

RESEARCH ARTICLE



CDT1 is a Potential Therapeutic Target for the Progression of NAFLD to HCC and the Exacerbation of Cancer



Xingyu He^{1, #}, Jun Ma^{1, #}, Xue Yan¹, Xiangyu Yang², Ping Wang², Lijie Zhang¹, Na Li¹ and Zheng Shi^{1, *}

¹Clinical Medical College & Affiliated Hospital of Chengdu University, Chengdu University, 610083, P.R. China;

²West China Hospital, Sichuan University, 610083, P.R. China

Abstract: Aims: This study aimed to identify potential therapeutic targets in the progression from non-alcoholic fatty liver disease (NAFLD) to hepatocellular carcinoma (HCC), with a focus on genes that could influence disease development and progression.

Background: Hepatocellular carcinoma, significantly driven by non-alcoholic fatty liver disease, represents a major global health challenge due to late-stage diagnosis and limited treatment options. This study utilized bioinformatics to analyze data from GEO and TCGA, aiming to uncover molecular biomarkers that bridge NAFLD to HCC. Through identifying critical genes and pathways, our research seeks to advance early detection and develop targeted therapies, potentially improving prognosis and personalizing treatment for NAFLD-HCC patients.

Objectives: Identify key genes that differ between NAFLD and HCC; Analyze these genes to understand their roles in disease progression; Validate the functions of these genes in NAFLD to HCC transition.

Methods: Initially, we identified a set of genes differentially expressed in both NAFLD and HCC using second-generation sequencing data from the GEO and TCGA databases. We then employed a Cox proportional hazards model and a Lasso regression model, applying machine learning techniques to the large sample data from TCGA. This approach was used to screen for key disease-related genes, and an external dataset was utilized for model validation. Additionally, pseudo-temporal sequencing analysis of single-cell sequencing data was performed to further examine the variations in these genes in NAFLD and HCC.

Results: The machine learning analysis identified IGSF3, CENPW, CDT1, and CDC6 as key genes. Furthermore, constructing a machine learning model for CDT1 revealed it to be the most critical gene, with model validation yielding an ROC value greater than 0.80. The single-cell sequencing data analysis confirmed significant variations in the four predicted key genes between the NAFLD and HCC groups.

Conclusion: Our study underscores the pivotal role of CDT1 in the progression from NAFLD to HCC. This finding opens new avenues for early diagnosis and targeted therapy of HCC, highlighting CDT1 as a potential therapeutic target.

Keywords: Nonalcoholic fatty liver disease, hepatocellular carcinoma, machine learning, CDT1, single-cell sequencing, early diagnosis.

1. INTRODUCTION

Hepatocellular carcinoma (HCC) is estimated to be one of the top four causes of cancer-related deaths worldwide [1]. Non-alcoholic fatty liver disease (NAFLD) is the most

common liver disorder and is attributed to many cases of HCC in numerous developed countries [2, 3]. Mechanisms of NAFLD include insulin resistance and oxidant stress, which leads to simple adipose lesions, chronic inflammation, and cellular fibrosis ultimately leading to a pro-cancer state in HCC [4, 5]. Studies have also reported that immunoregulatory mechanisms contribute to the disease outcome in the progression from NAFLD to fatty liver disease [6]. NAFLD can synergize with other risk factors such as

* Address correspondence to this author at the Clinical Medical College & Affiliated Hospital of Chengdu University, Chengdu University, 610083, P.R. China; E-mail: drshiz1002@hotmail.com

These authors contributed equally to this work.

metabolic dysfunction, chronic inflammation, genetic mutations, and immune dysregulation, accelerating the development of HCC [7]. Compared to viral HCC, patients with NAFLD-HCC are often challenging to detect in the early stages, frequently being diagnosed at an advanced stage of the pathology [8]. Moreover, many NAFLD-HCC patients face limitations in pharmacotherapy due to hepatic dysfunction induced by liver cirrhosis and metabolic syndrome-associated pathologies. Consequently, surgical resection remains the primary choice for HCC treatment for most patients, as recommended by local and global clinical guidelines [9]. Therefore, early diagnosis and prevention of the development of HCC emerge as the most effective strategies for improving patient prognosis, especially for those with NAFLD [10]. The screening and identification of gene functions are key steps in molecular biology and translational medicine research, helping us to understand the pathogenesis of diseases and guide the formulation of clinical treatment strategies. In this study, we collected a large amount of expression data from databases such as Gene Expression Omnibus (GEO) and The Cancer Genome Atlas (TCGA). Through a series of bioinformatics analyses, we identified molecular biomarkers related to the transition from NAFLD to HCC. The identification of new potential marker genes for risk assessment and targeted treatment may help improve patient survival rates and treatment outcomes.

The liver has a remarkable regenerative capacity as an adaptive response to continuous damage [11]. However, in the context of NAFLD, the repeated damage and regeneration processes occurring in chronic liver diseases (such as hepatitis, fatty liver disease, *etc.*) may lead to a dysregulation of cell proliferation [12]. Abnormal proliferation of liver cells not only increases the risk of genetic material errors but may also promote the activation of tumor promoters and the inactivation of tumor suppressors, thereby creating conditions for the occurrence of HCC [13]. Moreover, the chronic inflammatory environment serves as a promoting factor for carcinogenesis, further driving abnormal liver cell proliferation and tumor development by activating liver fibrosis and hepatic stellate cells [14]. Therefore, gaining a comprehensive understanding of the molecular mechanisms of abnormal liver cell division and proliferation in NAFLD, especially the intervention possibilities in this process by regulating the cell cycle, suppressing inflammatory responses, and preventing fibrosis progression, is crucial for developing strategies to prevent HCC [15, 16]. This requires not only focusing on the abnormal lipid metabolism of the liver but also on the precise regulatory mechanisms of cell proliferation during liver regeneration and how to maintain this balance in chronic liver diseases [17].

In conclusion, the objective of this study is to delineate the molecular landscape of the transition from NAFLD to HCC through multidimensional bioinformatics analysis, from the perspective of gene expression. By employing this approach, we aim to offer new insights and strategies for the early diagnosis and treatment of NAFLD-HCC and to provide a solid scientific foundation for personalized medicine for NAFLD, HCC, and related diseases. By identifying new

key genes and screening potential drugs, we hope to contribute to improving patient prognosis and enhancing treatment outcomes.

2. MATERIALS AND METHODS

2.1. Data Acquisition and Filtering

We conducted a search for gene expression profiles and related documentation of patients with NAFLD in the Gene Expression Omnibus (GEO) database (<http://www.ncbi.nlm.nih.gov/geo/>) using the term “nonalcoholic fatty liver disease.” The inclusion criteria were as follows: 1. The data included a control comparison between fatty liver groups and healthy groups. 2. All samples originated from human liver tissue. 3. The dataset's raw data or gene expression profiles were available in GEO. Following these criteria, we identified the NAFLD-related dataset (GSE207310, GSE213621, GSE193066) [18-20]. Additionally, data on HCC were obtained from The Cancer Genome Atlas (TCGA) (<https://portal.gdc.cancer.gov/>) database, selecting the TCGA-LIHC cohort. This cohort (LIRI-JP) includes RNA sequencing data and corresponding clinical information for 369 HCC samples. For validation datasets, a cohort of 68 HCC samples was selected from the International Cancer Genome Consortium (ICGC) database (<https://dcc.icgc.org/>). Furthermore, single-cell data from liver biopsy samples of 3 patients who had both NAFLD and HCC were selected from the GEO database (GSE189175) [21].

2.2. Identification of Co-expressed Differential Genes

First, we utilized the DESeq2 package [22] in RStudio software (version 4.2.3) to perform differential analysis on genes between the disease and control groups within the NAFLD dataset and the TCGA-LIHC dataset. The criteria for identifying differentially expressed genes (DEGs) were set as $|\text{Log2Fold Change}| > 1$ and $|\text{adj. } P \text{ Value}| < 0.01$ to identify differentially expressed genes (DEGs). Subsequently, we employed the 'ggplot2' package (<https://ggplot2.tidyverse.org>) to visualize the results of the differential analysis through volcano plots. Finally, we used the 'ggvenn' package (<https://github.com/yanlinlin82/ggvenn>) to integrate and identify the co-expressed DEGs between the NAFLD dataset and the HCC dataset.

2.3. Functional and Pathway Enrichment Analysis of Differentially Expressed Genes

To explore differentially expressed genes co-expressed in both diseases, we used the cluster profile package in RStudio (version 4.2.3) to perform Gene Ontology (GO) terms and GSEA enrichment analysis. A P -value < 0.05 was set as the cutoff threshold, and the Benjamini-Hochberg method was employed to perform multiple corrections on the raw p -values calculated during functional enrichment analysis [23, 24]. Additionally, we utilized other packages, including dplyr, org.Hs.eg.db (<https://bioconductor.org/packages/org.Hs.eg.db/>) and enrichplot. These packages are instrumental in gene annotation, database connections, and the conversion of gene identifiers to gene names [25, 26].

2.4. Prognostic Model Construction through Machine Learning

Before model construction, cases with a survival time of less than 30 days were excluded from both the TCGA-LIHC and ICGC (LIRI-JP) datasets. The glmnet package in R was used to fit Cox proportional hazards models and Lasso regression models, with cross-validation applied to the Cox proportional hazards model to determine the optimal regularization parameter. Here, 1000 iterations of 5-fold cross-validation were performed to ultimately identify biomarkers from the common differentially expressed genes [27]. Then, using the 'survival' package, risk scores for each sample were calculated based on these five key genes, with the risk score formula: Risky score $\sum_{i=1}^N (coef \times i \text{ expression})$, where coef represents the model coefficient for each key gene, and expression represents the expression level of each key gene. The median of the risk scores was used as the threshold to divide samples into low-risk and high-risk groups. Survival analysis between the two risk groups and the generation of the receiver operator characteristic (ROC) curve were conducted using the survminer (<https://rpkgs.datanovia.com/survminer/>) and timeROC packages [28].

2.5. Single-cell RNA-seq Integration Analysis and Pseudotime Analysis

Based on the single-cell dataset (GSE189175), which includes single-nucleus RNA-seq data from biopsy samples of both tumor and adjacent non-tumor tissues in patients with NAFLD and HCC, we performed principal component analysis (PCA) using the Seurat package in R [29]. Subsequently, the SingleR package (<https://comphealth.ucsf.edu/app/singleR>) and CellMarker were employed to annotate the cells. The annotation results were validated based on the PCA findings [30]. Within the PCA and its annotation results, hepatocytes were selected for pseudotime analysis. The Monocle 3 package was used to cluster, learn, and order the cells, determining the root node and differentiation trajectory of cell differentiation to delineate the cell differentiation path [31].

2.6. Immune Infiltration Analysis

In his study, the CIBERSORT algorithm, an advanced computational method, was employed to quantify the extent of immune cell infiltration in samples from patients with

HCC. The core advantage of CIBERSORT lies in its ability to estimate the relative abundance of 22 different types of immune cells from complex tissue expression data. This method relies on linear support vector regression (SVR) and uses gene expression signatures specific to each immune cell type, enabling accurate analysis of the composition of mixed cell populations [32]. We conducted a comprehensive analysis of each HCC patient sample, aiming to determine the extent of infiltration of different immune cells and their distribution proportions within the tumor microenvironment. Through CIBERSORT analysis, we successfully mapped the infiltration landscape of each type of immune cell in HCC, providing crucial insights into the role of immune cells in the progression of liver cancer. Additionally, this analysis revealed the complexity of the tumor immune microenvironment, laying the foundation for the development of future immunotherapy strategies [33].

3. RESULTS

3.1. Dataset Information

In this study, we selected the GEO dataset (GSE207310) and the TCGA-LIHC cohort data as the discovery cohorts. We selected the ICGC (LIRI-JP) cohort data, the NAFLD disease dataset (GSE213621, GSE193066), and the single-cell dataset (GSE189175) as external validation datasets. Detailed information about the data is presented in Table 1.

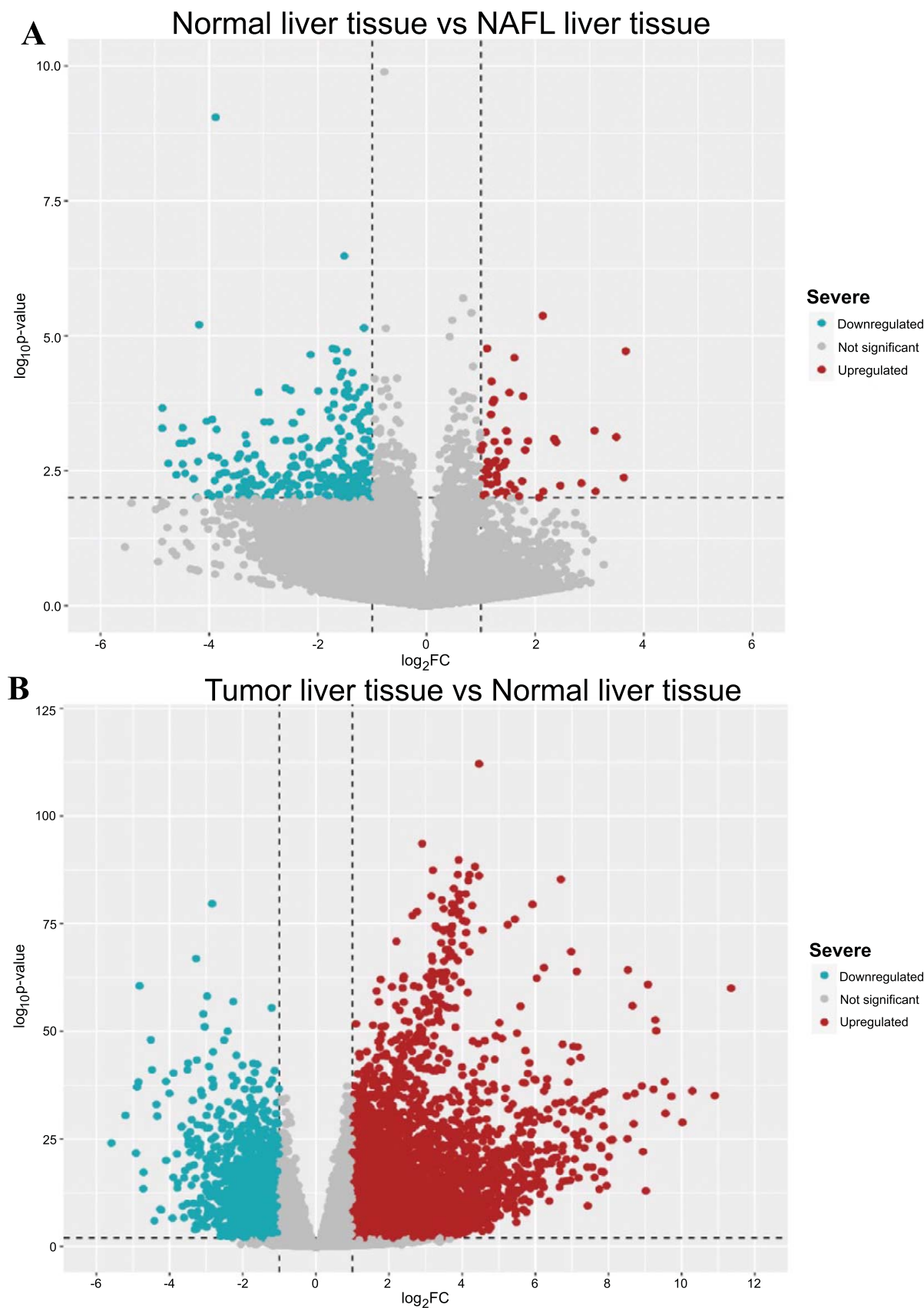
3.2. Differential Gene Expression Analysis and Gene Set Enrichment Analysis (GSEA)

In the analysis of the GEO dataset (GSE207310) comparing NAFLD cases with healthy controls, DEGs were identified in NAFLD patients, with 277 genes upregulated and 58 genes downregulated. These findings are represented in the form of a volcano plot (Fig. 1A). In the TCGA dataset, significant differences in gene expression between HCC cases and healthy controls were found, with 6926 genes upregulated and 1802 genes downregulated, also represented as a volcano plot (Fig. 1B). Using the GSEA method based on the TCGA database, we conducted further analysis on the differences between HCC and control groups. The top five ranked related pathways showed that DEGs in HCC are primarily related to the immunoglobulin complex, immunoglobulin receptor binding, and xenobiotic catabolic process (Fig. 1C).

Table 1. Data characteristics of patients with NAFLD and HCC.

Cohort	Sample Information	Disease	Tissue
GSE207310	15 cases with NAFLD, 5 healthy controls	NAFLD	Liver
GSE213621	299 cases with NAFLD, 69 healthy controls	NAFLD	Liver
GSE193066	106 patients with different disease stages of NAFLD	NAFLD	Liver
GSE189175	3 cases with HCC and NAFLD	HCC&NAFLD	Liver
TCGA	377 cases with HCC	HCC	Liver
LIRI-JP	68 cases with HCC	HCC	Liver

Abbreviations: NAFLD: Non-alcoholic fatty liver disease; HCC: Hepatocellular carcinoma.



(Fig. 1) contd....

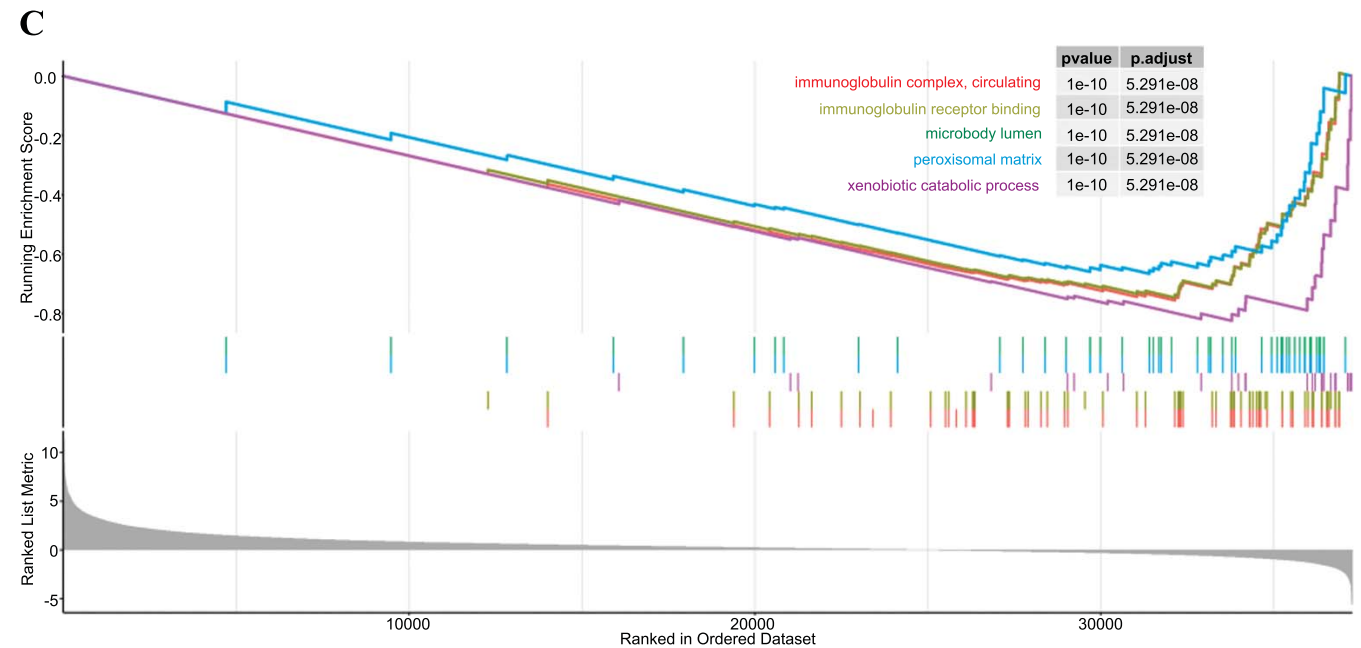


Fig. (1). Differential Gene Expression and Pathway Enrichment in Liver Tissues. (A) Volcano plot illustrating differential gene expression between normal liver tissue and non-alcoholic fatty liver (NAFLD) tissue. (B) Volcano plot illustrating differential gene expression between tumor liver tissue and normal liver tissue (C) Gene Set Enrichment Analysis (GSEA) plot. (A higher resolution / colour version of this figure is available in the electronic copy of the article).

3.3. Enrichment Analysis of Common DEGs

The Venn diagram illustrates that there are 46 common DEGs shared between 272 DEGs identified in NAFLD from the GEO dataset and 3078 DEGs identified in HCC from the TCGA dataset (Fig. 2A). This study utilized the clusterProfiler package in R software to perform GO analysis to elucidate the roles and functions of these DEGs commonly expressed in both NAFLD and HCC. The GO enrichment analysis encompassed three aspects: Biological Processes (BP), Cellular Components (CC), and Molecular Functions (MF), with a focus on the 46 genes shared between both diseases. The results revealed that these genes are primarily involved in lipid transporter activity, triglyceride binding, and cholesterol binding (Fig. 2B).

3.4. Construction and Evaluation of the HCC Prognostic Model

Common DEGs from both NAFLD and TCGA samples were collected for further analysis. After excluding TCGA samples with a survival time of less than 30 days, 340 samples remained. The expression data of the 46 common DEGs in these samples were then used to predict the risk scores using a Lasso Cox linear model in conjunction with TCGA prognostic time. After 10,000 iterations of 5-fold cross-validation, based on the optimal λ, a subset of 5 genes from the common DEGs was identified: IGSF3, CENPW, CDT1, CDC6, C20orf144 (Fig. 3A, D). The coefficients (coef values) corresponding to these 5 biomarkers are listed in Table 2.

The risk score for each patient sample in TCGA was calculated based on the expression levels of the 5 genes, weighted by their respective coefficients. The calculation is as follows: $Y(\text{risk score}) = (0.035 \times \text{IGSF3}) + (0.011 \times \text{CENPW}) + (0.002 \times \text{CDT1}) + (0.024 \times \text{CDC6}) + (0.411 \times \text{C20orf144})$. Based on the median risk score, the TCGA training dataset (n=340) was divided into high-risk (n=170) and low-risk (n=170) groups. The Kaplan-Meier curve showed a significant difference in overall survival (OS) between the high-risk and low-risk groups, with patients in the high-risk group exhibiting a lower probability of survival and a worse prognosis ($P < 0.001$, log-rank test) (Fig. 3B). The linear distribution of the risk scores for HCC patients calculated by the model is shown in Fig. (3E), and the survival status of HCC patients is displayed in a scatter plot in Fig. (3C). The sensitivity and specificity of the prognostic model were evaluated using time-dependent ROC curves, and the area under the curve (AUC) was calculated. The results showed that the model's AUC values for predicting the OS of patients at 1 year, 3 years, and 5 years were 0.758, 0.725, and 0.645, respectively (Fig. 3F).

Table 2. Coefficients of prognostic genes in the predictive model.

Gene ID	Coef
IGSF3	0.035
CENPW	0.011
CDT1	0.002
CDC6	0.024
C20orf144	0.411

3.5. External Dataset Validation of Model Predictive Capability

After constructing a prognostic model based on key genes, the predictive capability of the model was validated on the ICGC-JP dataset (n=65) for HCC patients. First, the risk score for each sample in the dataset was calculated. Based on the median risk score, the entire sample was divided into a high-risk group (n=32) and a low-risk group (n=33). Kaplan-Meier survival curves were plotted, reveal-

ing a significant difference in prognosis between the high and low-risk groups, with the high-risk group exhibiting a worse prognosis (Fig. 4A). A linear graph depicting the distribution of risk scores among HCC patients in the external dataset was generated (Fig. 4B). In the external dataset, ROC curve analysis showed that the AUC values of constructed model for predicting overall survival (OS) at 1 year, 3 years, and 5 years were 0.897, 0.803, and 0.848, respectively (Fig. 4C).

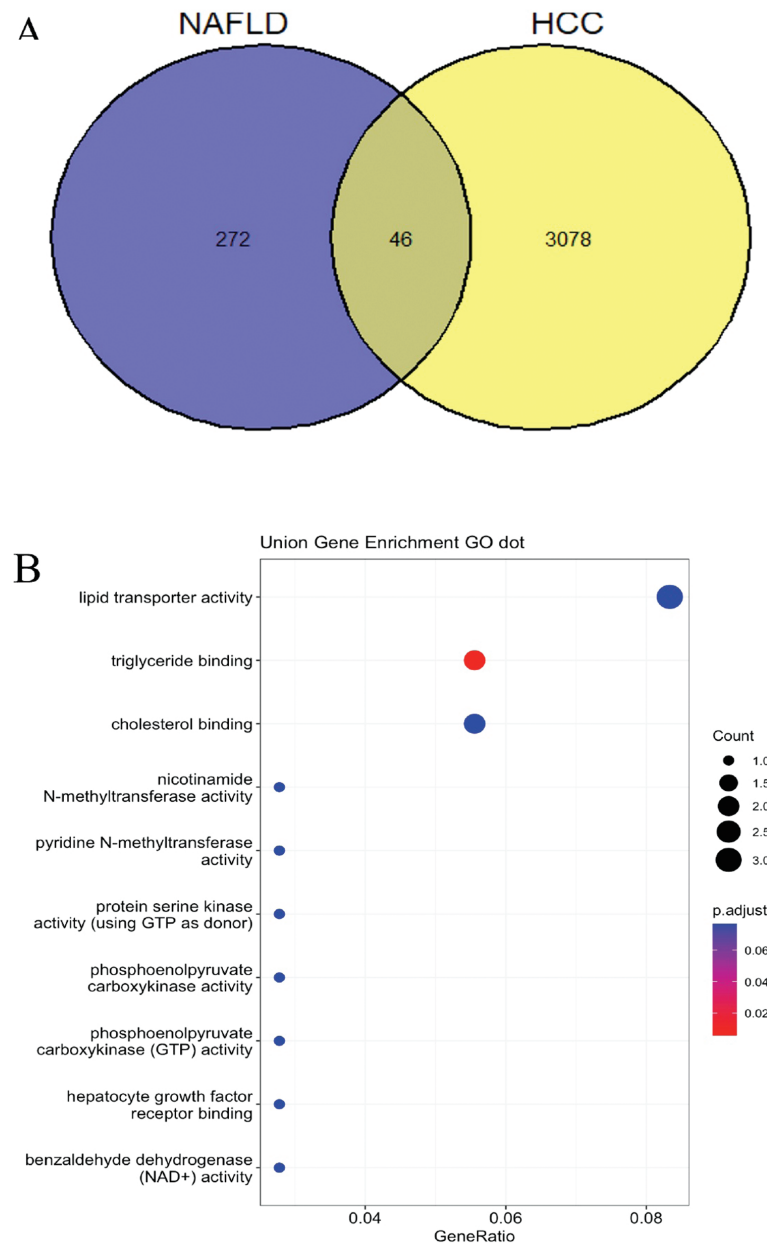
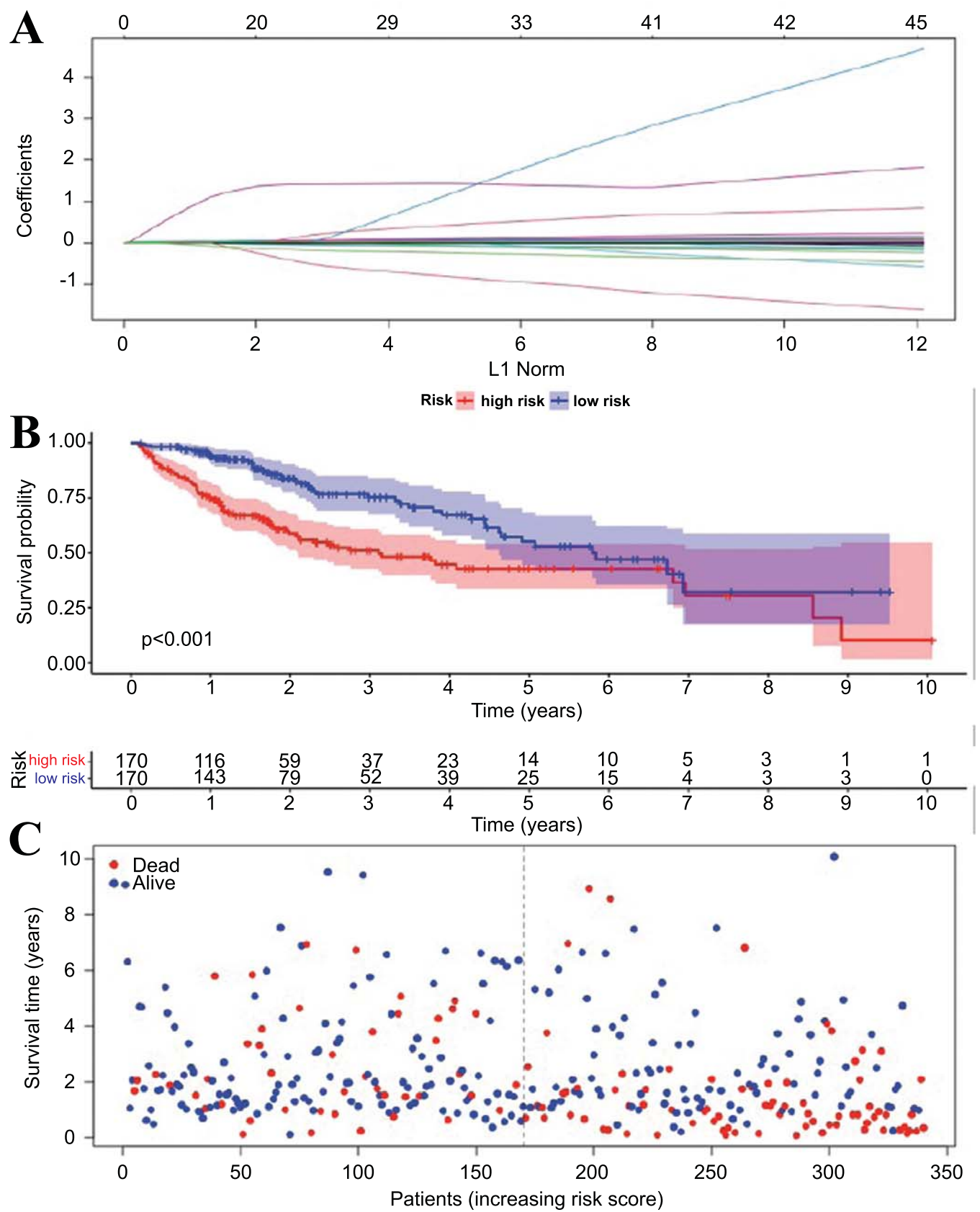


Fig. (2). Overlap of Gene Expression and Functional Enrichment in NAFLD and HCC. **(A)** Venn diagram illustrating the overlap of differentially expressed genes between non-alcoholic fatty liver disease (NAFLD) and hepatocellular carcinoma (HCC). **(B)** Dot plot of Gene Ontology (GO) term enrichment analysis for the union set of differentially expressed genes in both NAFLD and HCC. (A higher resolution / colour version of this figure is available in the electronic copy of the article).



(Fig. 3) contd....

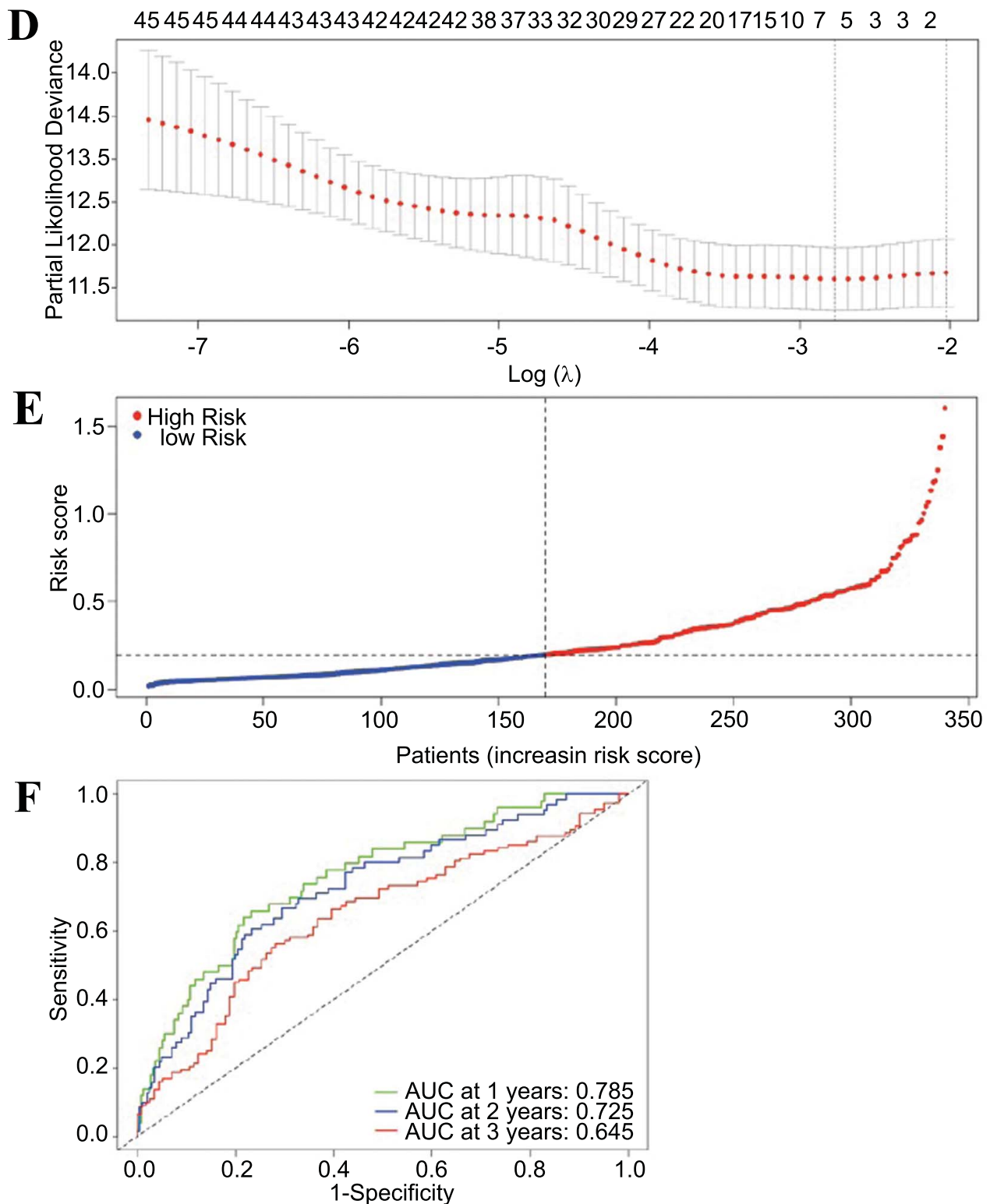


Fig. (3). Construction of LASSO Cox prognostic model and evaluation of the HCC Prognostic Model. (A) Lasso coefficient profiles of the prognostic features across the log(λ) sequence. (B) Kaplan-Meier survival curves stratified by high and low prognostic risk scores. (C) Scatter plot of survival status of patients against their prognostic risk score. (D) Tuning parameter (λ) selection in the lasso model used in cross-validation via mini-mum criteria and the 1-standard error rule. (E) Risk score distribution of patients. (F) Time-dependent Receiver Operating Characteristic (ROC) curves evaluating the predictive accuracy of the prognostic model at 1, 3, and 5 years. (A higher resolution / colour version of this figure is available in the electronic copy of the article).

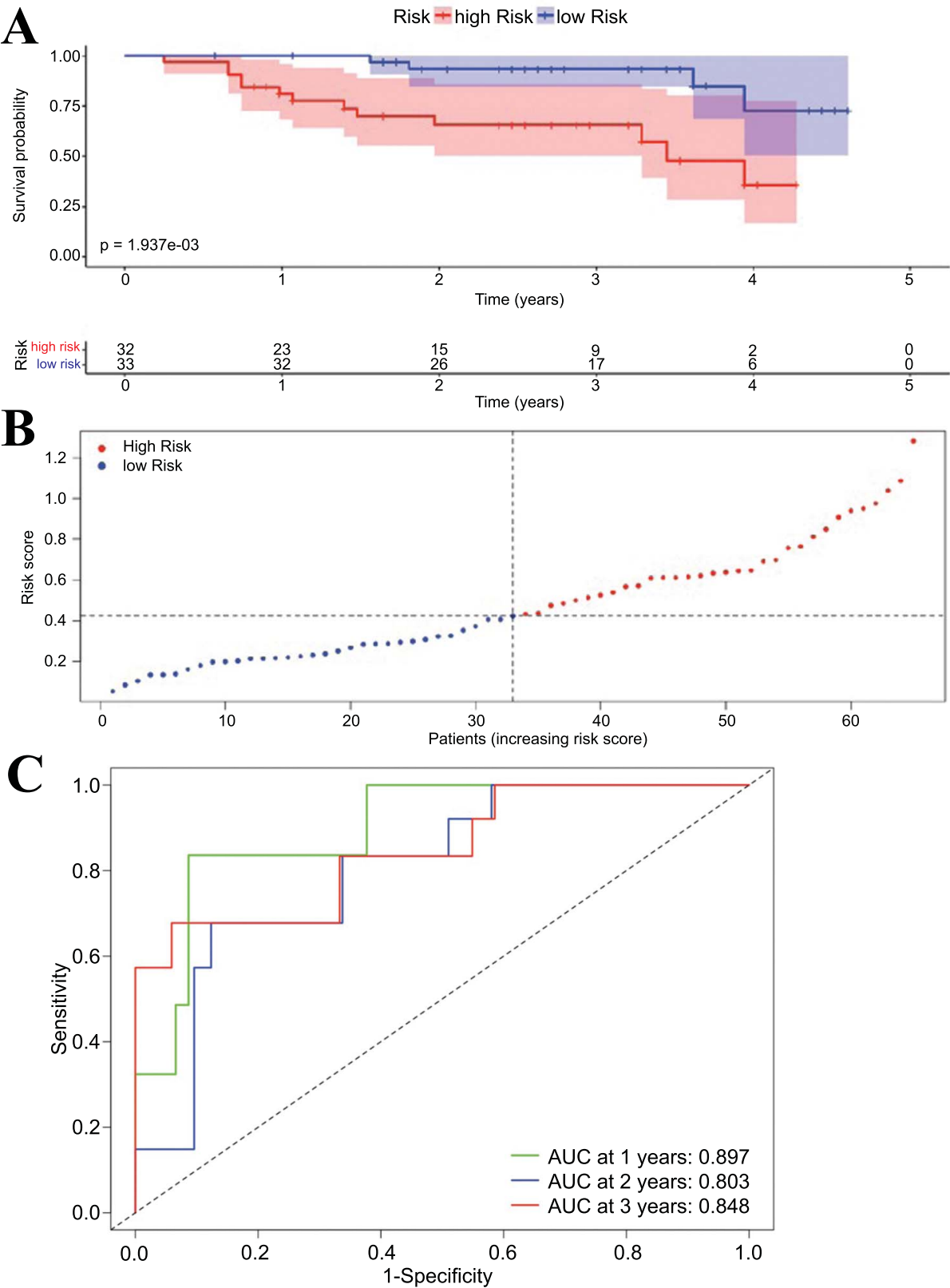


Fig. (4). External Validation of the Prognostic Model's Survival Predictions and Risk Stratification Accuracy. **(A)** Kaplan-Meier survival curves for external validation cohorts, stratified by derived risk scores. **(B)** Risk score distribution among the external validation cohort. **(C)** Time-dependent ROC curves showing the performance of the prognostic model in the external cohort at 1, 2, and 3 years. (A higher resolution / colour version of this figure is available in the electronic copy of the article).

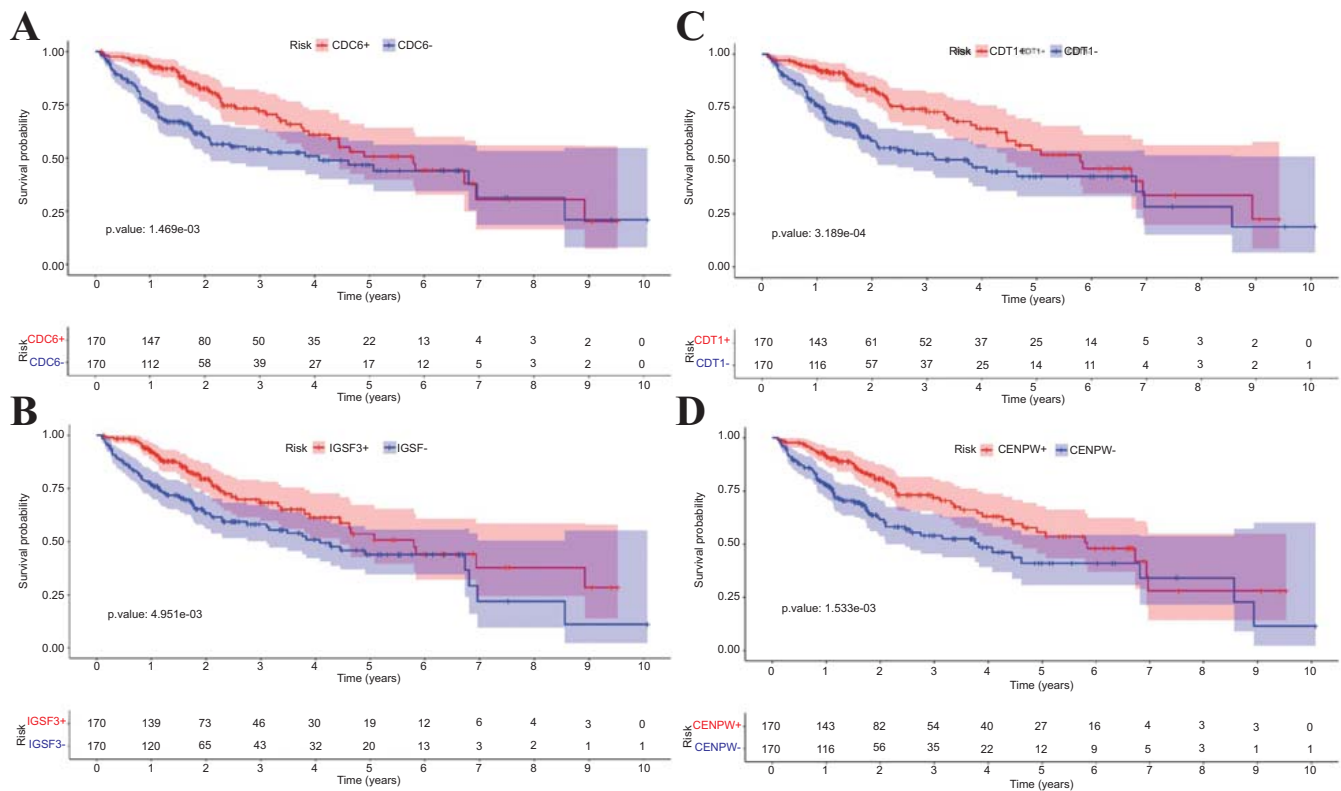


Fig. (5). Single gene Survival Analysis Stratified by Biomarker Expression. **(A)** Kaplan-Meier curve comparing the survival probability over time between patients expressing CDC6 biomarkers. **(B)** Kaplan-Meier curve comparing the survival probability over time between patients expressing IGSF3 biomarkers. **(C)** Survival probability over time for patients categorized by CDT1 expression. **(D)** Kaplan-Meier curve showcasing the difference in survival between patients with CENPW expression. The accompanying risk tables and *p*-value are displayed below the curves. (A higher resolution / colour version of this figure is available in the electronic copy of the article).

3.6. Impact of Single Gene Expression Levels on Prognosis

Furthermore, this study investigated the relationship between the expression levels of each gene and the overall survival time of patients. Kaplan-Meier survival curves were generated for individual genes in relation to HCC patients. It was found that among the biomarkers, CDC6 (*P*-value=1.469e-03), IGSF3 (*P*-value=4.951e-03), CDT1 (*P*-value=3.189e-04), CENPW (*P*-value=1.533e-03) and CENPW (*P*-value=1.533e-03) exhibited a significant relationship between their expression levels and prognosis, with CDT1 exerting the most pronounced impact (Fig. 5A-D).

3.7. Valuation Dataset for NAFLD from Bulk RNA-seq

In dataset, GSE213621, the expression levels of four key genes in NAFLD liver tissue samples were significantly higher in the NAFLD group compared to the normal group. These genes include IGSF3 (*P* = 5.9e-12), CENPW (*P* = 1.3e-07), CDT1 (*P* = 5.8e-07), and CDC6 (*P* = 2.5e-10) (Fig. 6A). The AUC of gene expression used to distinguish between normal and diseased states for CDT1 is 0.85, indicating a high accuracy of CDT1 in distinguishing between

normal and NAFLD groups (Fig. 6B). In the dataset GSE193066, patients with NAFLD were grouped into high, medium and low groups based on activity scores assessed by pathologists³⁴. The expression levels of the CDT1 gene increased significantly with higher NAFLD activity scores, suggesting that CDT1 could effectively predict the progression of NAFLD (Fig. 6C).

3.8. Changes in Key Gene Expression Levels in Hepatocytes during the Transition from NAFLD to HCC

Utilizing the single-cell dataset (GSE189175), we identified the top 2000 highly variable genes from both tumor cells and adjacent non-tumor cells. After conducting PCA, these genes were divided into 30 clusters. Based on the expression of lineage-specific markers, we identified distinct cell types, including hepatocytes, T cells, HSCs, gamma T cells, and B cells (Fig. 7A and B). Hepatocytes were selected for single-cell sorting to delineate the differentiation trajectory nodes in adjacent non-cancer cells, resulting in a differentiation trajectory map and pseudo time plot from adjacent non-tumor cells to liver cancer cells (Fig. 7C). Based on the results of single-cell pseudotime analysis, we examined the changes in the expression levels of the shared

pathogenic genes (IGSF3, CENPW, CDC6, CDT1) from adjacent non-tumor cells to liver cancer cells during the continuous process. The pseudotime expression changes from a

non-cancerous state to a liver cancer state were analyzed. The results showed differential expression levels of CDC6, CDT1, IGSF3, and CENPW (Fig. 7D and E).

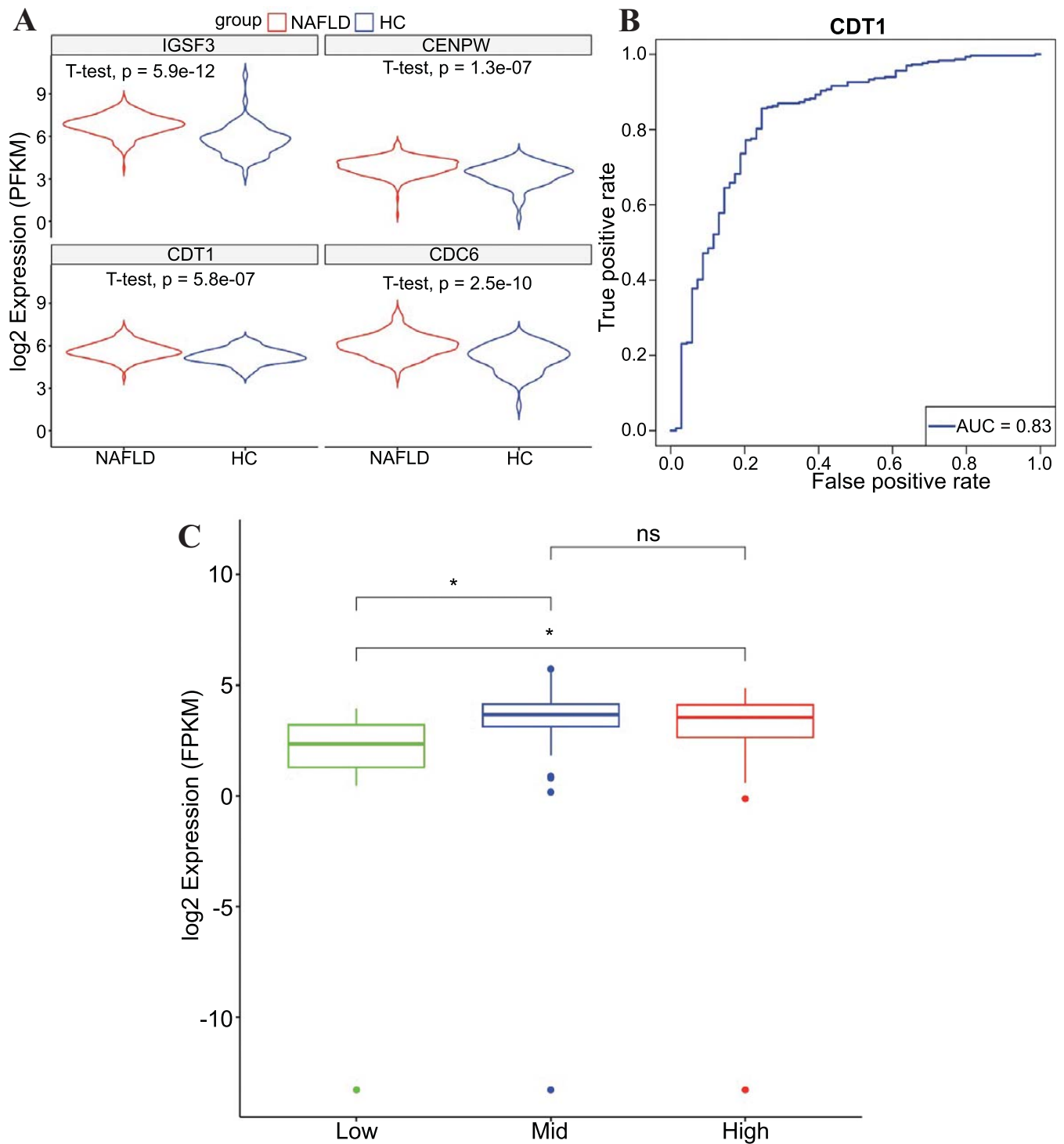
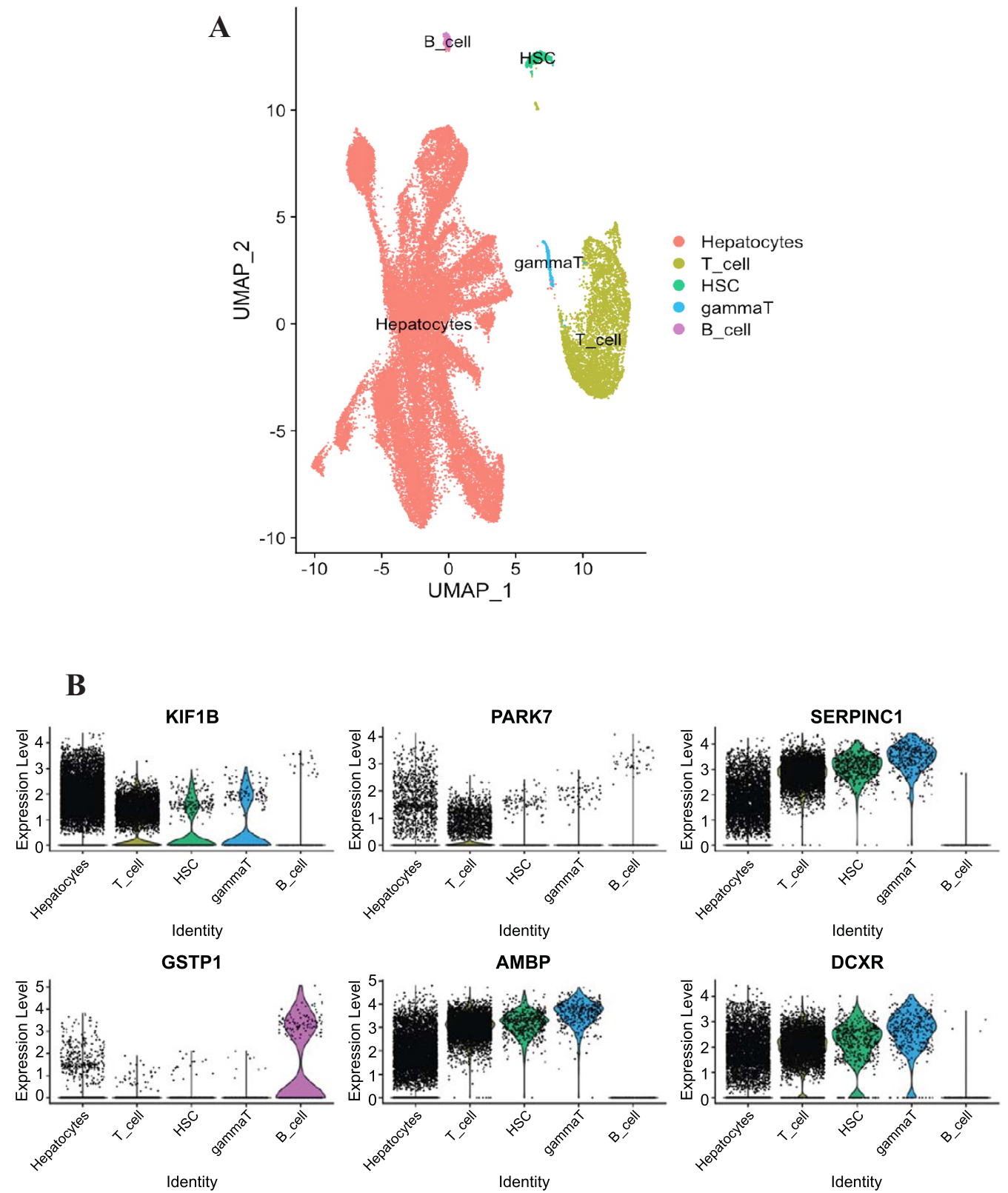
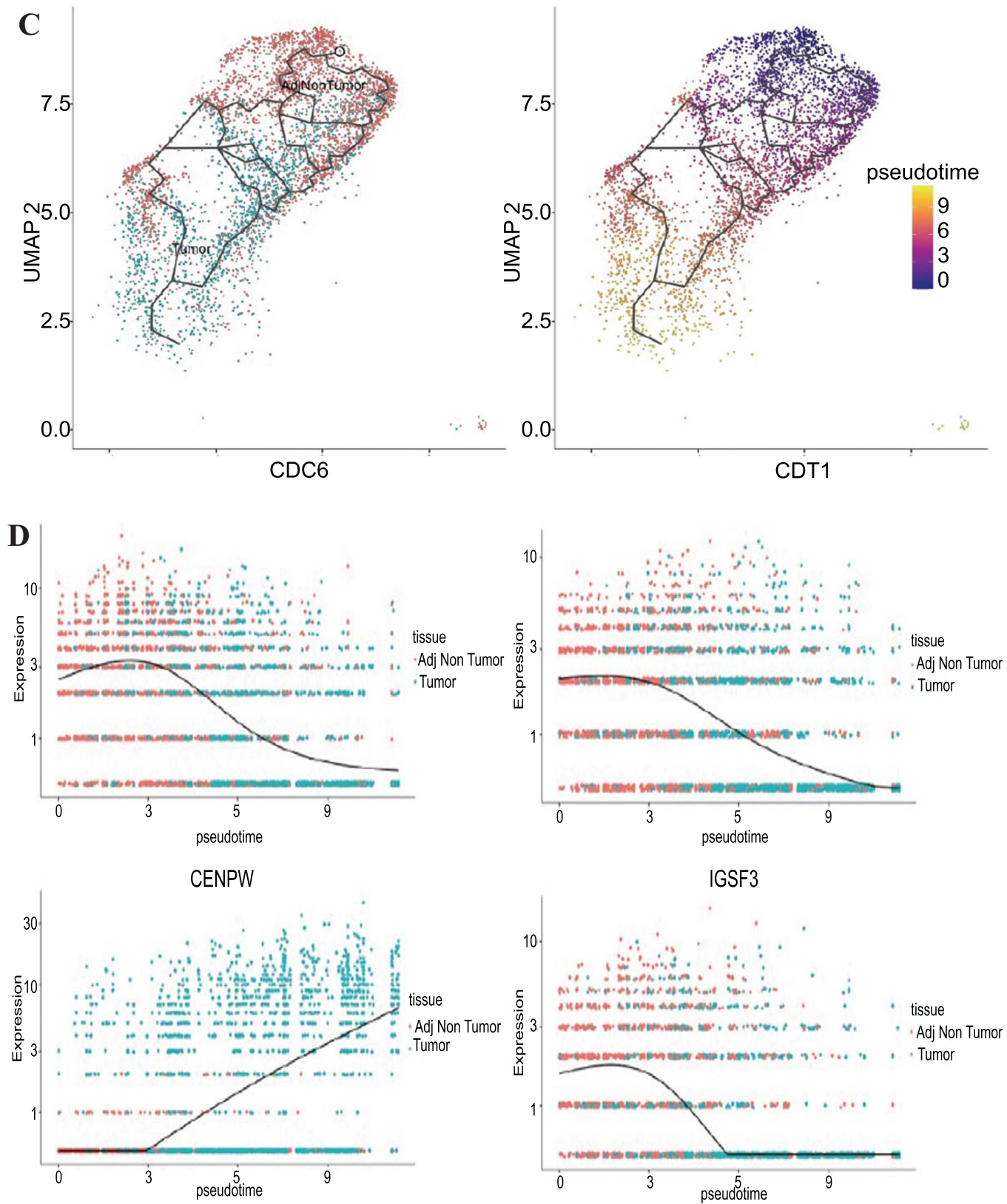


Fig. (6). Gene Expression in Multiple GEO Datasets. **(A)** NFALD Liver Tissue Samples Comparing Normal. **(B)** Distinguishing Power of CDT1 Gene Between Normal and Disease. **(C)** Expression of CDT1 Gene Across Differ-ent NAFLD Activity Scores. *: 0.01 < p < 0.05. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).



(Fig. 7) contd....



(Fig. 7) contd....

3.9. Immune Infiltration Characteristics in High-risk and Low-risk Patients

In this study, CIBERSORT was employed to comprehensively assess the proportions of 22 types of immune cells in the two risk subgroups. The results showed that the infiltration scores of B cells and T cells (CD8+ and CD4+), as indicated by three asterisks, were significantly higher in the high-risk group (P -value <0.001). This suggests a potential link between the co-expressed genes associated with the progression from NAFLD to HCC and an active immune response. NK cells and dendritic cells (DCs), which exhibited significantly higher infiltration scores in the high-risk group, may reflect an enhanced response of the immune system to the aggressive tumor biology associated with the high-risk gene expression profile. The infiltration levels of monocytes and macrophages (M0, M1, M2) showed no significant differences, indicating that the involvement of these immune cells may not be affected by gene expression risk stratification or may represent a common feature of the two risk groups. The results also noted the presence of variability and outliers, especially in NK cells and M2 macrophages, which could reflect individual differences in the tumor microenvironment or the potential complexity of immune-tumor interactions (Fig. 8).

4. DISCUSSION

NAFLD is one of the leading factors contributing to the development and progression of HCC worldwide [7, 34, 35]. NAFLD can progress to Non-Alcoholic Steatohepatitis (NASH), with the malignant type NASH-HCC being one primary predisposing factor for HCC [36]. Studies also indicate that NAFLD serves as a significant risk factor for HCC, a cancer type intricately involving inflammatory and metabolic pathways [37]. The primary characteristic of NAFLD is the accumulation of fat within the liver, which may be related to various biological factors including metabolic abnormalities, gut microbiota imbalance, and inflammation [38-41]. Previously, the precise pathogenesis of NAFLD progressing to metabolic risk factor-related HCC remained incompletely understood. The advent of high-throughput gene chips and transcriptome sequencing has completely changed the systematic analysis approach to disease research [42]. RNA sequencing and high-throughput microarrays aid in identifying reliable biomarkers, classifying diseases, and revealing the mechanisms behind disease progression [43-45]. The discovery of new biomarkers helps in predicting risk and determining the most suitable treatment for individual patients [46, 47]. Therefore, investigating the changes in gene expression during the transition from NAFLD to HCC is a crucial biological process.

In this study, we focused on analyzing DEGs that are commonly expressed in both NAFLD and HCC. Through GO analysis, we identified 46 genes that are co-expressed in both NAFLD and HCC, showing significant enrichment in biological processes related to lipid transporter activity, triglyceride binding, and cholesterol binding. This finding underscores the critical role of lipid metabolism-related functions

and pathways in the transition from NAFLD to HCC [48-50]. Abnormalities in lipid metabolism, such as the abnormal accumulation of fat in the liver and dysregulation of cholesterol metabolism, are commonly observed in this process [51, 52]. Genes involved in lipid transport, and triglyceride, and cholesterol binding may play an important role in regulating these lipid metabolic abnormalities. By participating in the transport and binding of lipids, these genes may influence the accumulation and metabolism of lipids in the liver, thereby contributing to the progression from NAFLD to HCC [53]. Furthermore, through GSEA, we found that DEGs associated with HCC are mainly related to the binding of immunoglobulin complexes, immunoglobulin receptors, and xenobiotic catabolic processes. This indicates that the occurrence and progression of HCC are closely related to the activity of the immune system and drug metabolism pathways. The involvement of immunoglobulin complexes and immunoglobulin receptors may play a pivotal role in the immune regulation and cell signaling mechanisms of liver cancer [54].

Through these analyses, we deepened our understanding of the roles of lipid metabolism and immune regulation mechanisms in the progression from NAFLD to HCC. Subsequently, the Lasso Cox linear model, in conjunction with TC-GA prognostic time, was used to predict the risk scores of the samples. Four gene subsets were identified from the shared DEGs in both NAFLD and HCC, including IGSF3, CENPW, CDT1, and CDC6. The identification of these gene subsets provided robust clues for revealing co-expressed pathogenic genes. Single-cell pseudotime analysis results then identified co-expressed pathogenic genes for NAFLD-HCC, including IGSF3, CENPW, CDC6, and CDT1. CDC6 and CDT1 are proteins necessary for mitosis, IGSF3 is involved in regulating the function of immune cells, and CENPW is a gene associated with DNA replication [55-58]. The overexpression of CDC6 and CDT1 can lead to excessive DNA replication and related instability, thereby promoting the process of carcinogenesis [59]. The expression of IGSF3 and CENPW can affect the timing of homologous chromosome pairing, responding to anomalies within the cell genome [60, 61]. Therefore, further exploration into the biological significance triggered by the aberrant expression of these genes could provide groundbreaking insights into the onset and progression of liver cancer. CDT1, functioning as a key regulator of DNA replication licensing, plays an important role in the pathogenesis of HCC [62]. The overexpression of CDT1 in various cancers, including HCC, induces abnormal replication and malignant transformation [63]. Research has shown that CDT1 transcription is significantly upregulated in HCC samples compared to healthy liver tissue, and elevated CDT1 expression is associated with the progression of clinical stages and poor prognosis in HCC [64]. Recent studies have explored the development of small molecule inhibitors targeting the CDT1/Geminin protein complex. One such compound, AF615, blocks cell cycle progression and reduces cell survival in various cancer cell lines by mimicking the cellular effects of abnormal CDT1/Geminin expression. This compound selectively

affects cancer cells with minimal impact on normal cells, highlighting its potential for cancer treatment targeting CDT1 [65].

In our research, through single-cell analysis technology, we delved into the liver tissues of patients with both NAFLD and HCC. We focused on the transition from peritumoral non-tumor cells to liver cancer cells, and we employed pseudotime analysis techniques to identify a series of key gene expression changes from the non-cancerous to the cancerous state of liver cells. Specifically, we observed a downregulation in the expression levels of the genes CDC6, CDT1, and IGSF3, while the expression of the CENPW gene was upregulated. These changes underscore the important roles these genes collectively play in the progression from fatty liver to liver cancer. CDC6 and CDT1 are key genes in DNA replication and cell cycle regulation [66, 67]. Their downregulation may disrupt the normal regulation of the DNA replication process and the cell cycle, creating conditions for the transition from NAFLD to HCC. IGSF3, a gene involved in cell adhesion and signal transduction, its downregulation may hinder normal cell communication, accelerating liver cell degeneration and tumor transformation [60]. On the other hand, the upregulation of CENPW might indicate abnormalities in chromosome separation and mitosis, associated with disordered proliferation and heightened aggressiveness of liver cells. The knockout of CENPW, combined with sorafenib treatment for HCC, exhibited significant effects [68]. This finding provides further evidence supporting the central role of cell cycle and DNA replication regulation in the development of liver cancer, while also highlighting the importance of cell communication in the exacerbation process of liver diseases.

The results of immune infiltration analysis suggest that the expression of hub genes defining high and low-risk groups in NAFLD-HCC may be related to the dynamics of the immune microenvironment. The gene expression profiles of the high-risk group indicate that the progression from NAFLD to HCC seems to be associated with an increase in immune cells. This may reflect immune surveillance mechanisms or an immune activation microenvironment conducive to tumor progression [69, 70]. The observed immune infiltration patterns, based on the expression of hub genes for classifying HCC patients into different risk groups, provide valuable insights into the immune status of HCC. The distinct immune characteristics associated with high-risk patients could impact the development of precision medicine approaches (including targeted immunotherapies) and serve as a foundation for further exploration of the immunobiology underlying the progression from NAFLD to HCC [71, 72].

Through the analysis of second-generation sequencing data and single-cell sequencing data, our study enhanced the comprehension of the dynamic gene expression changes during the progression from NAFLD to HCC. Moreover, we identified potential targets for future therapeutic strategies aimed at these key molecular pathways. Our research offers

insights for further studies into the molecular mechanisms of liver cancer development and progression. However, it is important to note that our findings are predictive results derived from data mining and analysis, which require further experimental validation. Additional population studies and clinical research will help to confirm the association between these genes and the development of fatty liver and liver cancer, providing novel targets for treatment and prevention. We anticipate ongoing advancements in this research field, bringing forth new perspectives for the prevention and treatment of liver cancer.

CONCLUSION

We utilized machine learning and single-cell RNA-seq analysis to investigate the relationship between NAFLD gene expression and the development of HCC. Ultimately, we identified 4 related hub genes (IGSF3, CENPW, CDT1, and CDC6). Subsequent analysis highlighted CDT1 as the gene most significantly associated with the prognosis of HCC. Additionally, single-cell RNA-seq analysis showed differences in gene expression levels of these hub genes during the progression from NAFLD to HCC.

AUTHORS' CONTRIBUTIONS

Both X.H and J.M designed the study and wrote the article. X.Y and P.W provided valuable suggestions for this study. Data collection and analysis were performed by X.Y. Material preparation was performed by L.Z and N.L. Z.S was responsible for the study design and critical revision of the article.

LIST OF ABBREVIATIONS

DCs	= Dendritic Cells
DEGs	= Differentially Expressed Genes
GEO	= Gene Expression Omnibus
GO	= Gene Ontology
HCC	= Hepatocellular Carcinoma
ICGC	= International Cancer Genome Consortium
NAFLD	= Non-alcoholic Fatty Liver Disease
NASH	= Non-Alcoholic Steatohepatitis
OS	= Overall Survival
PCA	= Principal Component Analysis
ROC	= Receiver Operator Characteristic
SVR	= Support Vector Regression
TCGA	= The Cancer Genome Atlas

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not applicable.

HUMAN AND ANIMAL RIGHTS

Not applicable.

CONSENT FOR PUBLICATION

Not applicable.

AVAILABILITY OF DATA AND MATERIALS

Gene Expression Omnibus (GEO): <https://www.ncbi.nlm.nih.gov/geo/>; The Cancer Genome Atlas (TCGA): <https://portal.gdc.cancer.gov/>; International Cancer Genome Consortium (ICGC): <https://dcc.icgc.org/>.

FUNDING

Studies were funded by the National Natural Science Foundation of China (No. 81900404), Natural Science Foundation of Sichuan Province (NO.24NSFSC2535), Innovation team project of Clinical Medical College & Affiliated Hospital of Chengdu University (Award Number: CDFY-CX202208), Sichuan Provincial Science and Technology Foundation (22NZZH0031), Longquanyi Talents program (No.90), Hospital level project of affiliated hospital of Chengdu University (Y202232, Y202234, Y202320).

CONFLICT OF INTEREST

The authors declare no conflict of interest, financial or otherwise.

ACKNOWLEDGEMENTS

Declared none.

REFERENCES

- [1] Fitzmaurice, C.; Allen, C.; Barber, R.M.; Barregard, L.; Bhutta, Z.A.; Brenner, H.; Dicker, D.J.; Chimed-Orchir, O.; Dandona, R.; Dandona, L. Global, Regional, and National Cancer Incidence, Mortality, Years of Life Lost, Years Lived With Disability, and Disability-Adjusted Life-years for 32 Cancer Groups, 1990 to 2015: A Systematic Analysis for the Global Burden of Disease Study. *JAMA Oncol.*, **2017**, 3(4), 524-548. <http://dx.doi.org/10.1001/jamaoncol.2016.5688>
- [2] Estes, C.; Anstee, Q.M.; Arias-Loste, M.T.; Bantel, H.; Bellentani, S.; Caballeria, J.; Colombo, M.; Craxi, A.; Crespo, J.; Day, C.P.; Eguchi, Y.; Geier, A.; Kondili, L.A.; Kroy, D.C.; Lazarus, J.V.; Loomba, R.; Manns, M.P.; Marchesini, G.; Nakajima, A.; Negro, F.; Petta, S.; Ratzl, V.; Romero-Gomez, M.; Sanyal, A.; Schattenberg, J.M.; Tacke, F.; Tanaka, J.; Trautwein, C.; Wei, L.; Zeuzem, S.; Razavi, H. Modeling NAFLD disease burden in China, France, Germany, Italy, Japan, Spain, United Kingdom, and United States for the period 2016–2030. *J. Hepatol.*, **2018**, 69(4), 896-904. <http://dx.doi.org/10.1016/j.jhep.2018.05.036> PMID: 29886156
- [3] Younossi, Z.M.; Koenig, A.B.; Abdelatif, D.; Fazel, Y.; Henry, L.; Wymer, M. Global epidemiology of nonalcoholic fatty liver disease—Meta-analytic assessment of prevalence, incidence, and outcomes. *Hepatology*, **2016**, 64(1), 73-84. <http://dx.doi.org/10.1002/hep.28431> PMID: 26707365
- [4] Danford, C.J.; Jiang, Z.G.; Salomaa, V.; Färkkilä, M.; Jula, A.; Männistö, S.; Lundqvist, A.; Valsta, L.; Perola, M.; Åberg, F. Insulin Resistance and Genetic Risk Predict Liver-Related Outcomes and Death in Nonalcoholic Fatty Liver Disease. *Hepatol. Commun.*, **2019**, 3(12), 1704-1705. <http://dx.doi.org/10.1002/hep4.1437> PMID: 31832576
- [5] 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(10), 589-604. <http://dx.doi.org/10.1038/s41575-019-0186-y> PMID: 31439937
- [6] Lawal, G.; Xiao, Y.; Rahnama-Azar, A.A.; Tsilimigras, D.I.; Kuang, M.; Bakopoulos, A.; Pawlik, T.M. The Immunology of Hepatocellular Carcinoma. *Vaccines (Basel)*, **2021**, 9(10), 1184. <http://dx.doi.org/10.3390/vaccines9101184> PMID: 34696292
- [7] Huang, D.Q.; El-Serag, H.B.; Loomba, R. Global epidemiology of NAFLD-related HCC: trends, predictions, risk factors and prevention. *Nat. Rev. Gastroenterol. Hepatol.*, **2021**, 18(4), 223-238. <http://dx.doi.org/10.1038/s41575-020-00381-6> PMID: 33349658
- [8] Ahn, S.Y.; Kim, S.B.; Song, I.H. Clinical Patterns and Outcome of Hepatocellular Carcinoma in Patients with Nonalcoholic Fatty Liver Disease. *Can. J. Gastroenterol. Hepatol.*, **2020**, 2020, 1-9. <http://dx.doi.org/10.1155/2020/4873875> PMID: 32566546
- [9] Galle, P.R.; Forner, A.; Llovet, J.M.; Mazzaferro, V.; Piscaglia, F.; Raoul, J.L.; Schirmacher, P.; Vilgrain, V. EASL Clinical Practice Guidelines: Management of hepatocellular carcinoma. *J. Hepatol.*, **2018**, 69(1), 182-236. <http://dx.doi.org/10.1016/j.jhep.2018.03.019> PMID: 29628281
- [10] Foerster, F.; Gairing, S.J.; Müller, L.; Galle, P.R. NAFLD-driven HCC: Safety and efficacy of current and emerging treatment options. *J. Hepatol.*, **2022**, 76(2), 446-457. <http://dx.doi.org/10.1016/j.jhep.2021.09.007> PMID: 34555422
- [11] Hora, S.; Wuestefeld, T. Liver Injury and Regeneration: Current Understanding, New Approaches, and Future Perspectives. *Cells*, **2023**, 12(17), 2129. <http://dx.doi.org/10.3390/cells12172129> PMID: 37681858
- [12] Pibiri, M.; Simbula, G. Role of the Hippo pathway in liver regeneration and repair: Recent advances. *Inflamm. Regen.*, **2022**, 42(1), 59. <http://dx.doi.org/10.1186/s41232-022-00235-5> PMID: 36471376
- [13] He, Z.; Yang, H.Y. Mechanism of hepatocytes differentiation and dedifferentiation in liver regeneration: Process and exploration. *Hepatobiliary Pancreat. Dis. Int.*, **2023**, 22(4), 331-332. <http://dx.doi.org/10.1016/j.hbpd.2022.11.008> PMID: 36456429
- [14] Marengo, A.; Rosso, C.; Bugianesi, E. Liver Cancer: Connections with Obesity, Fatty Liver, and Cirrhosis. *Annu. Rev. Med.*, **2016**, 67(1), 103-117. <http://dx.doi.org/10.1146/annurev-med-090514-013832> PMID: 26473416
- [15] Moayedfard, Z.; Sani, F.; Alizadeh, A.; Bagheri Lankarani, K.; Zarei, M.; Azarpira, N. The role of the immune system in the pathogenesis of NAFLD and potential therapeutic impacts of mesenchymal stem cell-derived extracellular vesicles. *Stem Cell Res. Ther.*, **2022**, 13(1), 242. <http://dx.doi.org/10.1186/s13287-022-02929-6> PMID: 35672797
- [16] Puengel, T.; Liu, H.; Guillot, A.; Heymann, F.; Tacke, F.; Peiseler, M. Nuclear Receptors Linking Metabolism, Inflammation, and Fibrosis in Nonalcoholic Fatty Liver Disease. *Int. J. Mol. Sci.*, **2022**, 23(5), 2668. <http://dx.doi.org/10.3390/ijms23052668> PMID: 35269812
- [17] Le, P.; Payne, J.Y.; Zhang, L.; Deshpande, A.; Rothberg, M.B.; Alkhouri, N.; Herman, W.; Hernandez, A.V.; Schleicher, M.; Ye, W.; Dasarthy, S. Disease State Transition Probabilities Across the Spectrum of NAFLD: A Systematic Review and Meta-Analysis of Paired Biopsy or Imaging Studies. *Clin. Gastroenterol. Hepatol.*, **2023**, 21(5), 1154-1168. <http://dx.doi.org/10.1016/j.cgh.2022.07.033> PMID: 35933075
- [18] Chen, T.; Dalton, G.; Oh, S.H.; Maeso-Diaz, R.; Du, K.; Meyers, R.A.; Guy, C.; Abdelmalek, M.F.; Henao, R.; Guarnieri, P.; Pullen, S.S.; Gregory, S.; Locker, J.; Brown, J.M.; Diehl, A.M. Hepatocyte Smoothed Activity Controls Susceptibility to Insulin Resistance and Nonalcoholic Fatty Liver Disease. *Cell. Mol. Gastroenterol. Hepatol.*, **2023**, 15(4), 949-970. <http://dx.doi.org/10.1016/j.jcmgh.2022.12.008> PMID: 36535507
- [19] Fujiwara, N.; Kubota, N.; Crouchet, E.; Koneru, B.; Marquez, C.A.; Jajoriya, A.K.; Panda, G.; Qian, T.; Zhu, S.; Goossens, N.; Wang, X.; Liang, S.; Zhong, Z.; Lewis, S.; Taouli, B.; Schwartz, M.E.; Fiel, M.I.; Singal, A.G.; Marrero, J.A.; Fobar, A.J.; Parikh, N.D.; Raman, I.; Li, Q.Z.; Taguri, M.; Ono, A.; Aikata, H.; Naka-

- hara, T.; Nakagawa, H.; Matsushita, Y.; Tateishi, R.; Koike, K.; Kobayashi, M.; Higashi, T.; Nakagawa, S.; Yamashita, Y.; Beppu, T.; Baba, H.; Kumada, H.; Chayama, K.; Baumert, T.F.; Hoshida, Y. Molecular signatures of long-term hepatocellular carcinoma risk in nonalcoholic fatty liver disease. *Sci. Transl. Med.*, **2022**, *14*(650)eabo4474
<http://dx.doi.org/10.1126/scitranslmed.abo4474> PMID: 35731891
- [20] Larsen, F.T.; Hansen, D.; Terkelsen, M.K.; Bendixen, S.M.; Avolio, F.; Wernberg, C.W.; Lauridsen, M.M.; Grønkaer, L.L.; Jacobsen, B.G.; Klinggaard, E.G.; Mandrup, S.; Di Caterino, T.; Siersbæk, M.S.; Chandran, V.I.; Graversen, J.H.; Krag, A.; Grøntved, L.; Ravnskaer, K. Stellate cell expression of SPARC-related modular calcium-binding protein 2 is associated with human non-alcoholic fatty liver disease severity. *JHEP Reports*, **2023**, *5*(2)100615
<http://dx.doi.org/10.1016/j.jhepr.2022.100615> PMID: 36687468
- [21] Alvarez, M.; Benhammou, J.N.; Darci-Maher, N.; French, S.W.; Han, S.B.; Sinsheimer, J.S.; Agopian, V.G.; Pisegna, J.R.; Pajukanta, P. Human liver single nucleus and single cell RNA sequencing identify a hepatocellular carcinoma-associated cell-type affecting survival. *Genome Med.*, **2022**, *14*(1), 50.
<http://dx.doi.org/10.1186/s13073-022-01055-5> PMID: 35581624
- [22] Love, M.I.; Huber, W.; Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.*, **2014**, *15*(12), 550.
<http://dx.doi.org/10.1186/s13059-014-0550-8> PMID: 25516281
- [23] Liberzon, A.; Birger, C.; Thorvaldsdóttir, H.; Ghandi, M.; Mesirov, J.P.; Tamayo, P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.*, **2015**, *1*(6), 417-425.
<http://dx.doi.org/10.1016/j.cels.2015.12.004> PMID: 26771021
- [24] Zhou, Y.; Zhou, B.; Pache, L.; Chang, M.; Khodabakhshi, A.H.; Tanaseichuk, O.; Benner, C.; Chanda, S.K. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.*, **2019**, *10*(1), 1523.
<http://dx.doi.org/10.1038/s41467-019-09234-6> PMID: 30944313
- [25] Wickham, H. The Split-Apply-Combine Strategy for Data Analysis. *J. Stat. Softw.*, **2011**, *40*(1)
<http://dx.doi.org/10.18637/jss.v040.i01>
- [26] Yu, G.; Wang, L.G.; He, Q.Y. ChIPseeker: An R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics*, **2015**, *31*(14), 2382-2383.
<http://dx.doi.org/10.1093/bioinformatics/btv145> PMID: 25765347
- [27] Friedman, J.; Hastie, T.; Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J. Stat. Softw.*, **2010**, *33*(1), 1-22.
<http://dx.doi.org/10.18637/jss.v033.i01> PMID: 20808728
- [28] Heagerty, P.J.; Lumley, T.; Pepe, M.S. Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics*, **2000**, *56*(2), 337-344.
<http://dx.doi.org/10.1111/j.0006-341X.2000.00337.x> PMID: 10877287
- [29] Zhao, X.; Wu, S.; Fang, N.; Sun, X.; Fan, J. Evaluation of single-cell classifiers for single-cell RNA sequencing data sets. *Brief. Bioinform.*, **2020**, *21*(5), 1581-1595.
<http://dx.doi.org/10.1093/bib/bbz096> PMID: 31675098
- [30] Zhang, X.; Lan, Y.; Xu, J.; Quan, F.; Zhao, E.; Deng, C.; Luo, T.; Xu, L.; Liao, G.; Yan, M.; Ping, Y.; Li, F.; Shi, A.; Bai, J.; Zhao, T.; Li, X.; Xiao, Y. CellMarker: A manually curated resource of cell markers in human and mouse. *Nucleic Acids Res.*, **2019**, *47*(D1), D721-D728.
<http://dx.doi.org/10.1093/nar/gky900> PMID: 30289549
- [31] Cao, J.; Spielmann, M.; Qiu, X.; Huang, X.; Ibrahim, D.M.; Hill, A.J.; Zhang, F.; Mundlos, S.; Christiansen, L.; Steemers, F.J.; Trapnell, C.; Shendure, J. The single-cell transcriptional landscape of mammalian organogenesis. *Nature*, **2019**, *566*(7745), 496-502.
<http://dx.doi.org/10.1038/s41586-019-0969-x> PMID: 30787437
- [32] Newman, A.M.; Liu, C.L.; Green, M.R.; Gentles, A.J.; Feng, W.; Xu, Y.; Hoang, C.D.; Diehn, M.; Alizadeh, A.A. Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods*, **2015**, *12*(5), 453-457.
<http://dx.doi.org/10.1038/nmeth.3337> PMID: 25822800
- [33] Ali, H.R.; Chlon, L.; Pharoah, P.D.P.; Markowitz, F.; Caldas, C. Patterns of Immune Infiltration in Breast Cancer and Their Clinical Implications: A Gene-Expression-Based Retrospective Study. *PLoS Med.*, **2016**, *13*(12)e1002194
<http://dx.doi.org/10.1371/journal.pmed.1002194> PMID: 27959923
- [34] Kleiner, D.E.; Brunt, E.M.; Van Natta, M.; Behling, C.; Contos, M.J.; Cummings, O.W.; Ferrell, L.D.; Liu, Y.C.; Torbenson, M.S.; Unalp-Arida, A.; Yeh, M.; McCullough, A.J.; Sanyal, A.J.; Network, N.S.C.R. Design and validation of a histological scoring system for nonalcoholic fatty liver disease. *Hepatology*, **2005**, *41*(6), 1313-1321.
<http://dx.doi.org/10.1002/hep.20701> PMID: 15915461
- [35] Tan, D.J.H.; Ng, C.H.; Lin, S.Y.; Pan, X.H.; Tay, P.; Lim, W.H.; Teng, M.; Syn, N.; Lim, G.; Yong, J.N.; Quek, J.; Xiao, J.; Dan, Y.Y.; Siddiqui, M.S.; Sanyal, A.J.; Muthiah, M.D.; Loomba, R.; Huang, D.Q. Clinical characteristics, surveillance, treatment allocation, and outcomes of non-alcoholic fatty liver disease-related hepatocellular carcinoma: A systematic review and meta-analysis. *Lancet Oncol.*, **2022**, *23*(4), 521-530.
[http://dx.doi.org/10.1016/S1470-2045\(22\)00078-X](http://dx.doi.org/10.1016/S1470-2045(22)00078-X) PMID: 35255263
- [36] Takino, J.; Nagamine, K.; Hori, T.; Sakasai-Sakai, A.; Takeuchi, M. Contribution of the toxic advanced glycation end-products-receptor axis in nonalcoholic steatohepatitis-related hepatocellular carcinoma. *World J. Hepatol.*, **2015**, *7*(23), 2459-2469.
<http://dx.doi.org/10.4254/wjh.v7.i23.2459> PMID: 26483867
- [37] Zoller, H.; Tilg, H. Nonalcoholic fatty liver disease and hepatocellular carcinoma. *Metabolism*, **2016**, *65*(8), 1151-1160.
<http://dx.doi.org/10.1016/j.metabol.2016.01.010> PMID: 26907206
- [38] Friedman, S.L.; Neuschwander-Tetri, B.A.; Rinella, M.; Sanyal, A.J. Mechanisms of NAFLD development and therapeutic strategies. *Nat. Med.*, **2018**, *24*(7), 908-922.
<http://dx.doi.org/10.1038/s41591-018-0104-9> PMID: 29967350
- [39] Jou, J.; Choi, S.; Diehl, A. Mechanisms of disease progression in nonalcoholic fatty liver disease. *Semin. Liver Dis.*, **2008**, *28*(4), 370-379.
<http://dx.doi.org/10.1055/s-0028-1091981> PMID: 18956293
- [40] Kotsiliti, E.; Leone, V.; Schuehle, S.; Govaere, O.; Li, H.; Wolf, M.J.; Horvatic, H.; Bierwirth, S.; Hundertmark, J.; Inverso, D.; Zizmare, L.; Sarusi-Portuguez, A.; Gupta, R.; O'Connor, T.; Gianou, A.D.; Shiri, A.M.; Schlesinger, Y.; Beccaria, M.G.; Rennert, C.; Pfister, D.; Öllinger, R.; Gadajalova, I.; Ramadori, P.; Rahbari, M.; Rahbari, N.; Healy, M.E.; Fernández-Vaquero, M.; Yahoo, N.; Janzen, J.; Singh, I.; Fan, C.; Liu, X.; Rau, M.; Feuchtenberger, M.; Schwaneck, E.; Wallace, S.J.; Cockell, S.; Wilson-Kanamori, J.; Ramachandran, P.; Kho, C.; Kendall, T.J.; Leblond, A.L.; Keppler, S.J.; Bielecki, P.; Steiger, K.; Hofmann, M.; Rippe, K.; Zitzelsberger, H.; Weber, A.; Malek, N.; Luedde, T.; Vucur, M.; Augustin, H.G.; Flavell, R.; Parnas, O.; Rad, R.; Pabst, O.; Henderson, N.C.; Huber, S.; Macpherson, A.; Knolle, P.; Claassen, M.; Geier, A.; Trautwein, C.; Unger, K.; Elinav, E.; Waisman, A.; Abdullah, Z.; Haller, D.; Tacke, F.; Anstee, Q.M.; Heikenwalder, M.; Intestinal, B. Intestinal B cells license metabolic T-cell activation in NASH microbiota/antigen-independently and contribute to fibrosis by IgA-FcR signalling. *J. Hepatol.*, **2023**, *79*(2), 296-313.
<http://dx.doi.org/10.1016/j.jhep.2023.04.037> PMID: 37224925
- [41] Palma, R.; Pronio, A.; Romeo, M.; Scognamiglio, F.; Ventriglia, L.; Ormando, V.M.; Lamazza, A.; Pontone, S.; Federico, A.; D'Allo, M. The Role of Insulin Resistance in Fueling NAFLD Pathogenesis: From Molecular Mechanisms to Clinical Implications. *J. Clin. Med.*, **2022**, *11*(13), 3649.
<http://dx.doi.org/10.3390/jcm11133649> PMID: 35806934
- [42] Bustoros, M.; Sklaventis-Pistofidis, R.; Park, J.; Redd, R.; Zhitomirsky, B.; Dunford, A.J.; Saleem, K.; Tai, Y.T.; Anand, S.; Mouhieddine, T.H.; Chavda, S.J.; Boehner, C.; Elagina, L.; Neuse, C.J.; Cha, J.; Rahmat, M.; Taylor-Weiner, A.; Van Allen, E.; Kumar, S.; Kastritis, E.; Leshchiner, I.; Morgan, E.A.; Laubach, J.; Casneuf, T.; Richardson, P.; Munshi, N.C.; Anderson, K.C.; Trippa, L.; Aguet, F.; Stewart, C.; Dimopoulos, M.A.; Yong, K.; Bergsagel, P.L.; Manier, S.; Getz, G.; Ghobrial, I.M. Genomic Profiling of Smoldering Multiple Myeloma Identifies Patients at a High Risk of Disease Progression. *J. Clin. Oncol.*, **2020**, *38*(21), 2380-2389.
<http://dx.doi.org/10.1200/JCO.20.00437> PMID: 32442065
- [43] Mortazavi, A.; Williams, B.A.; McCue, K.; Schaeffer, L.; Wold,

- B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods*, **2008**, 5(7), 621-628.
http://dx.doi.org/10.1038/nmeth.1226 PMID: 18516045
- [44] Montgomery, S.B.; Sammeth, M.; Gutierrez-Arcelus, M.; Lach, R.P.; Ingle, C.; Nisbett, J.; Guigo, R.; Dermitzakis, E.T. Transcriptome genetics using second generation sequencing in a Caucasian population. *Nature*, **2010**, 464(7289), 773-777.
http://dx.doi.org/10.1038/nature08903 PMID: 20220756
- [45] Nagalakshmi, U.; Waern, K.; Snyder, M. RNA-Seq: A method for comprehensive transcriptome analysis. *Curr. Protoc. Mol. Biol.*, **2010**, 1-13.
http://dx.doi.org/10.1002/0471142727.mb0411s89
- [46] Wang, Y.; Tong, Z.; Zhang, W.; Buzdin, A.; Mu, X.; Yan, Q.; Zhao, X.; Chang, H.H.; Duhon, M.; Zhou, X.; Zhao, G.; Chen, H.; Li, X. FDA-Approved and Emerging Next Generation Predictive Biomarkers for Immune Checkpoint Inhibitors in Cancer Patients. *Front. Oncol.*, **2021**, 11683419
http://dx.doi.org/10.3389/fonc.2021.683419 PMID: 34164344
- [47] Moiseev, A.; Albert, E.; Lubarsky, D.; Schroeder, D.; Clark, J. Transcriptomic and Genomic Testing to Guide Individualized Treatment in Chemoresistant Gastric Cancer Case. *Biomedicines*, **2020**, 8(3), 67.
http://dx.doi.org/10.3390/biomedicines8030067 PMID: 32210001
- [48] Bort, A.; Sánchez, B.G.; de Miguel, I.; Mateos-Gómez, P.A.; Diaz-Laviada, I. Dysregulated lipid metabolism in hepatocellular carcinoma cancer stem cells. *Mol. Biol. Rep.*, **2020**, 47(4), 2635-2647.
http://dx.doi.org/10.1007/s11033-020-05352-3 PMID: 32125560
- [49] Blencowe, M.; Karunanayake, T.; Wier, J.; Hsu, N.; Yang, X. Network Modeling Approaches and Applications to Unravelling Non-Alcoholic Fatty Liver Disease. *Genes (Basel)*, **2019**, 10(12), 966.
http://dx.doi.org/10.3390/genes10120966 PMID: 31771247
- [50] Li, N.; Li, X.; Ding, Y.; Liu, X.; Diggle, K.; Kisseleva, T.; Brenner, D.A. SREBP Regulation of Lipid Metabolism in Liver Disease, and Therapeutic Strategies. *Biomedicines*, **2023**, 11(12), 3280.
http://dx.doi.org/10.3390/biomedicines11123280 PMID: 38137501
- [51] Tenen, D.G.; Chai, L.; Tan, J.L. Metabolic alterations and vulnerabilities in hepatocellular carcinoma. *Gastroenterol. Rep. (Oxf.)*, **2021**, 9(1), 1-13.
http://dx.doi.org/10.1093/gastro/goaa066 PMID: 33747521
- [52] Vitullo, M.; Gnodi, E.; Rosini, G.; Meneveri, R.; Giovannoni, R.; Barisani, D. Current Therapeutic Approaches Targeting Lipid Metabolism in NAFLD. *Int. J. Mol. Sci.*, **2023**, 24(16), 12748.
http://dx.doi.org/10.3390/ijms241612748 PMID: 37628929
- [53] Sozio, M.; Liangpunsakul, S.; Crabb, D. The role of lipid metabolism in the pathogenesis of alcoholic and nonalcoholic hepatic steatosis. *Semin. Liver Dis.*, **2010**, 30(4), 378-390.
http://dx.doi.org/10.1055/s-0030-1267538 PMID: 20960377
- [54] Fan, J.; Wang, L.; Chen, M.; Zhang, J.; Li, J.; Song, F.; Gu, A.; Yin, D.; Yi, Y. Analysis of the expression and prognosis for leukocyte immunoglobulin-like receptor subfamily B in human liver cancer. *World J. Surg. Oncol.*, **2022**, 20(1), 92.
http://dx.doi.org/10.1186/s12957-022-02562-w PMID: 35321724
- [55] Caillat, C.; Perrakis, A. Cdt1 and geminin in DNA replication initiation. *Subcell Biochem*, **2012**, 62, 61-87.
http://dx.doi.org/10.1007/978-94-007-4572-8_5
- [56] Fujita, M. Cdt1 revisited: Complex and tight regulation during the cell cycle and consequences of deregulation in mammalian cells. *Cell Div.*, **2006**, 1(1), 22.
http://dx.doi.org/10.1186/1747-1028-1-22 PMID: 17042960
- [57] Walmsley, S.R. Pressure regulates immune-cell function. *Nature*, **2019**, 573(7772), 41-42.
http://dx.doi.org/10.1038/d41586-019-02339-4 PMID: 31481766
- [58] Cheon, Y.; Lee, S. CENP-W inhibits CDC25A degradation by destabilizing the SCF β -TrCP-1 complex at G2/M. *FASEB J*, **2018**.
http://dx.doi.org/10.1096/fj.201701358RRR
- [59] Yang, C.; Xie, N.; Luo, Z.; Ruan, X.; Zhang, Y.; Wang, W.; Huang, Y. The Effect of High CDC6 Levels on Predicting Poor Prognosis in Colorectal Cancer. *Chemotherapy*, **2022**, 67(1), 47-56.
http://dx.doi.org/10.1159/000519913 PMID: 35034031
- [60] Sheng, P.; Zhu, H.; Zhang, W.; Xu, Y.; Peng, W.; Sun, J.; Gu, M.; Jiang, H. The immunoglobulin superfamily member 3 (IGSF3) promotes hepatocellular carcinoma progression through activation of the NF- κ B pathway. *Ann. Transl. Med.*, **2020**, 8(6), 378-378.
http://dx.doi.org/10.21037/atm.2020.02.14 PMID: 32355822
- [61] Zhou, Z.; Zhou, Z.; Huang, Z.; He, S.; Chen, S. Histone-fold centromere protein W (CENP-W) is associated with the biological behavior of hepatocellular carcinoma cells. *Bioengineered*, **2020**, 11(1), 729-742.
http://dx.doi.org/10.1080/21655979.2020.1787776 PMID: 32635817
- [62] Zhang, H. Regulation of DNA Replication Licensing and Re-Replication by Cdt1. *Int. J. Mol. Sci.*, **2021**, 22(10), 5195.
http://dx.doi.org/10.3390/ijms22105195 PMID: 34068957
- [63] Cai, C.; Zhang, Y.; Hu, X.; Hu, W.; Yang, S.; Qiu, H.; Chu, T. CDT1 Is a Novel Prognostic and Predictive Biomarkers for Hepatocellular Carcinoma. *Front. Oncol.*, **2021**, 11721644
http://dx.doi.org/10.3389/fonc.2021.721644 PMID: 34631549
- [64] Jia, W.; Xie, L.; Wang, X.; Zhang, Q.; Wei, B.; Li, H.; Qin, S.; Chen, S.; Liu, J.; Tan, Y.; Zheng, S.; Liang, X.; Yang, X. The impact of MCM6 on hepatocellular carcinoma in a Southern Chinese Zhuang population. *Biomed. Pharmacother.*, **2020**, 127110171
http://dx.doi.org/10.1016/j.biopha.2020.110171 PMID: 32403044
- [65] Karantzeli, N.; Petropoulos, M.; De Marco, V.; Egan, D.A.; Fish, A.; Christodoulou, E.; Will, D.W.; Lewis, J.D.; Perrakis, A.; Lygerou, Z.; Taraviras, S. Small Molecule Inhibitor Targeting CDT1/Geminin Protein Complex Promotes DNA Damage and Cell Death in Cancer Cells. *Front. Pharmacol.*, **2022**, 13860682
http://dx.doi.org/10.3389/fphar.2022.860682 PMID: 35548337
- [66] Lim, N.; Townsend, P.A. Cdc6 as a novel target in cancer: Oncogenic potential, senescence and subcellular localisation. *Int. J. Cancer*, **2020**, 147(6), 1528-1534.
http://dx.doi.org/10.1002/ijc.32900 PMID: 32010971
- [67] Moriyama, M.; Kanda, T.; Midorikawa, Y.; Matsumura, H.; Masuzaki, R.; Nakamura, H.; Ogawa, M.; Matsuoka, S.; Shibata, T.; Yamazaki, M.; Kuroda, K.; Nakayama, H.; Higaki, T.; Kanemaru, K.; Miki, T.; Sugitani, M.; Takayama, T. The proliferation of atypical hepatocytes and CDT1 expression in noncancerous tissue are associated with the postoperative recurrence of hepatocellular carcinoma. *Sci. Rep.*, **2022**, 12(1), 20508.
http://dx.doi.org/10.1038/s41598-022-25201-6 PMID: 36443564
- [68] Ho, C.M.; Lin, K.T.; Shen, R.; Gu, D.L.; Lee, S.S.; Su, W.H.; Jou, Y.S. Prognostic comparative genes predict targets for sorafenib combination therapies in hepatocellular carcinoma. *Comput. Struct. Biotechnol. J.*, **2022**, 20, 1752-1763.
http://dx.doi.org/10.1016/j.csbj.2022.04.008 PMID: 35495118
- [69] Fujiwara, N.; Nakagawa, H. Clinico-histological and molecular features of hepatocellular carcinoma from nonalcoholic fatty liver disease. *Cancer Sci.*, **2023**, 114(10), 3825-3833.
http://dx.doi.org/10.1111/cas.15925 PMID: 37545384
- [70] Dai, K.; Liu, C.; Guan, G.; Cai, J.; Wu, L. Identification of immune infiltration-related genes as prognostic indicators for hepatocellular carcinoma. *BMC Cancer*, **2022**, 22(1), 496.
http://dx.doi.org/10.1186/s12885-022-09587-0 PMID: 35513781
- [71] Wang, Y.; Yuan, Z.; Zhu, Q.; Ma, J.; Lu, Q.; Xiao, Z. Disulfidoptosis-related genes of prognostic signature and immune infiltration features in hepatocellular carcinoma supported by bulk and single-cell RNA sequencing data. *J. Gastrointest. Oncol.*, **2024**, 15(1), 377-396.
http://dx.doi.org/10.21037/jgo-23-949 PMID: 38482234
- [72] Zhou, D.; Wang, Y.; Wei, W.; Zhou, W.; Gu, J.; Kong, Y.; Yang, Q.; Wu, Y. m6A regulator-mediated methylation modification highlights immune infiltration patterns for predicting risk in hepatocellular carcinoma. *J. Cancer Res. Clin. Oncol.*, **2023**, 149(7), 3661-3680.
http://dx.doi.org/10.1007/s00432-022-04255-z PMID: 35972694