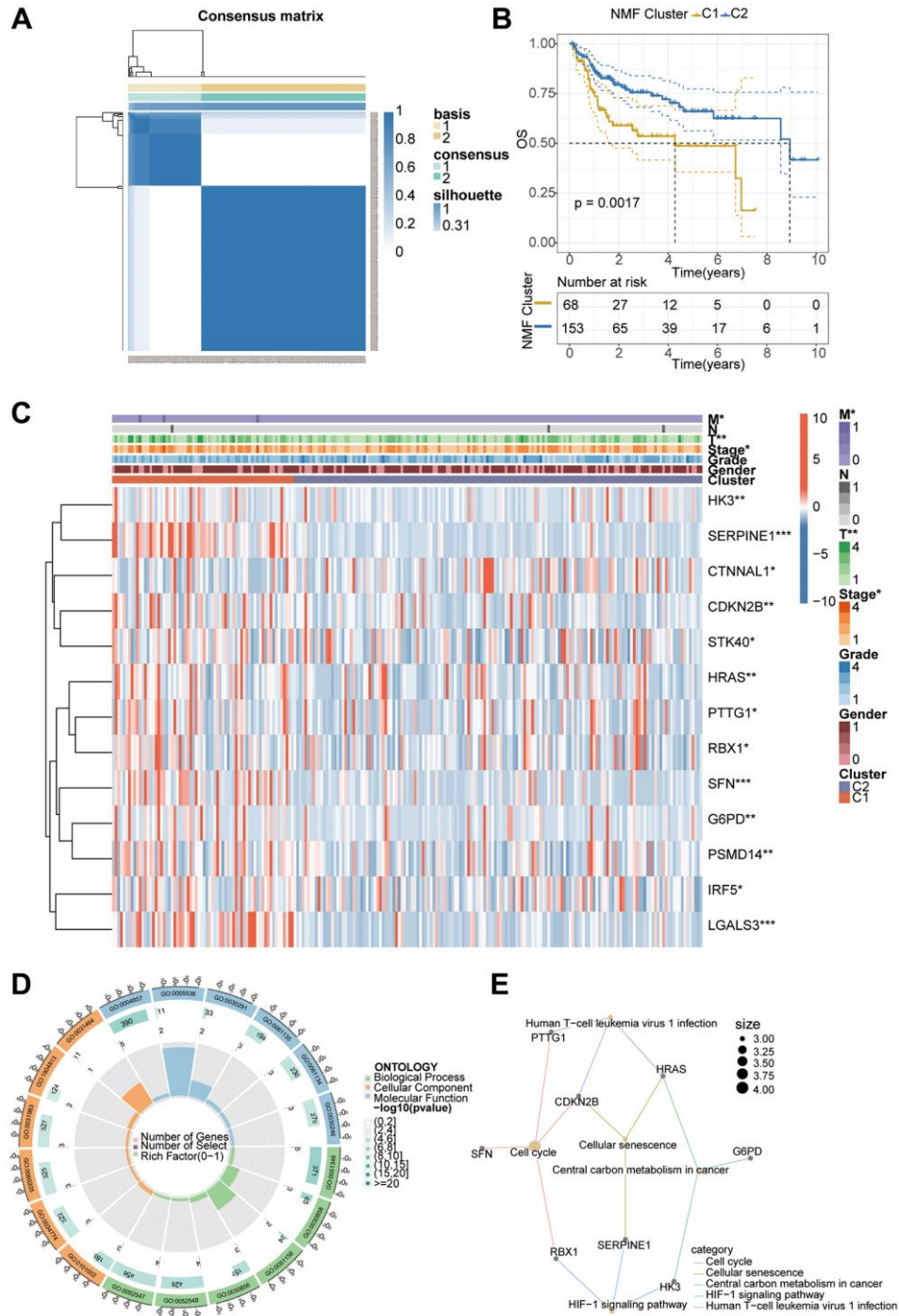


Figure 2. NMF staging and correlation analysis of patients with hepatocellular carcinoma. (A) Patients were classified into two clusters using the NMF algorithm. (B) Prognostic analysis revealed a poorer prognosis for patients with the C1 cluster. (C) Exploration of differential genes between different subtypes by differential analysis. (D, E) GO and KEGG analysis to investigate the underlying mechanisms and pathways.

Single-cell data analysis

We downloaded single-cell sequencing data (GSE146115) from the GEO database for liver cancer tissues, and collated the data for a total of 3200 single-cell comprehensive transcriptional profiles. The PCA and tSNE dimensionality reduction analysis of the samples allowed us to divide the samples into 12 clusters, and the subsequent heat map shows the

differential genes between the different clusters (Figure 3A, 3B). We then annotated the 12 clusters of cells and could find that the cells were clearly divided into four classes. In addition to the main component hepatocellular carcinoma cells, there are macrophages, T cells and NK cells (Figure 3C). Subsequently, the expression of typing difference genes between different cells in liver cancer tissues was analyzed. The scatter plot clearly shows that RBX1 is expressed in the highest

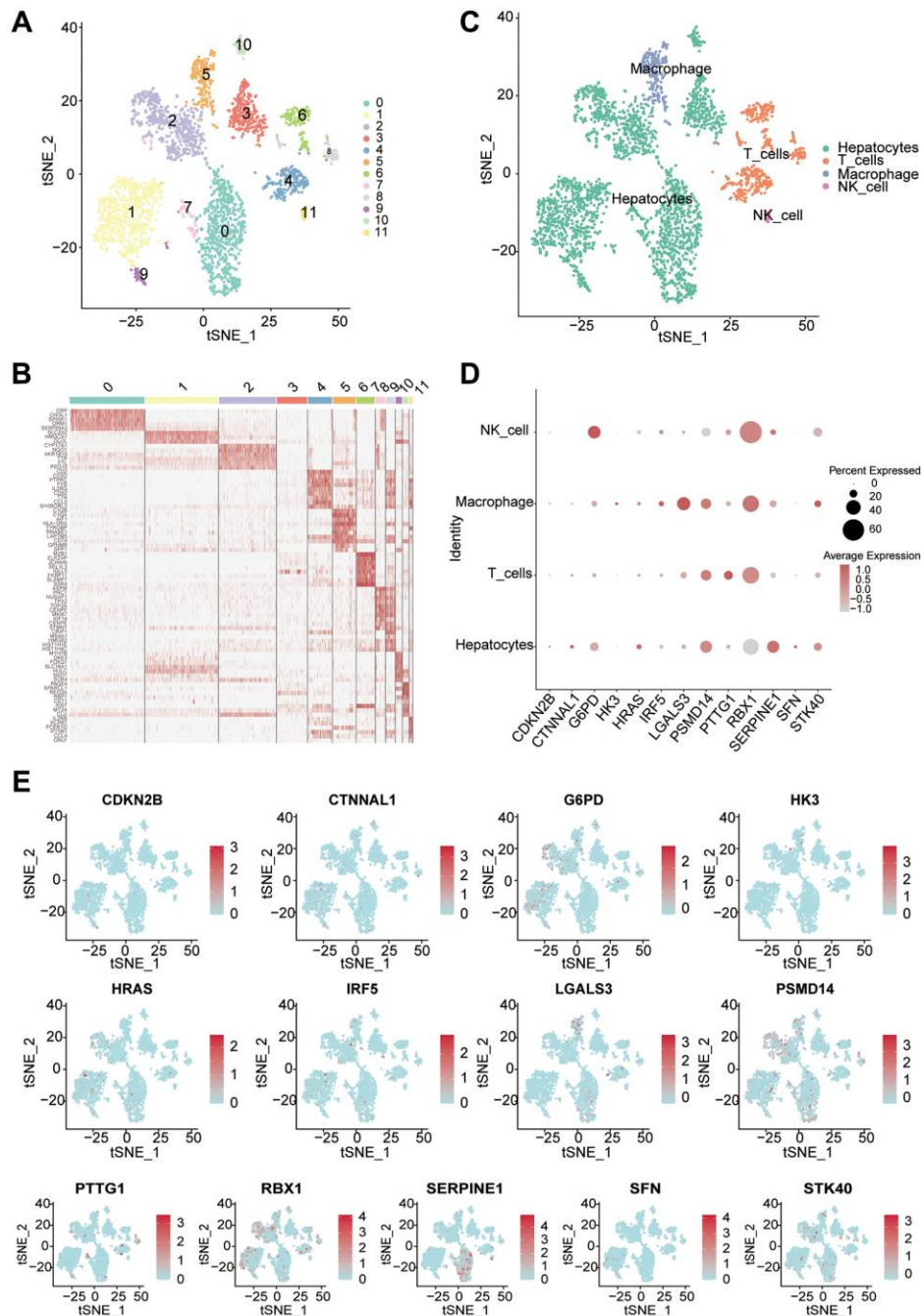


Figure 3. Single-cell analysis reveals the expression of differential genes in different cell types. (A) Using the tSNE algorithm to dimensionalize the samples into 12 clusters. (B) Heat map clearly showing the major differentially expressed genes in different clusters. (C) The Seurat package annotates different clusters with a total of 4 classes of cells. (D) The bubble diagram shows the expression of difference genes in different cells. (E) Visualization of 13 differential genes by single-cell sequencing.

amount in all cells. In addition, the expression of RBX1 was again significantly higher in NK cells than in other cells (Figure 3D, 3E).

Immunoassay of NMF clustering

To explore the immune cell landscape between different NMF clusters, we performed the analysis of immune infiltration by the Cibersort algorithm. First, we can see

how the various types of immune cells are different between the two clusters. Among them, B cell naive, plasma cells, and Tregs had significant differences between the two types, and the expression in C2 clusters was higher than that in C1 cluster (Figure 4A, 4B). The estimate algorithm allowed exploring the differences in immune scores, stromal scores and total scores between the two clusters, and we found that immune scores were significantly different between the two clusters

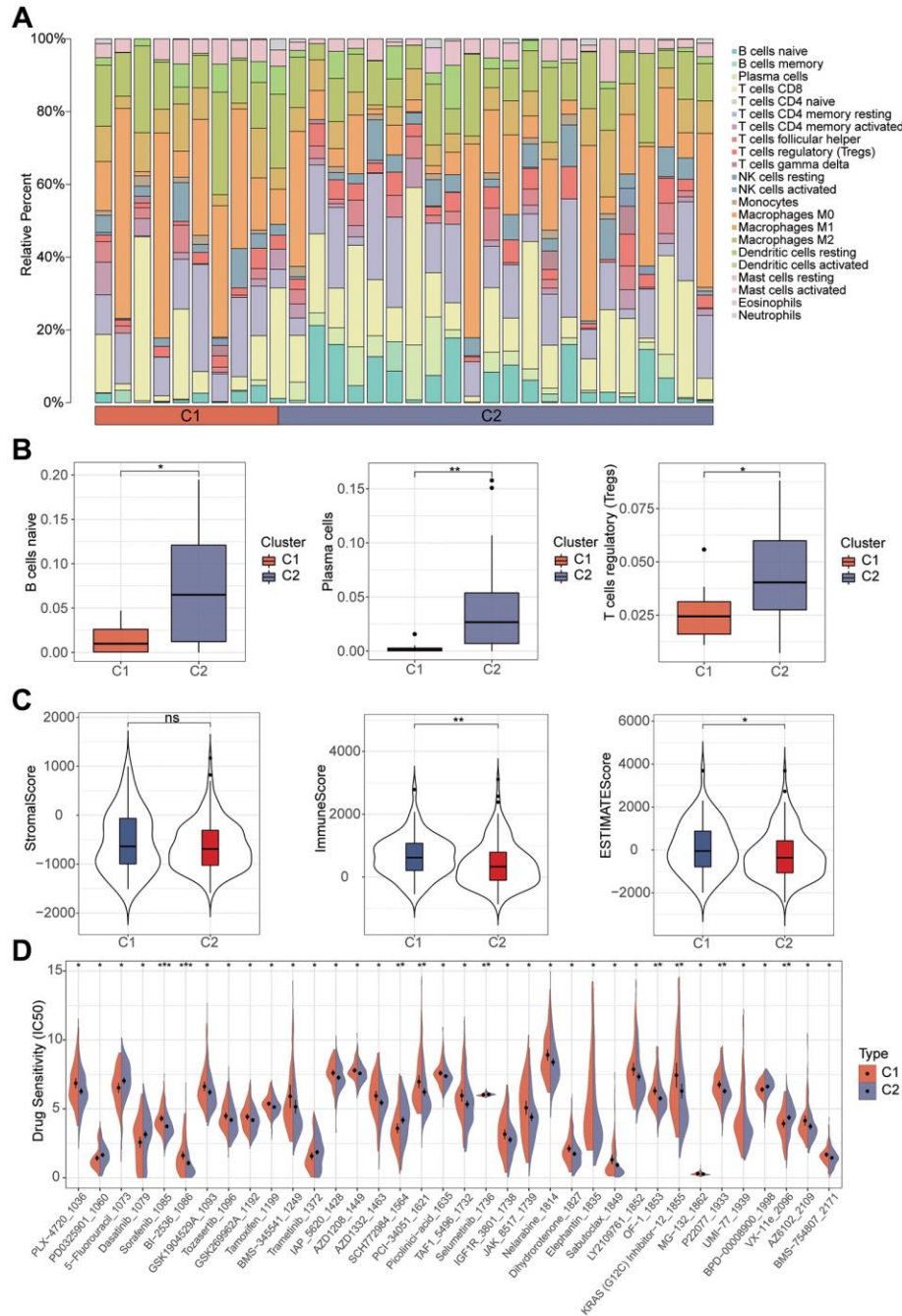


Figure 4. Immunoassay and drug sensitivity analysis of samples with different cluster. (A) The histogram shows the expression of immune cells between the different typologies. **(B)** Expression of B cell naive, Plasma cell and Tregs between different clusters. **(C)** Exploring the correlation of different clusters with the tumor microenvironment. **(D)** Analysis of the differences between the various chemotherapeutic agents in the different clusters.

(Figure 4C). We then investigated whether there were differences in drug sensitivity of chemotherapeutic drugs between the two clusters and plotted graphs for those indicators that were significant (Figure 4D).

Analysis of prognostic models

Considering that aging-related genes may be highly correlated with the prognosis of hepatocellular

carcinoma patients, we used these genes to construct a prognostic model. First, we clustered the aging genes by WGCNA, and we could see that the genes were divided into a total of 6 modules. Based on the *p*-values, we observed that the MEturquoise module exhibits the highest correlation with patient prognosis (Figure 5A, 5B). For differential and prognostic analysis of aging-related genes in this module, we screened a total of 38 hub genes. Through the string database, we explored the

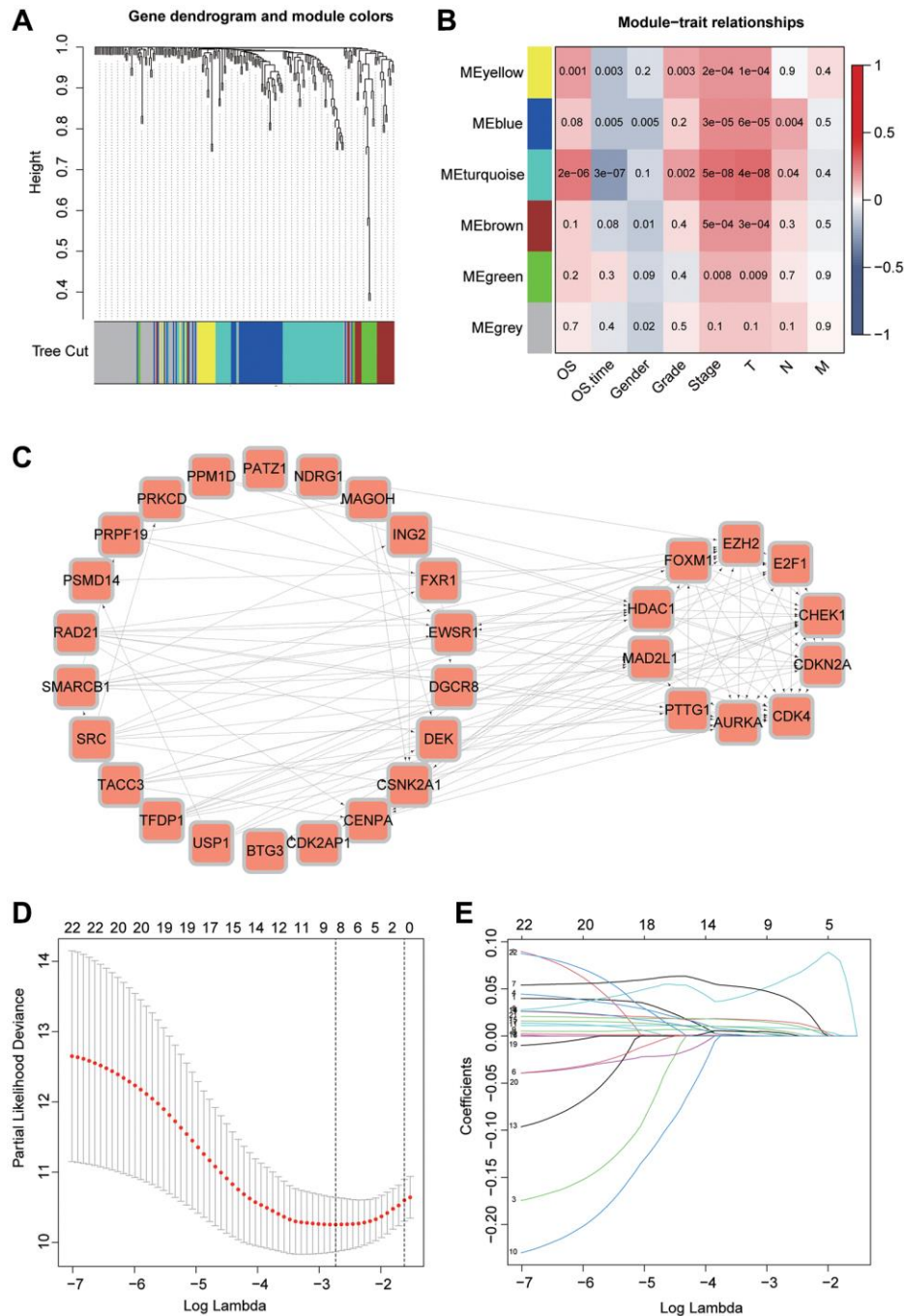


Figure 5. WGCNA combined with Lasso algorithm to construct a prognostic model. (A) Clustering of genes using the WGCNA algorithm. (B) Clinical and prognostic analysis of the genes in different modules. (C) Study of associations between genes of the MEturquoise module using the STRING database. (D, E) Selection of valuable genes by Lasso algorithm to construct prognostic models.

correlation between these genes (Figure 5C). We identified FXR1, NDRG1 and TFDP1 to construct the model by lasso analysis. The risks core of each sample was calculated according to Equation (Figure 5D, 5E). We validated the model using 3 datasets considering the correlation between this risk score and patient prognosis. The TCGA dataset is first divided into training and validation sets using the R package “caret” while GSE14520 is also used as the validation set. It can be seen that patients with high-risk scores in all three

datasets have a poor prognosis (Figure 6A). To further investigate the survival status of both groups, we found that patients in the high-risk group had a higher mortality rate. And the heat map revealed that the expression of all three key genes used for modeling were significantly higher in the high-risk group than in the low-risk group (Figure 6B). ROC curves can be seen for the TCGA training set model with AUC values of 0.797, 0.749 and 0.740 for 1, 2 and 3 years. While the validation set AUC values are 0.698, 0.626, and 0.627

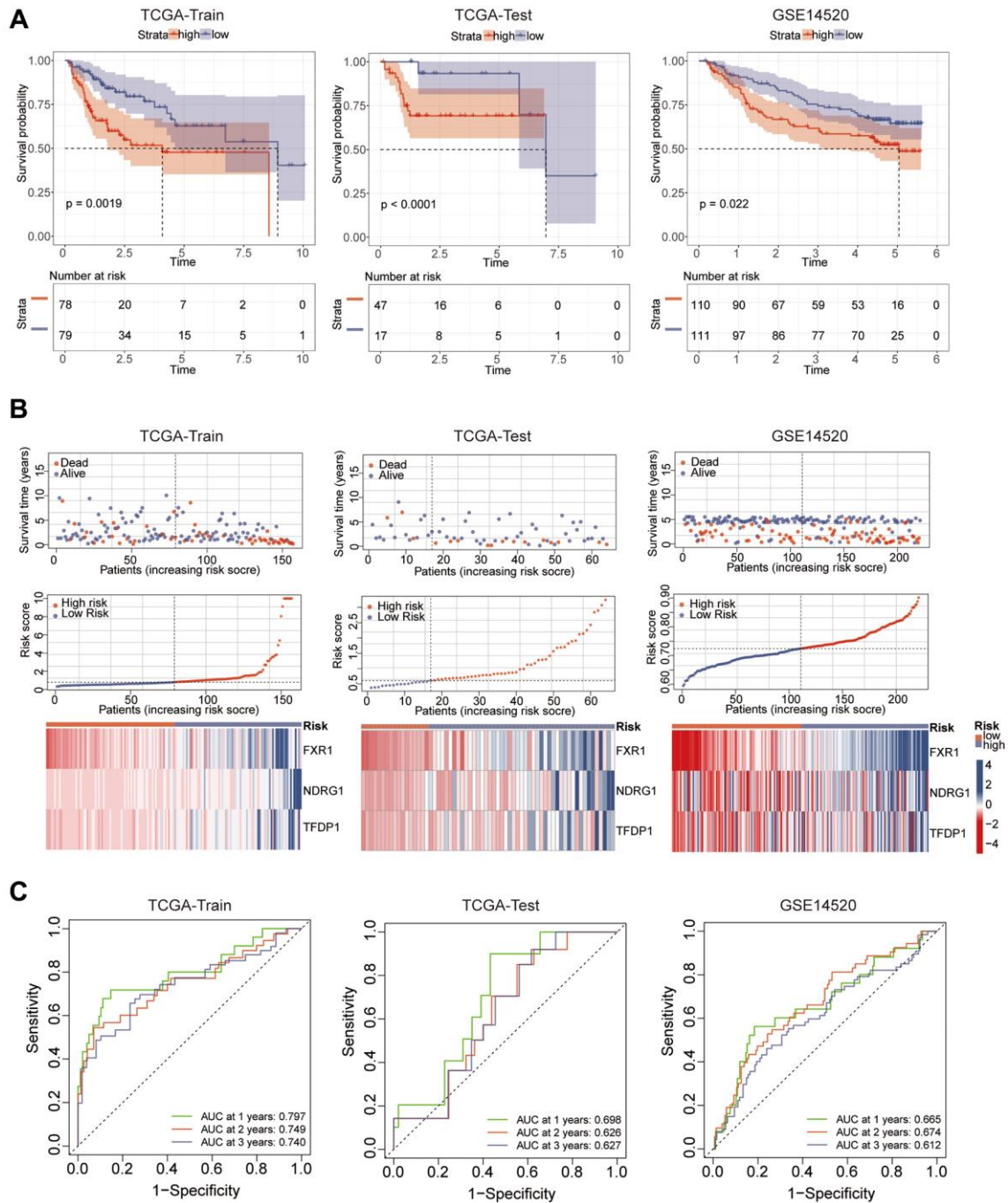


Figure 6. Prognostic analysis and model efficacy validation. (A) Prognosis between high and low risk groups was analyzed by KM curves. (B) Risk curves showing the differences between high and low risk groups. (C) ROC curves analyzing the specific efficacy of the model.

were significantly lower in the high-risk group than in the low-risk group (Figure 7B). There is growing evidence that increased expression of stem cell-associated biomarkers in tumor cells is highly correlated with drug resistance, cancer recurrence and tumor proliferation [21]. Our study found that this risk score was positively correlated with the tumor stemness score (Figure 7C). To understand the differences in immunotherapy between high and low risk groups, we predicted whether risk scores were associated with immunotherapy by MMR and immune checkpoint analysis. MMR were all positively correlated with risk scores, with MSH2 having the highest correlation (Figure 7D). The immune checkpoint analysis also found several indicators correlated with risk score, NRP1, TNFSF4, TNFSF15, TNFSF18, CD276, CD80 and HHLA2 were strongly correlated with risk score ($P < 0.001$). However, PD1, PDL1, and CTLA4 did not show correlation (Figure 7E). Subsequently, we verified with the dataset that risk score was a better predictor of the effectiveness of immunotherapy and that patients in the high-risk group had better immunotherapy outcomes (Figure 7F, 7G). To study the expression of key genes for constructing the model, we analyzed the expression of TFDP1, NDRG1 and FXR1 by single cell sequencing. Among them, TFDP1 expression was low in all four types of cells, while FXR1 was mainly expressed in NK cells. NDRG1 was highly expressed in hepatocellular carcinoma cells, macrophages and NK cells (Figure 8A, 8B). Subsequently, we investigated the sensitivity of high and low risk groups to different chemotherapeutic drugs. Interestingly, we found significant differences in drug sensitivity between the two groups for Camptothecin, Cisplatin, Gemcitabine, Irinotecan, Oxaliplatin and Vinblastine, but not for sorafenib and 5-Fluorouracil (Figure 8C).

Experimental validation of key genes

We used the GSE39791 dataset for validation and could see that the expression of TFDP1, NDRG1 and FXR1 in liver cancer tissues was significantly higher than that in normal tissues (Figure 9A). Subsequently, we performed experimental validation using LO2 normal liver cells as well as the HEPG2, BEL-7402, and HCC-LM3 liver cancer cell lines. And the expression of the three key genes in different hepatocellular carcinoma cells was higher than that in normal hepatocytes (Figure 9B). Finally, IHC results showed that the expression of FXR1 and NDRG1 was significantly higher in liver cancer tissues than in normal tissues (Figure 9C).

DISCUSSION

With the development of immunotherapy, new treatment options are bringing hope to liver cancer

patients. Immunotherapy works by enhancing the body's immune system to attack tumor cells, and unlike traditional treatment methods, it does not destroy normal cells, thus reducing many side effects [22]. One common immunotherapy method is the use of immune checkpoint inhibitors. Immune checkpoints are proteins that can help the body's immune system recognize and attack tumor cells. However, certain tumor cells can use immune checkpoints to evade immune system attacks, leading to tumor growth and spread. The role of immune checkpoint inhibitors is to block these immune checkpoints, making tumor cells unable to escape immune attacks [23]. In the treatment of liver cancer, the use of immune checkpoint inhibitors has made some progress. Clinical studies have shown that some immune checkpoint inhibitors can enhance T cell immune responses, helping patients suppress tumor growth and spread, while also prolonging survival and improving quality of life [24].

This study first performed NMF classification using TCGA liver cancer transcriptome database, and the results showed that the samples were well divided into two clusters. KM analysis showed a significant difference in prognosis between the two clusters, and we then analyzed the potential mechanisms through GO and KEGG analyses. The MF analysis mainly enriched in protein serine/threonine kinase inhibitor activity, which has been found to be closely related to the occurrence and development of liver cancer in previous studies [25, 26]. Interesting KEGG pathway enrichment included Cell cycle, Cellular senescence, and HIF-1 signaling pathway. The disruption of the cell cycle is closely related to the occurrence and development of liver cancer, and HIF-1 abnormal activation plays an important role in the development of liver cancer, including promoting tumor cell growth, metabolism, and immune evasion [27, 28]. Considering that aging-related genes may be related to the prognosis of patients, we used the WGCNA algorithm to select an aging gene module that was correlated with patient prognosis. Then, we used the aging-related genes in this module to construct a lasso model, and the selected genes mainly included TFDP1, NDRG1, and FXR1. Previous studies have demonstrated that overexpression of TFDP1 can promote tumor cell growth, thereby accelerating the progression and deterioration of certain liver cancers [29]. NDRG1 can enhance the interaction between fibroblasts and tumor cells, leading to the development of hepatocellular carcinoma [30, 31]. FXR1 can promote the proliferation, invasion, and migration of hepatocellular carcinoma, and its action is mediated by Smad2/3 [32]. These studies provide additional evidence for the reliability of the genes we screened and are consistent with our predicted results. Afterwards, we performed survival analysis of the model using three

datasets and found that patients in the high-risk group had significantly worse prognosis than those in the low-risk group. Additionally, the model demonstrated good prediction performance at 1, 2, and 3 years based on the ROC curves.

Furthermore, we used the Cibersort algorithm to explore the correlation between the model risk score and 21 immune cell types. We found a positive correlation

between the risk score and B cell memory and macrophages M0, and a negative correlation with B cell naive and plasma cell. Previous studies have shown that a decrease in B cell naive is closely related to the occurrence and prognosis of liver cancer, which may be associated with immune escape and tolerance in the liver cancer microenvironment [33]. The quantity and function of plasma cells in liver cancer patients may also affect the immune status and treatment

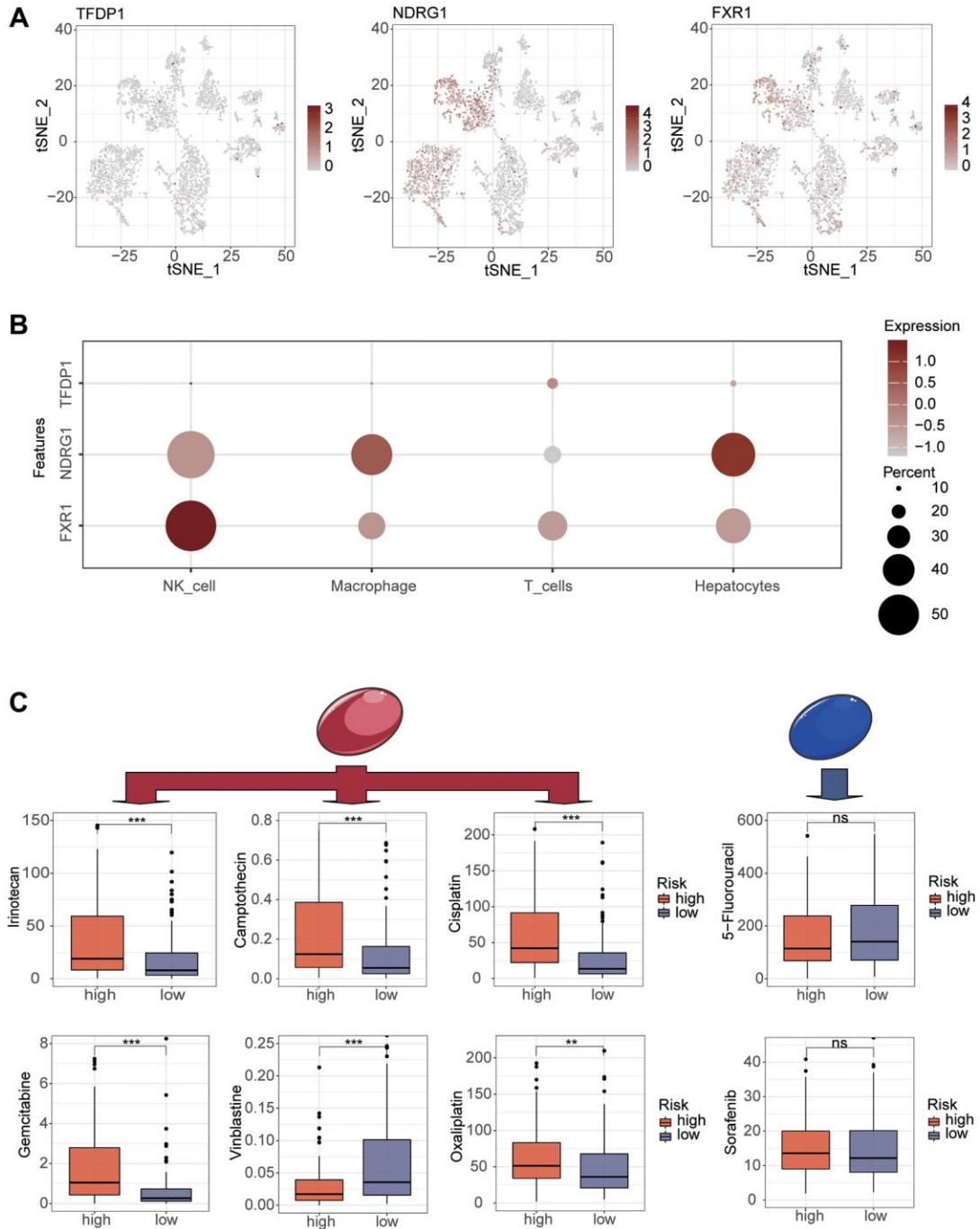


Figure 8. Single-cell analysis of modeled key genes and drug sensitivity analysis of different risk groups. (A, B) Single-cell analysis reveals the expression of modeling key genes in different cells. **(C)** Use the oncopredict package to explore chemotherapeutic agents that are significantly different between high and low risk groups.

effectiveness [34]. Tumor stem cells are tumor cells with stem cell-like characteristics. Studies have shown that the existence and characteristics of liver cancer stem cells make liver cancer highly recurrent and resistant to treatment [35]. The stemness index is an indicator of the similarity between tumor cells and stem cells. Our study found a good correlation between this

risk score and tumor stemness, demonstrating that the score can predict the degree of tumor stemness to a certain extent. This has important implications for subsequent treatment. Immunotherapy is an emerging cancer treatment method, but researchers have found that its efficacy varies greatly among different solid tumor patients [22, 23]. Therefore, we explored the

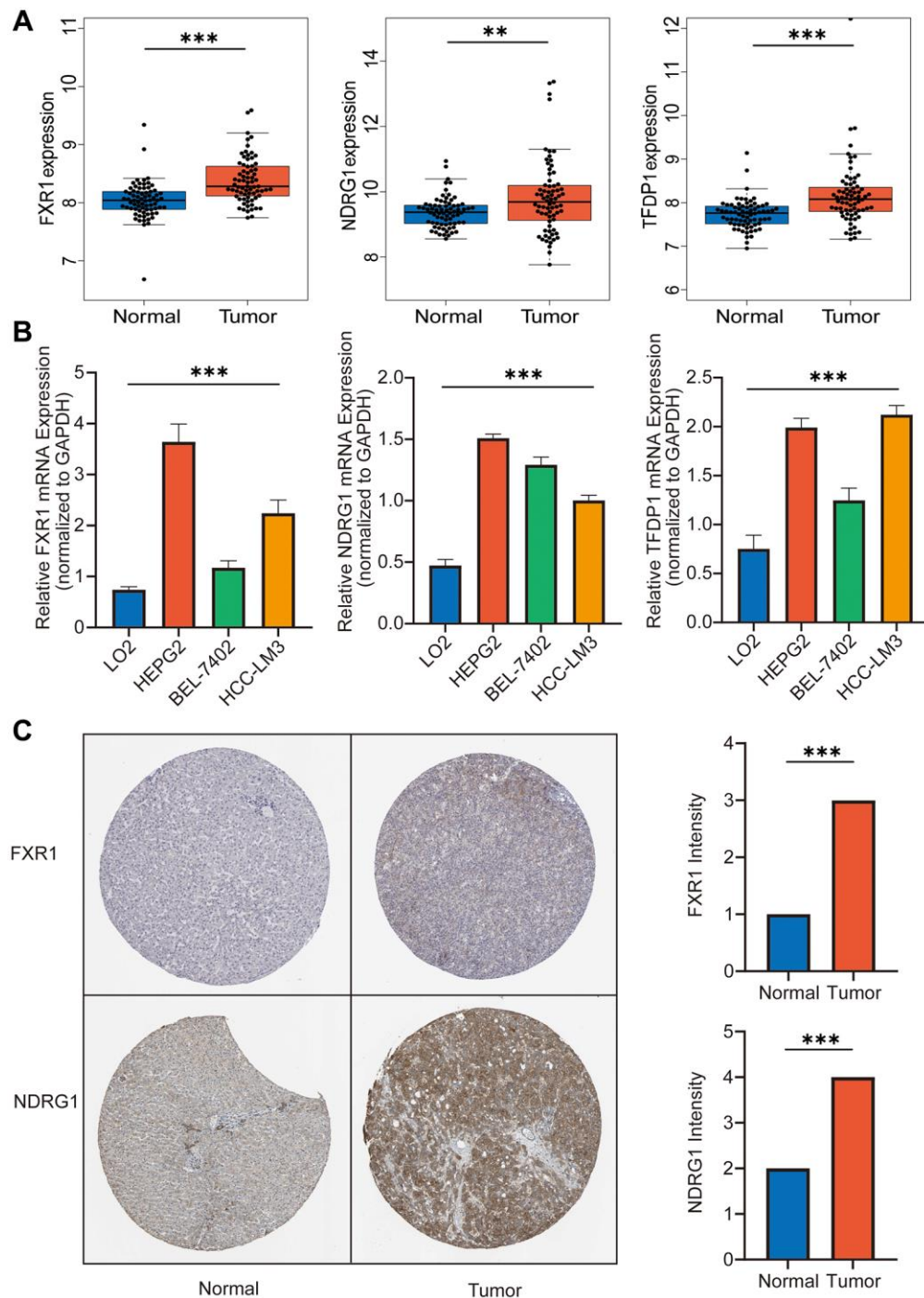


Figure 9. Experiment *in vitro* and *in vivo*. (A) Expression of FXR1, NDRG1 and TFDP1 was significantly higher in hepatocellular carcinoma tissues than in normal liver tissues by GEO dataset. (B) The expression of FXR1, NDRG1 and TFDP1 was found to be significantly higher in three types of hepatocellular carcinoma cells than in normal hepatocytes by RT-qPCR assay. (C) The expression of FXR1 and NDRG1 was found to be significantly higher in hepatocellular carcinoma tissues than in normal liver tissues using the HPA database.

correlation between this risk score and immunotherapy. MMR is a predictive indicator of immunotherapy, and our study found that the correlation between the risk score and indicators such as EPCAM, MLH1, MSH2, MSH6, and PMS2 was extremely high. Among them, the correlation between MSH2 and risk score was the highest, and studies have shown that mutations and abnormal expression of the MSH2 gene are closely related to the occurrence and development of various cancers [36]. Subsequently, we further verified the effectiveness of immunotherapy in patients with different risk scores using data sets, and found that patients with higher risk scores had better responses to immunotherapy. Single-cell sequencing technology can reveal the heterogeneity and evolutionary trajectories of different subclones within a tumor, which helps to deepen our understanding of the molecular mechanisms underlying cancer initiation and progression [37]. In order to further explore the expression patterns of key molecules in different cell types, we performed single-cell analysis using the GSE146115 dataset. We found that FXR1 had the highest expression level in NK cells. According to Zhang et al., natural killer cells (NK cells) play an important role in liver cancer immunotherapy, and their activity and infiltration level are closely related to the clinical prognosis of liver cancer [34]. NDRG1 was primarily expressed in liver cells. As for TFDP1, its expression levels were relatively low in various cell types, and it was mainly expressed in T cells. Studies have shown that T cells play an important role in tumor immunotherapy, and their immune surveillance and killing abilities are closely related to tumor progression and prognosis [22]. Subsequently, we performed drug sensitivity analysis and found that Camptothecin, Cisplatin, Gemcitabine, Irinotecan, Oxaliplatin and Vinblastine showed significant differences between the high risk and low risk groups. This has significant clinical implications for our diagnosis and treatment. Finally, through RT-qPCR experiments, we found that the expression levels of key genes involved in the construction of the model were higher in different liver cancer cells than in normal liver cells. IHC validation confirmed the expression patterns of these key genes in cancer tissue and normal tissue.

Certainly, our study has several limitations that need to be acknowledged. Firstly, we utilized multiple public databases for our joint analysis, but some of these databases lack clinical and immunotherapy data, which may result in certain omissions in our analysis. We plan to conduct further prospective studies to collect samples and data from our own hospital to conduct more in-depth research. Secondly, downstream mechanisms of the genes used to construct our model were not explored, which may lead to some bias in our prediction

of targeted drugs. Further research is needed to address this issue.

In summary, our study has important clinical implications. The results obtained through the integration of bulk and single-cell sequencing data from multiple datasets are more reliable. The risk score constructed using WGCNA and LASSO can serve as a reliable and independent biomarker for predicting the prognosis of liver cancer patients. Single-cell sequencing analysis can help us further explore the expression patterns of hub genes in different cells. Functional enrichment analysis can assist in mechanism exploration and downstream analysis. In addition, our study explores the correlation between MMR and immune checkpoints through classification and model construction, which is helpful in assessing the effectiveness of immunotherapy. Furthermore, the difference in drug sensitivity between high and low risk groups determined by the risk model is useful for developing personalized chemotherapy regimens for patients.

CONCLUSION

In summary, we have constructed an aging-related model, which we hope can serve as a reference for predicting patient survival and guiding liver cancer-related treatments.

AUTHOR CONTRIBUTIONS

JS and YS have constructed and devised the research. HG and BL performed data analysis and wrote the manuscript. BL and YH acquired samples and performed the experiments. All authors reviewed the manuscript.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest related to this study.

FUNDING

No funding was provided to this study.

REFERENCES

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2018; 68:394–424.
<https://doi.org/10.3322/caac.21492>
PMID:[30207593](https://pubmed.ncbi.nlm.nih.gov/30207593/)

2. Gilles H, Garbutt T, Landrum J. Hepatocellular Carcinoma. *Crit Care Nurs Clin North Am.* 2022; 34:289–301.
<https://doi.org/10.1016/j.cnc.2022.04.004>
PMID:[36049848](https://pubmed.ncbi.nlm.nih.gov/36049848/)
3. Johnson P, Zhou Q, Dao DY, Lo YMD. Circulating biomarkers in the diagnosis and management of hepatocellular carcinoma. *Nat Rev Gastroenterol Hepatol.* 2022; 19:670–81.
<https://doi.org/10.1038/s41575-022-00620-y>
PMID:[35676420](https://pubmed.ncbi.nlm.nih.gov/35676420/)
4. Piñero F, Dirchwolf M, Pessôa MG. Biomarkers in Hepatocellular Carcinoma: Diagnosis, Prognosis and Treatment Response Assessment. *Cells.* 2020; 9:1370.
<https://doi.org/10.3390/cells9061370>
PMID:[32492896](https://pubmed.ncbi.nlm.nih.gov/32492896/)
5. He Q, Yang J, Jin Y. Immune infiltration and clinical significance analyses of the coagulation-related genes in hepatocellular carcinoma. *Brief Bioinform.* 2022; 23:bbac291.
<https://doi.org/10.1093/bib/bbac291>
PMID:[35849048](https://pubmed.ncbi.nlm.nih.gov/35849048/)
6. Xue S, Lu F, Sun C, Zhao J, Zhen H, Li X. LncRNA ZEB1-AS1 regulates hepatocellular carcinoma progression by targeting miR-23c. *World J Surg Oncol.* 2021; 19:121.
<https://doi.org/10.1186/s12957-021-02176-8>
PMID:[33865414](https://pubmed.ncbi.nlm.nih.gov/33865414/)
7. Lee W, Kim SJ. Protective effects of isoflavones on alcoholic liver diseases: Computational approaches to investigate the inhibition of ALDH2 with isoflavone analogues. *Front Mol Biosci.* 2023; 10:1147301.
<https://doi.org/10.3389/fmolb.2023.1147301>
PMID:[36923641](https://pubmed.ncbi.nlm.nih.gov/36923641/)
8. Cervello M, Emma MR, Augello G, Cusimano A, Giannitrapani L, Soresi M, Akula SM, Abrams SL, Steelman LS, Gulino A, Belmonte B, Montalto G, McCubrey JA. New landscapes and horizons in hepatocellular carcinoma therapy. *Aging (Albany NY).* 2020; 12:3053–94.
<https://doi.org/10.18632/aging.102777>
PMID:[32018226](https://pubmed.ncbi.nlm.nih.gov/32018226/)
9. Anwanwan D, Singh SK, Singh S, Saikam V, Singh R. Challenges in liver cancer and possible treatment approaches. *Biochim Biophys Acta Rev Cancer.* 2020; 1873:188314.
<https://doi.org/10.1016/j.bbcan.2019.188314>
PMID:[31682895](https://pubmed.ncbi.nlm.nih.gov/31682895/)
10. Bruix J, Han KH, Gores G, Llovet JM, Mazzaferro V. Liver cancer: Approaching a personalized care. *J Hepatol.* 2015; 62:S144–56.
<https://doi.org/10.1016/j.jhep.2015.02.007>
PMID:[25920083](https://pubmed.ncbi.nlm.nih.gov/25920083/)
11. Huang A, Yang XR, Chung WY, Dennison AR, Zhou J. Targeted therapy for hepatocellular carcinoma. *Signal Transduct Target Ther.* 2020; 5:146.
<https://doi.org/10.1038/s41392-020-00264-x>
PMID:[32782275](https://pubmed.ncbi.nlm.nih.gov/32782275/)
12. Wang K, Wang C, Jiang H, Zhang Y, Lin W, Mo J, Jin C. Combination of Ablation and Immunotherapy for Hepatocellular Carcinoma: Where We Are and Where to Go. *Front Immunol.* 2021; 12:792781.
<https://doi.org/10.3389/fimmu.2021.792781>
PMID:[34975896](https://pubmed.ncbi.nlm.nih.gov/34975896/)
13. Oura K, Morishita A, Tani J, Masaki T. Tumor Immune Microenvironment and Immunosuppressive Therapy in Hepatocellular Carcinoma: A Review. *Int J Mol Sci.* 2021; 22:5801.
<https://doi.org/10.3390/ijms22115801>
PMID:[34071550](https://pubmed.ncbi.nlm.nih.gov/34071550/)
14. Hu D, Yuan S, Zhong J, Liu Z, Wang Y, Liu L, Li J, Wen F, Liu J, Zhang J. Cellular senescence and hematological malignancies: From pathogenesis to therapeutics. *Pharmacol Ther.* 2021; 223:107817.
<https://doi.org/10.1016/j.pharmthera.2021.107817>
PMID:[33587950](https://pubmed.ncbi.nlm.nih.gov/33587950/)
15. He S, Sharpless NE. Senescence in Health and Disease. *Cell.* 2017; 169:1000–11.
<https://doi.org/10.1016/j.cell.2017.05.015>
PMID:[28575665](https://pubmed.ncbi.nlm.nih.gov/28575665/)
16. Calcinotto A, Kohli J, Zagato E, Pellegrini L, Demaria M, Alimonti A. Cellular Senescence: Aging, Cancer, and Injury. *Physiol Rev.* 2019; 99:1047–78.
<https://doi.org/10.1152/physrev.00020.2018>
PMID:[30648461](https://pubmed.ncbi.nlm.nih.gov/30648461/)
17. Alimirah F, Pulido T, Valdovinos A, Alptekin S, Chang E, Jones E, Diaz DA, Flores J, Velarde MC, Demaria M, Davalos AR, Wiley CD, Limbad C, et al. Cellular Senescence Promotes Skin Carcinogenesis through p38MAPK and p44/42MAPK Signaling. *Cancer Res.* 2020; 80:3606–19.
<https://doi.org/10.1158/0008-5472.CAN-20-0108>
PMID:[32641409](https://pubmed.ncbi.nlm.nih.gov/32641409/)
18. Rielland M, Cantor DJ, Graveline R, Hajdu C, Mara L, de Diego Diaz B, Miller G, David G. Senescence-associated SIN3B promotes inflammation and pancreatic cancer progression. *J Clin Invest.* 2014; 124:2125–35.
<https://doi.org/10.1172/JCI72619>
PMID:[24691445](https://pubmed.ncbi.nlm.nih.gov/24691445/)
19. Liu B, Zhou Z, Jin Y, Lu J, Feng D, Peng R, Sun H, Mu X, Li C, Chen Y. Hepatic stellate cell activation and senescence induced by intrahepatic microbiota disturbances drive progression of liver cirrhosis toward hepatocellular carcinoma. *J Immunother Cancer.* 2022; 10:e003069.
<https://doi.org/10.1136/jitc-2021-003069>

- PMID:[34996812](https://pubmed.ncbi.nlm.nih.gov/34996812/)
20. Shen J, Sun W, Liu J, Li J, Li Y, Gao Y. Metabolism-related signatures is correlated with poor prognosis and immune infiltration in hepatocellular carcinoma *via* multi-omics analysis and basic experiments. *Front Oncol.* 2023; 13:1130094.
<https://doi.org/10.3389/fonc.2023.1130094>
PMID:[36860325](https://pubmed.ncbi.nlm.nih.gov/36860325/)
 21. Xiong J, Yan L, Zou C, Wang K, Chen M, Xu B, Zhou Z, Zhang D. Integrins regulate stemness in solid tumor: an emerging therapeutic target. *J Hematol Oncol.* 2021; 14:177.
<https://doi.org/10.1186/s13045-021-01192-1>
PMID:[34715893](https://pubmed.ncbi.nlm.nih.gov/34715893/)
 22. Kohrt HE, Tumei PC, Benson D, Bhardwaj N, Brody J, Formenti S, Fox BA, Galon J, June CH, Kalos M, Kirsch I, Kleen T, Kroemer G, et al, and Cancer Immunotherapy Trials Network (CITN). Immunodynamics: a cancer immunotherapy trials network review of immune monitoring in immuno-oncology clinical trials. *J Immunother Cancer.* 2016; 4:15.
<https://doi.org/10.1186/s40425-016-0118-0>
PMID:[26981245](https://pubmed.ncbi.nlm.nih.gov/26981245/)
 23. O'Donnell JS, Teng MWL, Smyth MJ. Cancer immunoediting and resistance to T cell-based immunotherapy. *Nat Rev Clin Oncol.* 2019; 16:151–67.
<https://doi.org/10.1038/s41571-018-0142-8>
PMID:[30523282](https://pubmed.ncbi.nlm.nih.gov/30523282/)
 24. Raskov H, Orhan A, Christensen JP, Gögenur I. Cytotoxic CD8⁺ T cells in cancer and cancer immunotherapy. *Br J Cancer.* 2021; 124:359–67.
<https://doi.org/10.1038/s41416-020-01048-4>
PMID:[32929195](https://pubmed.ncbi.nlm.nih.gov/32929195/)
 25. Li W, Xiao J, Zhou X, Xu M, Hu C, Xu X, Lu Y, Liu C, Xue S, Nie L, Zhang H, Li Z, Zhang Y, et al. STK4 regulates TLR pathways and protects against chronic inflammation-related hepatocellular carcinoma. *J Clin Invest.* 2015; 125:4239–54.
<https://doi.org/10.1172/JCI81203>
PMID:[26457732](https://pubmed.ncbi.nlm.nih.gov/26457732/)
 26. Llovet JM, Pinyol R, Kelley RK, El-Khoueiry A, Reeves HL, Wang XW, Gores GJ, Villanueva A. Molecular pathogenesis and systemic therapies for hepatocellular carcinoma. *Nat Cancer.* 2022; 3:386–401.
<https://doi.org/10.1038/s43018-022-00357-2>
PMID:[35484418](https://pubmed.ncbi.nlm.nih.gov/35484418/)
 27. Zhang RY, Liu ZK, Wei D, Yong YL, Lin P, Li H, Liu M, Zheng NS, Liu K, Hu CX, Yang XZ, Chen ZN, Bian H. UBE2S interacting with TRIM28 in the nucleus accelerates cell cycle by ubiquitination of p27 to promote hepatocellular carcinoma development. *Signal Transduct Target Ther.* 2021; 6:64.
<https://doi.org/10.1038/s41392-020-00432-z>
PMID:[33589597](https://pubmed.ncbi.nlm.nih.gov/33589597/)
 28. Hu W, Zheng S, Guo H, Dai B, Ni J, Shi Y, Bian H, Li L, Shen Y, Wu M, Tian Z, Liu G, Hossain MA, et al. PLAGL2-EGFR-HIF-1/2 α Signaling Loop Promotes HCC Progression and Erlotinib Insensitivity. *Hepatology.* 2021; 73:674–91.
<https://doi.org/10.1002/hep.31293>
PMID:[32335942](https://pubmed.ncbi.nlm.nih.gov/32335942/)
 29. Yasui K, Okamoto H, Arai S, Inazawa J. Association of over-expressed TFDP1 with progression of hepatocellular carcinomas. *J Hum Genet.* 2003; 48:609–13.
<https://doi.org/10.1007/s10038-003-0086-3>
PMID:[14618416](https://pubmed.ncbi.nlm.nih.gov/14618416/)
 30. Luo Q, Wang CQ, Yang LY, Gao XM, Sun HT, Zhang Y, Zhang KL, Zhu Y, Zheng Y, Sheng YY, Lu L, Jia HL, Yu WQ, et al. FOXQ1/NDRG1 axis exacerbates hepatocellular carcinoma initiation via enhancing crosstalk between fibroblasts and tumor cells. *Cancer Lett.* 2018; 417:21–34.
<https://doi.org/10.1016/j.canlet.2017.12.021>
PMID:[29248714](https://pubmed.ncbi.nlm.nih.gov/29248714/)
 31. Cheng J, Xie HY, Xu X, Wu J, Wei X, Su R, Zhang W, Lv Z, Zheng S, Zhou L. NDRG1 as a biomarker for metastasis, recurrence and of poor prognosis in hepatocellular carcinoma. *Cancer Lett.* 2011; 310:35–45.
<https://doi.org/10.1016/j.canlet.2011.06.001>
PMID:[21763068](https://pubmed.ncbi.nlm.nih.gov/21763068/)
 32. Zhao K, Gao J, Shi J, Shi C, Pang C, Li J, Guo W, Zhang S. FXR1 promotes proliferation, invasion and migration of hepatocellular carcinoma *in vitro* and *in vivo*. *Oncol Lett.* 2023; 25:22.
<https://doi.org/10.3892/ol.2022.13608>
PMID:[36466996](https://pubmed.ncbi.nlm.nih.gov/36466996/)
 33. Zhang Z, Ma L, Goswami S, Ma J, Zheng B, Duan M, Liu L, Zhang L, Shi J, Dong L, Sun Y, Tian L, Gao Q, Zhang X. Landscape of infiltrating B cells and their clinical significance in human hepatocellular carcinoma. *Oncoimmunology.* 2019; 8:e1571388.
<https://doi.org/10.1080/2162402X.2019.1571388>
PMID:[30906667](https://pubmed.ncbi.nlm.nih.gov/30906667/)
 34. Zhang PF, Gao C, Huang XY, Lu JC, Guo XJ, Shi GM, Cai JB, Ke AW. Cancer cell-derived exosomal circUHRF1 induces natural killer cell exhaustion and may cause resistance to anti-PD1 therapy in hepatocellular carcinoma. *Mol Cancer.* 2020; 19:110.
<https://doi.org/10.1186/s12943-020-01222-5>
PMID:[32593303](https://pubmed.ncbi.nlm.nih.gov/32593303/)
 35. Liu YC, Yeh CT, Lin KH. Cancer Stem Cell Functions in

Hepatocellular Carcinoma and Comprehensive Therapeutic Strategies. *Cells*. 2020; 9:1331.

<https://doi.org/10.3390/cells9061331>

PMID:[32466488](https://pubmed.ncbi.nlm.nih.gov/32466488/)

36. Bonadona V, Bonaïti B, Olschwang S, Grandjouan S, Huiart L, Longy M, Guimbaud R, Buecher B, Bignon YJ, Caron O, Colas C, Noguès C, Lejeune-Dumoulin S, et al, and French Cancer Genetics Network. Cancer risks associated with germline mutations in MLH1, MSH2, and MSH6 genes in Lynch syndrome. *JAMA*. 2011; 305:2304–10.

<https://doi.org/10.1001/jama.2011.743>

PMID:[21642682](https://pubmed.ncbi.nlm.nih.gov/21642682/)

37. Navin N, Kendall J, Troge J, Andrews P, Rodgers L, McIndoo J, Cook K, Stepansky A, Levy D, Esposito D, Muthuswamy L, Krasnitz A, McCombie WR, et al. Tumour evolution inferred by single-cell sequencing. *Nature*. 2011; 472:90–4.

<https://doi.org/10.1038/nature09807>

PMID:[21399628](https://pubmed.ncbi.nlm.nih.gov/21399628/)