

# Development of a Bile Acid-Related Gene Signature for Predicting Survival in Patients with Hepatocellular Carcinoma

**Gang Wang**

Zhejiang University School of Medicine

**Jun Guan**

Zhejiang University School of Medicine

**Qin Yang**

zhejiang university school of medicine

**Fengtian Wu**

Zhejiang University School of Medicine

**Junwei Shao**

Zhejiang University School of Medicine

**Qihui Zhou**

Zhejiang University School of Medicine

**Zixuan Guo**

Zhejiang University School of Medicine

**Yanli Ren**

Zhejiang University School of Medicine

**Haihong Zhu**

Zhejiang University School of Medicine

**Zhi Chen** (✉ [zjuchenzhi@zju.edu.cn](mailto:zjuchenzhi@zju.edu.cn))

Zhejiang University <https://orcid.org/0000-0002-0848-1502>

---

## Research article

**Keywords:** bile acid, hepatocellular carcinoma, enterohepatic cycle, TCGA, prognosis, gene signature

**Posted Date:** February 10th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-215296/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

## Abstract

**Background:** Hepatocellular carcinoma (HCC) is one of the most common diseases, threatening millions of patients annually. Increasing evidence supports that bile acid (BA) has an impact on HCC through inflammation, DNA damage or other mechanisms.

**Methods:** With the data from The Cancer Genome Atlas portal, a total of 127 BA-associated genes were analyzed in HCC tumor and non-tumor samples, and then, using univariate and multivariate Cox regression, genes with correlations to the prognosis of HCC patients were identified. Then, a prediction model with identified genes was constructed for evaluating the risk of HCC patients for prognosis.

**Results:** Twenty-six genes with differential expression between HCC and control tissue samples were identified, of which 19 genes had up-regulated expression and 7 genes had down-regulated expression in tumor samples. Three genes, NPC1, ABCC1 and SLC51B, were extrapolated to construct a prediction model for prognosis of HCC patients.

**Conclusion:** The three-gene prediction model was more reliable than the pathological staging characters of the tumor for the prognosis and survival of HCC patients. Additionally, the up-regulated genes facilitating the transport of BAs are associated with poor prognosis of HCC patients and genes of de novo synthesis of BAs benefits HCC patients.

## Introduction

Liver cancer is one of the most common cancers in the world, and is the third highest cause of cancer mortality with 0.84 million new cases and 0.78 million death annually [1]. China has a heavy burden of HBV infections, resulting in a high incidence of liver cancer, which is one of the leading causes of cancer deaths in China [2]. Additionally, it was reported that the incidence rate of liver cancer is increasing and faster than other types of cancers [3]. It is known that 70% of the whole blood supply of the liver is circulated from the portal vein, in which blood is transported from the intestine and contains nutrients, metabolites, products of gut bacteria as well as bile acids [4].

In the past decades, more and more investigating interests are focusing on the correlations between gut microbiota and HCC [5]. The gut microbiota produces numerous metabolites everyday in human body, including lipopolysaccharides (LPS), short chain fatty acids (SCFAs), and bile acids (BAs), which could promote haptic inflammation, resulting in liver diseases, such as liver fibrosis, liver cirrhosis, and even HCC [5]. Among of them, BAs are the major components of a cycle between enterointestine and liver, named enterohepatic cycling, which maintains the pool of BAs that are essential for the homeostasis of cholesterol and human nutrition-uptake [6–9]. BAs are a group of heterogeneous cholic acids with a steroid nucleus, commonly consisting of cholic acid (CA), chenodeoxycholic acid (CDCA), deoxycholic acid (DCA), lithocholic acid (LCA) and ursodeoxycholic acid (UDCA); CA, CDCA, DCA and LCA, which are free bile acids, can be conjugated by glycine (G) or taurine (T) to formulate (G/T)-(CA/CDCA/DCA/LCA) conjugated bile acids that are the major components in human BAs [10]. BAs are primarily synthesized

from cholesterol through multiple modified enzymatic steps involving 17 enzymes including rate-limiting enzyme cholesterol 7 $\alpha$ -hydroxylase (CYP7A1), sterol 27 $\alpha$ -hydroxylase (CYP27A1), oxysterol and steroid 7 $\alpha$ -hydroxylase (CYP7B1) and other related enzymes in hepatocytes (reviewed in [10]). BAs form the majority of the solid components in gallbladder bile, usually sodium ( $\text{Na}^+$ ) and potassium ( $\text{K}^+$ ) salts of the BAs [10, 11].

In humans, bile including BAs and other constituents are usually stored in the gallbladder and, after a meal, are released into the intestinal lumen as emulsifiers to solubilize fats and fat-soluble vitamins and facilitate uptake of nutrients [7, 11]. In the human intestine, 95% of BAs in the gut are absorbed into the intestinal blood and circulated into the liver through the portal vein [12]. In the liver, several transporters are involved in the process of BAs cross-membrane transport within the gut-liver-gallbladder circle. In the gut, most BAs (~ 95%) are up-taken into the ileum by the transporter apical sodium-dependent bile acid transporter (ASBT) and exported into the portal vein blood by organic solute transporter  $\alpha$  and  $\beta$  heterodimer (OST $\alpha$ /OST $\beta$ ) [13]. In the liver, hepatocytes up-take BAs from the venous blood through transporters  $\text{Na}^+$  taurocholate cotransporting polypeptide (NTCP/SLC10A1) and organic anion transporting polypeptide (OATPs) family members (OATP-A/SLC21A3, OATP-C/SLC21A9 and OATP-8/SLC21A8 in human); the former are  $\text{Na}^+$ -dependent and the later are  $\text{Na}^+$ -independent [14]. The efflux of BAs in hepatocytes involves multidrug resistance (MRP) subfamily, such as MRP-1, -2, -3 and - 6 [14]. Additionally, bile canalicular excretion of BAs requires two types of transporters, bile salt export pump BSEP (ABCB11) and the ATP-binding cassette (ABC) superfamily members, including ABCA (1, 2 and 3), ABCG (2, 5 and 8) and MRP2 (ABCC2), that pump BAs out [14].

The disorder of bile acids synthesis or recycling process could result in diseases such as gall stone, obesity, liver diseases and other metabolic syndromes [15–17]. Furthermore, BAs can act as agents that initiate signaling pathways and directly induce cellular injuries involved in many types of cancers including liver cancer [18–21]. Many reports suggested that BAs have a key role in cell proliferation, inflammation, cell invasion, cellular metabolism and other cellular behaviors [19, 20, 22, 23]. Here, we explore the genes involved in metabolic pathways and transportation systems of BAs in liver cells and are expected to get more important makers for the prognosis and survival of hepatocellular carcinoma (HCC) patients.

## Materials And Methods

### Genes involved in transport and metabolism of BAs

The bile acids metabolic gene name list was obtained from the Gene Set Enrichment Analysis (GSEA) website (<http://software.broadinstitute.org/gsea/msigdb>, file name: Bile acids metabolism gene set, 112 genes) [24]. Additionally, 15 genes, implicated in transporters for hepatocyte influx and outflux of BAs, such as ABCB11, ABCC1, ABCC2, ABCC3, ABCC6, ABCG2, ABCG5, SLC10A1, SLC10A2, SLC51A, SLC51B, SLC01B3 and SLC02B1, as well as synthesis of BAs, like ACOT8 and BAAT, were added to the list [14]. Thus, 127 BA-associated genes were enrolled in the present study (Table S1).

# Data acquisition

The data of HCC patients were downloaded from The Cancer Genome Atlas (TCGA) portal (<https://portal.gdc.cancer.gov/>), and only TCGA-LIHC (Liver Hepatocellular Carcinoma) data were included. As a result, 424 sample sequencing files and 377 clinical data files of HCC patients were downloaded directly from the website. All HCC patients had completed clinical character information and follow-up data ( $\geq 1$  month). Among the sequencing samples, there were 50 non-tumor samples and 374 HCC samples, and both groups were used to analyze for different expression of BA-associated genes. The data of follow-up time was used for evaluating the effects of BA-associated genes on the survival of HCC patients. P-value  $< 0.05$  and fold-change  $> |2|$  were considered as significant.

# Data processing

Perl software (version 5.28) and R software (version 4.0) were employed for data processing. Perl was used for the combination of clinical data and gene expression data based on patient ID, and for merging the splicing file from individual samples or patients.

Identification of BA-associated genes with differential expression between the tumor and non-tumor groups were conducted by R software with the “limma” package. For the probability of type I error, adjusted P value  $< 0.05$  was applied as the threshold for significant differentially expressed genes [25, 26]. Additionally, fold-change of gene expression was  $> |2|$ . The graphs of heatmap, volcano and boxplot for illustrating the levels of genes were constructed using packages “pheatmap”, “limma” and “ggpubr” in R.

Enrichment and annotation of genes with GO (Gene Ontology) terms and KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways were performed by R software with packages “clusterprofiler”, “org.Hs.eg.db”, “enrichplot”, “ggplot2” or “GOplot”. During the processes, adjusted P-value  $< 0.05$  was the threshold used for annotation and enrichment of GO terms and KEGG pathways [27, 28].

# Construction of prediction model for prognosis of HCC patients

In the present study, survival analysis of identified genes was analyzed by packages “survival” and “survminer” implemented in R. Additionally, “survivalROC” package in R was used for obtaining the receiver-operator characteristic (ROC) curve for the prediction of prognostic value.

Univariate COX regression analysis was used for exploring the correlation of gene expression and survival of patients ( $P < 0.05$  as the threshold). Multivariate COX regression analysis was used for the identification of BA-associated genes, which were candidate genes for the construction of the prediction model. The coefficients of selected genes were used for constructing the model. The risk score of every

patient = coefficient1 × expression value of gene1 + coefficient2 × expression value of gene2 + ... ... coefficient(n) × expression value of gene(n) (Table S2). The median value of risk score divided the patients into two groups, the high-risk group and the low-risk group. Two packages in R, "survival" and "survminer", were employed to construct the survival curve of the two groups. The timeline of the ROC curve was from 0 to 10 years. The ROC curve with high- and low-risk groups was constructed using clinical characters, including age, gender, stage, grade and TNM, and the risk-score model under a timeline of multiple years.

## Correlation analysis of clinical characters and the model genes

According to gender (male and female), stage (I, II, III and IV), grade (1, 2, 3 and 4) and TNM (tumor (T), nodes (N), and metastases (M)), HCC patients were divided into two groups, respectively. Then, the expression value of the three genes in the prediction model and the risk score of the patients in each of the two groups were statistically analyzed for significant differences. P-values < 0.05 were considered significant.

## Immunohistochemistry

Anti-SLC51B antibody (Cat. ab121285) and Niemann pick c1 antibody (Cat. ab134113) were purchased from Abcam (USA), and MRP1/ABCC1 antibody (Cat. 72202) was purchased from Cell signaling technology Inc. The liver tumor tissue array was purchased from Superchip, Inc. (Shanghai, China). The immunohistochemistry processes were performed as described in previous work [29].

## Statistical analyses

All statistical analyses were performed using R software with various packages, and the differences were considered significant if the P-value was < 0.05 or 0.01. Among the data process, KEGG and GO analysis used a significance threshold of under the adjusted P-value < 0.05, whereas the Wilcox test was used to identify differentially expressed BA-associated genes.

## Results

### Identification of 26 BA-related genes with significant differential expression between tumor and non-tumor in HCC patients

To explore the differential expression of the genes involved in the metabolism and transport of bile acids in HCC, we firstly downloaded 112 genes from the GSEA website, and then another 15 genes associated with the metabolism and transport of BAs in the liver were added to the list, resulting in a total of 127

genes (Table S1). Then, we compared the expression of the 127 genes in tumor and non-tumor samples; 26 genes, including ABCA2, ABCA3, ABCA4, ABCC1, ABCD1, ACSL1, AKR1D1, ALDH8A1, BBOX1, CYP39A1, CYP7A1, DIO2, EFHC1, FADS1, FADS2, GNPAT, KLF1, LIPE, NPC1, PEX6, PFKM, SLC27A5, SLC29A1, SLC35B2, SLC51B and SLC01B3, were identified under the criteria of fold change  $> |2|$  and adjusted P-value  $< 0.05$  (false discovery rate, FDR  $< 0.05$ ) (Fig. 1A) [26]. Among the identified genes, 19 genes (ABCA2, ABCA3, ABCA4, ABCC1, ABCD1, CYP7A1, DIO2, EFHC1, FADS1, FADS2, GNPAT, KLF1, LIPE, NPC1, PEX6, PFKM, SLC29A1, SLC35B2 and SLC51B) were up-regulated in tumor samples, whereas 7 genes (ACSL1, AKR1D1, ALDH8A1, BBOX1, CYP39A1, SLC27A5 and SLC01B3) had up-regulated levels in non-tumor samples ( $P < 0.05$ ) (Fig. 1B and C). The results suggested that most of the genes with differential expression are transporter proteins, like solute carrier (SLC) family and ATP-binding cassette (ABC) family proteins, which carry bile acids or metabolic mediates shuttling in the liver.

### Annotation and enrichment of identified genes by GO and KEGG analyses

From the 26 identified genes, the functions and involved pathways were further explored using GO and KEGG functional enrichment analyses (adjusted  $P < 0.05$ ). We found that these genes were involved in primary bile acid biosynthesis, ABC transporters, PPAR signaling pathway and monocarboxylic acid synthesis. The details are displayed in Fig. 2 and Fig. S1. The enrichment analysis and annotation showed that the genes with increased expression in HCC were related to ABC transporters, cholesterol metabolism and biosynthesis of fatty acids (KEGG), as well as lipid transport and localization (GO). Meanwhile, the genes with down-regulated expression in HCC were involved in BA metabolism process (GO) and primary BA biosynthesis (KEGG) (Fig. 2 and Fig. S2). Additionally, some of genes were also implicated in PPAR signaling pathway, peroxisome and metabolism of sterol. These results support that reduced BA synthesis and increased transporters of BAs and lipid (including fatty acids) have important roles in HCC.

## Identification Of Survival-related Genes Related To Ba Metabolism

To study whether identified genes have effects on survival time, univariate Cox regression was performed using package "Survival" in R (cutoff value:  $P < 0.05$ ). A total of 376 HCC patients with follow-up time were included in the cohort. Consequently, 12 genes were identified; 4 of these genes, SLC27A5, CYP7A1, AKR1D1 and ALDH8A1, benefit HCC patients, and the remaining 8 genes, EFHC1, ABCC1, NPC1, ABCD1, SLC35B2, ABCA4, ABCA3 and SLC51B, increase the risk of poor prognosis (Fig. 3). Furthermore, multivariate Cox regression was conducted, in which all 25 genes were included. Three genes, NPC1, ABCC1 and SLC51B, were determined to have significant relationships with the survival time of HCC patients (Fig. 4A and Table 1). These results support that high expression of NPC1, ABCC1 and SLC51B may increase the risk of poor prognosis and be negatively correlated with the survival of HCC patients (hazard ratio of 1.48, 1.30 and 1.16, respectively). Additionally, the three genes were detected on HCC

tissue b immunohistochemistry and all of that had higher expression levels compared with control tissues (Fig.S3).

Table 1  
Three genes identified for construction of the prediction model

<b>Genes</b>	<b>Co-ef</b>	<b>HR</b>	<b>HR.95L</b>	<b>HR.95H</b>	<b>P-value</b>
ABCC1	0.262516	1.300197	1.053752	1.60428	0.014355
NPC1	0.39572	1.485454	1.074874	2.052867	0.016512
SLC51B	0.15545	1.168183	1.046335	1.304221	0.005677
<b>Genes</b>	<b>Co-ef</b>	<b>HR</b>	<b>HR.95L</b>	<b>HR.95H</b>	<b>P-value</b>
ABCC1	0.262516	1.300197	1.053752	1.60428	0.014355
NPC1	0.39572	1.485454	1.074874	2.052867	0.016512
SLC51B	0.15545	1.168183	1.046335	1.304221	0.005677

## Discussion

Though increasing publications supporting that enterohepatic circulation are involved in carcinogenesis of HCC, few reports have focused on the correlations between the genes involved in the transport and metabolism of BAs in the liver and HCC systematically [31–33]. In our work, based on TCGA data, we identified 26 BA-associated genes with differential expression between tumor and non-tumor tissues from 127 collected genes (Fig. 1). Next, we performed univariate and multivariate COX regression analyses and identified 11 genes and 3 genes, respectively, that were significant for the prognosis of HCC patients. These genes are implicated in the metabolism and transport of BAs, cholesterol and fatty acids (Fig. 2). Subsequently, we constructed a prediction model of survival of HCC patients using the three genes (NPC1, ABCC1 and SLC51B) from multivariate Cox regression. The model value (risk score) had higher confidence for predicting the prognosis of HCC patients than several tissue pathological characters like tumor staging, TNM, and grading (Fig. 5).

Within the list of 27 differentially expressed genes, there are 19 genes with high expression and 7 genes with low expression in HCC. Most of the down-regulated genes are related to catalytic enzymes, which participate in enzymatic steps of BA biosynthesis or upstream molecules like cholesterol and fatty acids. For instance, SLC27A5 is involved in the formation of conjugated BAs (glycine-/taurina-BAs), AKR1D1 catalyzes the key step in the alternative pathway of BA synthesis, and CYP39A1 takes part in the synthesis of cholesterol [7, 10, 34, 35]. Meanwhile, ACDL1 and BBOX1 engage in synthesis and transport of lipids in liver cells, respectively [36, 37]. Otherwise, the major up-regulated genes tend to encode transporter proteins including ABC family members, SLC family members and NPC1. These transporters are engaged in transporting BAs resorbed from portal vein blood into and out of liver cells [17]. Additionally, BAs and associated molecules also require transporters when shuttling into the different

compartments in cells. These transporters including NTCP (SLC10A1), OATP1 (SLC01A2), OATP2 (SLC21A6), OST- $\alpha$  (SLC51A), OST- $\beta$  (SLC51B), MRP3 (ABCC3), MPR4 (ABCC4), MRP2 (ABCC2), BSEP, MDR1A and others transporters involved in lipid and cholesterol metabolism.

After the exploration of BA-related genes and the prognosis of HCC patients, we identified 12 genes and 3 genes with univariate and multivariate Cox regression, respectively. Among the former 12 genes, four genes were negatively correlated with the prognosis of patients, and the other 8 genes were positively correlated. CYP7A1 and SLC27A5 are the most important genes controlling the two way, classic and alternative ways of BA synthesis [9–11, 32]. AKR1D1 is also a catalytic enzyme in the synthesis of BAs [10]. ALDH8A1 is involved in the metabolism of tryptophan, which can be linked to fatty acid and cholesterol [38]. Evidence suggests that evaluating the de novo synthesis of BAs are beneficial to HCC patients. At the same time, the 8 genes positively related to poor prognosis of HCC patients (ABCC1, ABCA3, ABCA4, SLC35B2, SLC51B, ABCD1 and NPC1) are all transporter proteins, and EFHC1 is a calcium regulation protein [39]. The results indicates that fluent transportation of BAs in the enterohepatic cycle potentially deteriorates the condition of HCC patients. NPC1, ABCC1 and SLC51B were determined by multivariate Cox regression and used to construct a prediction model for the prognosis of HCC patients. With this model, we can evaluate the risk score of patients. The three genes are a signature for the prediction of the prognosis of HCC patients.

Why are transporter genes, which facilitate the cycling of BAs, risk factors for HCC patients? BAs effectively facilitate digestion and absorption of lipids, nutrients, lipid-soluble vitamins and even drugs; they also regulate cholesterol homeostasis and could be recycled multiple times during the digestive phase [10, 40]. The de novo synthesis of BAs is initiated from cholesterol catalyzed by CYP7A1 or CYP27A1 (and CYP7B1) and subsequently in other enzymes, which then results in the modification of the side chain and steroid nucleus, and finally ends up with CA and CDCA [10, 17, 40]. Later, CA and CDCA is conjugated to glycine and taurine, forming glycine-taurina-cholic acid and glycine-taurina-chenodeoxycholic acid [32]. Then, gut bacteria convert the primary BAs to secondary BAs, DCA, UDCA and LCA [13, 32]. In hepatic sinusoids, BAs are transported into hepatocytes with transporter proteins, including NTCP and OATPs, which then sustain the balance of the pool of BAs in the liver [17]. Meanwhile, the lost part in faeces will be replenished by newly de novo synthesized BAs [32]. All BAs are discharged into the bile duct with transporters, such as MRP2 and MDR1A, and the cycle starts again [17].

Apart from the functions described above, BAs have been identified as signaling molecules involved in physical functions through binding to the receptors including a membrane receptor, Takeda G-protein receptor 5 (TGR5), and, a nuclear hormone receptor, farnesoid X receptor- $\alpha$  (FXRa) [17]. Although there are several other receptors that BAs can bind to, including sphingosine-1-phosphate receptor 2 (S1PR2), muscarinic receptors M2 and M3, formyl peptide receptor 1 (FMLP), liver X receptors  $\alpha$  and  $\beta$  (LXR  $\alpha\beta$ ), VDR (NR1H1), pregnane X receptor (PXR) and constitutive androstane receptor (CAR), TGR5 and FXRa are specific BAs and are expressed in liver cells at high levels [7, 41]. When BAs bind to and activate TGR5 and FXRa, a series of cascade signals introduce the physical orders into immune cells and hepatocytes

and then affect the metabolic routes, which include enzymes and transporters in BA metabolism, metabolism of lipids and glucose, and even inflammation and cancers [7, 17].

BAs have a close relationship with HCC. BAs engage with the carcinogenesis of HCC via direct activation of signaling pathways by binding to receptors to induce changes in the local immune microenvironment, change the metabolic status of liver cells and even result in DNA damage.

Deficiency of the FXR gene was found to increase proinflammatory cytokine IL-1 $\beta$  level and elevate oncogene c-Myc level [31]. FXR is distributed in hepatocytes and the FXR agonists can be CA, CDCA, LCA and DCA [41]. TGR5 is usually expressed on the cell surface of the ileum, gallbladder, adipose tissues and macrophages, and the two agonists in BAs are the secondary BAs, LCA and DCA [41]. Evidence supports that TGR5 seems to be an inflammatory suppressor [33]. Activation of TGR5 could suppress LPS-induced production of proinflammatory cytokines, and antagonize NF- $\kappa$ B signaling via prevention of phosphorylation of I $\kappa$ B $\alpha$  (an inhibitor of NF- $\kappa$ B translocation), which is the most important inflammatory pathway [33]. Additionally, cytokines and chemokines, including IL-1 $\beta$ , TNF $\alpha$ , IL-6, IFN and MCP-1, have evaluated levels as results of TGR5 deficiency [33]. FXR is also involved in and negatively regulates the NF- $\kappa$ B pathway, which has a close connection with liver chronic inflammation [33]. In summary, TGR5 and FXR play important roles in protecting the liver from inflammation, chronic injury and even cancer.

However, cholesterol and BAs are toxic to tissues and cells. Hydrophobic and hydrophilic BAs have different roles in stimulating cells. A new report showed that cholesterol accumulation in hepatocytes increased local inflammation and fibrosis through stabilizing TAZ, a transcription factor controlling the expression of genes associated with inflammation and fibrosis [42]. CYP7A1 drives BA metabolism, which can reduce the deposition of cholesterol in the liver.

DCA and CDCA induce oxidative DNA damage, which can then be involved in tumor initiation, and antioxidants could reduce the carcinogenic effects [43]. Additionally, DCA also increases nitrosative stress, resulting in damage to the cellular membrane and DNA [44]. The changes of the constitution of BAs are associated with liver fibrosis and, for example, decreased CA and increased CDCA are accompanied with non-alcoholic steatohepatitis (NASH), which maybe the results of alternated gut microbiota in patients [45]. The interaction of BAs and gut microorganisms is complicated for probe who is the causes. Meanwhile, alteration of composition of BAs or gut microbiota will affect liver situation dramatically.

In our present work, we observed two aspects of an apparent phenomenon; on one hand, up-regulated genes facilitating the transport of BAs are associated with poor prognosis of HCC patients, and on the other hand, de novo synthesis of BAs benefits HCC patients. The reasons may include: (1) reabsorbed BAs may hold carcinogens of HCC and promote carcinogenesis of HCC; (2) transporters facilitate the accumulation of BAs in hepatocytes, which can induce carcinoma development [46]; and (3) recycled BAs feedback to the synthesis of BAs, which may disrupt the balance of cholesterol and BA metabolism, thereby promoting the aberrant proliferation of liver cells [47]. At the same time, dysregulation of BA metabolism could activate inflammatory signalling in the liver and increase the risk of HCC development [48]. BA recycling could carry lots of metabolites of bacteria in the gut into the liver, and as a result, molecules of

the metabolites change the microenvironment of the liver, which leads to carcinogenesis [49]. Perspectively, more and more reports will emphasize the importance of the enterohepatic cycle in the development of liver diseases, including carcinoma.

## Abbreviations

HCC: Hepatocellular carcinoma; BA: bile acid; TCGA: The Cancer Genome Atlas portal; LPS: lipopolysaccharides; SCFA: short chain fatty acids; CA: cholic acid; CDCA: chenodeoxycholic acid; DCA: deoxycholic acid; LCA:lithocholic acid; UDCA: ursodeoxycholic acid; CYP7A1: cholesterol 7 $\alpha$ -hydroxylase; CYP27A1: sterol 27A-hydroxylase; CYP7B1: oxysterol and steroid 7 $\alpha$ -hydroxylase; ASBT: apical sodium-dependent bile acid transporter; OST $\alpha$ /OST $\beta$ : organic solute transporter  $\alpha$  and  $\beta$  heterodimer; NTCP: Na $^+$  taurocholate cotransporting polypeptide; OATPs: organic anion transporting polypeptide; MRP: multidrug resistance; BSEP: bile salt export pump; ABC: ATP-binding cassette; GSEA : Gene Set Enrichment Analysis; GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; ROC: receiver-operator characteristic curve; NPC1: Niemann pick c1; SLC: solute carrier.

## Declarations

### *Ethics approval and consent to participate*

Not applicable.

### *Consent for publication*

Not applicable.

### *Availability of data and materials*

All data generated or analyzed during this study are included in this published article and its supplementary information files.

### *Competing interests*

The authors have no competing interests to declare.

### *Funding*

This work was supported by grants: the National Science and Technology Major Project of China (2017ZX10202203) and the Open Research Fund of Zhejiang First-foremost Key Subject-Chinese medicine, Zhejiang Chinese Medicine University.

### *Author Contributions*

G.W and Z.C raised the idea of the review; G.W, J.G, Q.Y processed the data. G.W, J.G, Q.Y, F.W, Z.G, Y.R and H.Z prepared the writing of the manuscript; Z.C and H.Z proofread and edited the manuscript.

## Acknowledgements

Not Applicable.

## Author details

<sup>1</sup> State Key Laboratory for Diagnosis and Treatment of Infectious Diseases, National Clinical Research Center for Infectious Diseases, Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, The First Affiliated Hospital, College of Medicine, Zhejiang University, Hangzhou, China

## References

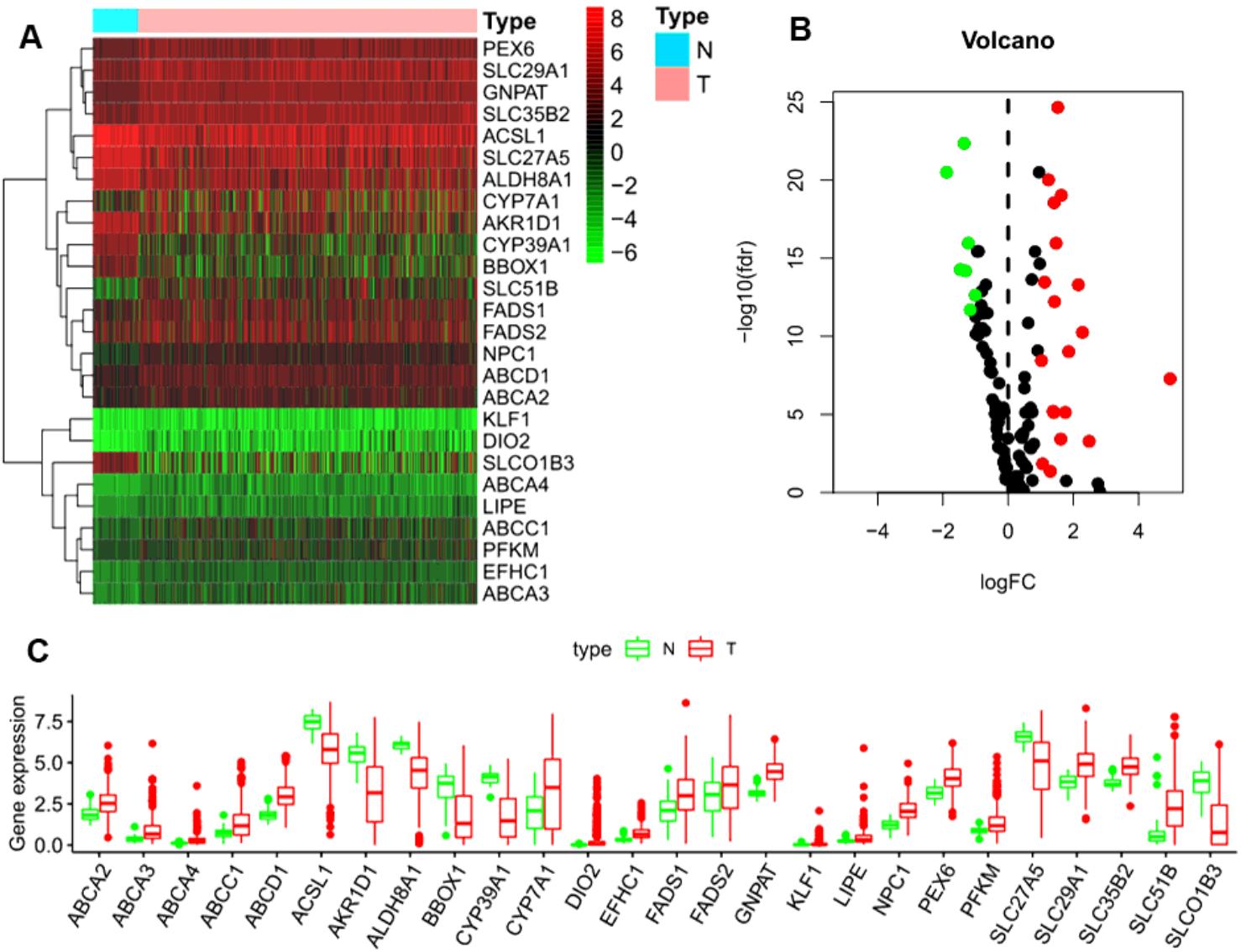
1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *Cancer J Clin.* 2018;68(6):394–424.
2. Chen W, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, Jemal A, Yu XQ, He J. Cancer statistics in China, 2015. *Cancer J Clin.* 2016;66(2):115–32.
3. Cronin KA, Lake AJ, Scott S, Sherman RL, Noone AM, Howlader N, Henley SJ, Anderson RN, Firth AU, Ma J, et al: **Annual Report to the Nation on the Status of Cancer, part I: National cancer statistics.** *Cancer* 2018, **124**(13):2785–2800.
4. Ma C, Han M, Heinrich B, Fu Q, Zhang Q, Sandhu M, Agdashian D, Terabe M, Berzofsky JA, Fako V, et al: **Gut microbiome-mediated bile acid metabolism regulates liver cancer via NKT cells.** *Science* 2018, **360**(6391).
5. Effenberger M, Tilg H. The intestinal microbiota and hepatocellular carcinoma. *memo-Magazine of European Medical Oncology.* 2020;13(2):223–6.
6. Hofmann AF. The continuing importance of bile acids in liver and intestinal disease. *Arch Intern Med.* 1999;159(22):2647–58.
7. Martinot E, Sedes L, Baptissart M, Lobaccaro JM, Caira F, Beaudoin C, Volle DH. Bile acids and their receptors. *Mol Aspects Med.* 2017;56:2–9.
8. Russell DW. Fifty years of advances in bile acid synthesis and metabolism. *Journal of lipid research.* 2009;50 **Suppl**:120–5.
9. Vaz FM, Ferdinandusse S. Bile acid analysis in human disorders of bile acid biosynthesis. *Mol Aspects Med.* 2017;56:10–24.
10. Sarenac TM, Mikov M. Bile Acid Synthesis: From Nature to the Chemical Modification and Synthesis and Their Applications as Drugs and Nutrients. *Front Pharmacol.* 2018;9:939.
11. Staels B, Fonseca VA. Bile acids and metabolic regulation: mechanisms and clinical responses to bile acid sequestration. *Diabetes Care.* 2009;32(Suppl 2):237–45.
12. Begley M, Gahan CG, Hill C. The interaction between bacteria and bile. *FEMS MicroBiol Rev.* 2005;29(4):625–51.

13. Dawson PA, Karpen SJ. Intestinal transport and metabolism of bile acids. *Journal of lipid research*. 2015;56(6):1085–99.
14. Trauner M, Boyer JL. Bile salt transporters: molecular characterization, function, and regulation. *Physiological reviews*. 2003;83(2):633–71.
15. Cruz-Ramon V, Chinchilla-Lopez P, Ramirez-Perez O, Mendez-Sanchez N. Bile Acids in Nonalcoholic Fatty Liver Disease: New Concepts and Therapeutic Advances. *Annals of hepatology*. 2017;16(Suppl 1):58–67.
16. Vitek L. Bile Acids in the Treatment of Cardiometabolic Diseases. *Annals of hepatology*. 2017;16(Suppl 1):43–52.
17. Thomas C, Pellicciari R, Pruzanski M, Auwerx J, Schoonjans K. Targeting bile-acid signalling for metabolic diseases. *Nature reviews Drug discovery*. 2008;7(8):678–93.
18. Jia B, Jeon CO. Promotion and induction of liver cancer by gut microbiome-mediated modulation of bile acids. *PLoS pathogens*. 2019;15(9):e1007954.
19. Sarenac T, Mikov M. Cervical Cancer, Different Treatments and Importance of Bile Acids as Therapeutic Agents in This Disease. *Front Pharmacol*. 2019;10:484.
20. Wu YC, Chiu CF, Hsueh CT, Hsueh CT. The role of bile acids in cellular invasiveness of gastric cancer. *Cancer cell international*. 2018;18:75.
21. Ocvirk S, O'Keefe SJ. Influence of Bile Acids on Colorectal Cancer Risk: Potential Mechanisms Mediated by Diet - Gut Microbiota Interactions. *Current nutrition reports*. 2017;6(4):315–22.
22. Zeng H, Umar S, Rust B, Lazarova D, Bordonaro M. **Secondary Bile Acids and Short Chain Fatty Acids in the Colon: A Focus on Colonic Microbiome, Cell Proliferation, Inflammation, and Cancer.** *International journal of molecular sciences* 2019, 20(5).
23. Lee W, Um J, Hwang B, Lee YC, Chung BC, Hong J. Assessing the progression of gastric cancer via profiling of histamine, histidine, and bile acids in gastric juice using LC-MS/MS. *J Steroid Biochem Mol Biol*. 2019;197:105539.
24. Reimand J, Isserlin R, Voisin V, Kucera M, Tannus-Lopes C, Rostamianfar A, Wadi L, Meyer M, Wong J, Xu C, et al. Pathway enrichment analysis and visualization of omics data using g:Profiler, GSEA, Cytoscape and EnrichmentMap. *Nature protocols*. 2019;14(2):482–517.
25. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *J R Stat Soc B*. 1995;57(1):289–300.
26. Asomaning N, Archer KJ. High-throughput DNA methylation datasets for evaluating false discovery rate methodologies. *Comput Stat Data Anal*. 2012;56(6):1748–56.
27. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet*. 2000;25(1):25–9.
28. Kanehisa M, Goto S, Kawashima S, Nakaya A. The KEGG databases at GenomeNet. *Nucleic acids research*. 2002;30(1):42–6.

29. Chen B, Huang S, Pisanic TR, Stark A, Tao Y, Cheng B, Li Y, Wei Y, Zhao W, Wang T-H. Rab8 GTPase regulates Klotho-mediated inhibition of cell growth and progression by directly modulating its surface expression in human non-small cell lung cancer. *EBioMedicine*. 2019;49:118–32.
30. Pons F, Varela M, Llovet JM. Staging systems in hepatocellular carcinoma. *Hpb*. 2005;7(1):35–41.
31. Jansen PL. Endogenous bile acids as carcinogens. *Journal of hepatology*. 2007;47(3):434–5.
32. Chiang JYL, Ferrell JM. Bile Acid Metabolism in Liver Pathobiology. *Gene Expr*. 2018;18(2):71–87.
33. Wang X, Fu X, Van Ness C, Meng Z, Ma X, Huang W. Bile Acid Receptors and Liver Cancer. *Current pathobiology reports*. 2013;1(1):29–35.
34. Chen M, Drury JE, Penning TM. Substrate specificity and inhibitor analyses of human steroid 5beta-reductase (AKR1D1). *Steroids*. 2011;76(5):484–90.
35. Steinberg SJ, Mihalik SJ, Kim DG, Cuevas DA, Watkins PA. The human liver-specific homolog of very long-chain acyl-CoA synthetase is cholate:CoA ligase. *J Biol Chem*. 2000;275(21):15605–8.
36. Hirche F, Fischer M, Keller J, Eder K. Determination of carnitine, its short chain acyl esters and metabolic precursors trimethyllysine and gamma-butyrobetaine by quasi-solid phase extraction and MS/MS detection. *Journal of chromatography B Analytical technologies in the biomedical life sciences*. 2009;877(22):2158–62.
37. Ohkuni A, Ohno Y, Kihara A: **Identification of acyl-CoA synthetases involved in the mammalian sphingosine 1-phosphate metabolic pathway.** *Biochemical and biophysical research communications* 2013, **442**(3–4):195–201.
38. Miyazawa S, Sakurai T, Shindo Y, Imura M, Hashimoto T. The effect of tryptophan administration on fatty acid synthesis in the livers of rats under various nutritional conditions. *J BioChem*. 1975;78(1):139–47.
39. Suzuki T, Delgado-Escueta AV, Aguan K, Alonso ME, Shi J, Hara Y, Nishida M, Numata T, Medina MT, Takeuchi T, et al. Mutations in EFHC1 cause juvenile myoclonic epilepsy. *Nat Genet*. 2004;36(8):842–9.
40. Chiang JY. Bile acids: regulation of synthesis. *Journal of lipid research*. 2009;50(10):1955–66.
41. Fiorucci S, Distrutti E. The Pharmacology of Bile Acids and Their Receptors. *Handb Exp Pharmacol*. 2019;256:3–18.
42. Wang X, Cai B, Yang X, Sonubi OO, Zheng Z, Ramakrishnan R, Shi H, Valenti L, Pajvani UB, Sandhu J, et al. Cholesterol Stabilizes TAZ in Hepatocytes to Promote Experimental Non-alcoholic Steatohepatitis. *Cell Metabol*. 2020;31(5):969–86 e967.
43. Rosignoli P, Fabiani R, De Bartolomeo A, Fuccelli R, Pelli MA, Morozzi G. Genotoxic effect of bile acids on human normal and tumour colon cells and protection by dietary antioxidants and butyrate. *Eur J Nutr*. 2008;47(6):301–9.
44. Dall'Agnol M, Bernstein C, Bernstein H, Garewal H, Payne CM. Identification of S-nitrosylated proteins after chronic exposure of colon epithelial cells to deoxycholate. *Proteomics*. 2006;6(5):1654–62.

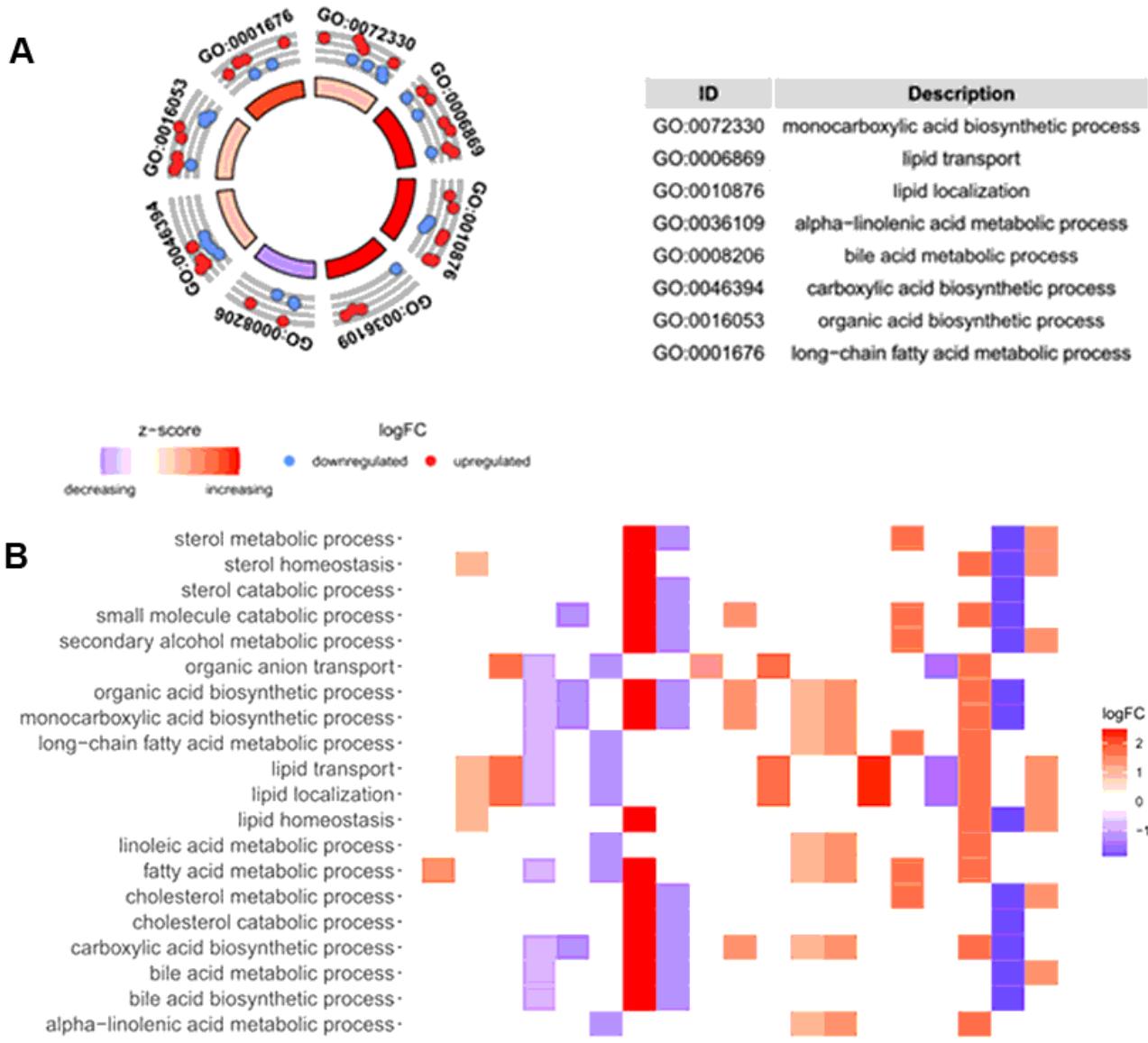
45. Sydor S, Best J, Messerschmidt I, Manka P, Vilchez-Vargas R, Brodesser S, Lucas C, Wegehaupt A, Wenning C, Assmuth S, et al. Altered Microbiota Diversity and Bile Acid Signaling in Cirrhotic and Noncirrhotic NASH-HCC. *Clinical translational gastroenterology*. 2020;11(3):e00131.
46. Lozano E, Sanchez-Vicente L, Monte MJ, Herraez E, Briz O, Banales JM, Marin JJ, Macias RI. Cocarcinogenic effects of intrahepatic bile acid accumulation in cholangiocarcinoma development. *Molecular cancer research: MCR*. 2014;12(1):91–100.
47. Manieri E, Folgueira C, Rodriguez ME, Leiva-Vega L, Esteban-Lafuente L, Chen C, Cubero FJ, Barrett T, Cavanagh-Kyros J, Seruggia D, et al: **JNK-mediated disruption of bile acid homeostasis promotes intrahepatic cholangiocarcinoma**. *Proc Natl Acad Sci U S A* 2020.
48. Sun L, Beggs K, Borude P, Edwards G, Bhushan B, Walesky C, Roy N, Manley MW Jr, Gunewardena S, O'Neil M, et al. Bile acids promote diethylnitrosamine-induced hepatocellular carcinoma via increased inflammatory signaling. *American journal of physiology Gastrointestinal liver physiology*. 2016;311(1):G91–104.
49. Jia W, Xie G, Jia W. Bile acid-microbiota crosstalk in gastrointestinal inflammation and carcinogenesis. *Nature reviews Gastroenterology hepatology*. 2018;15(2):111–28.

## Figures



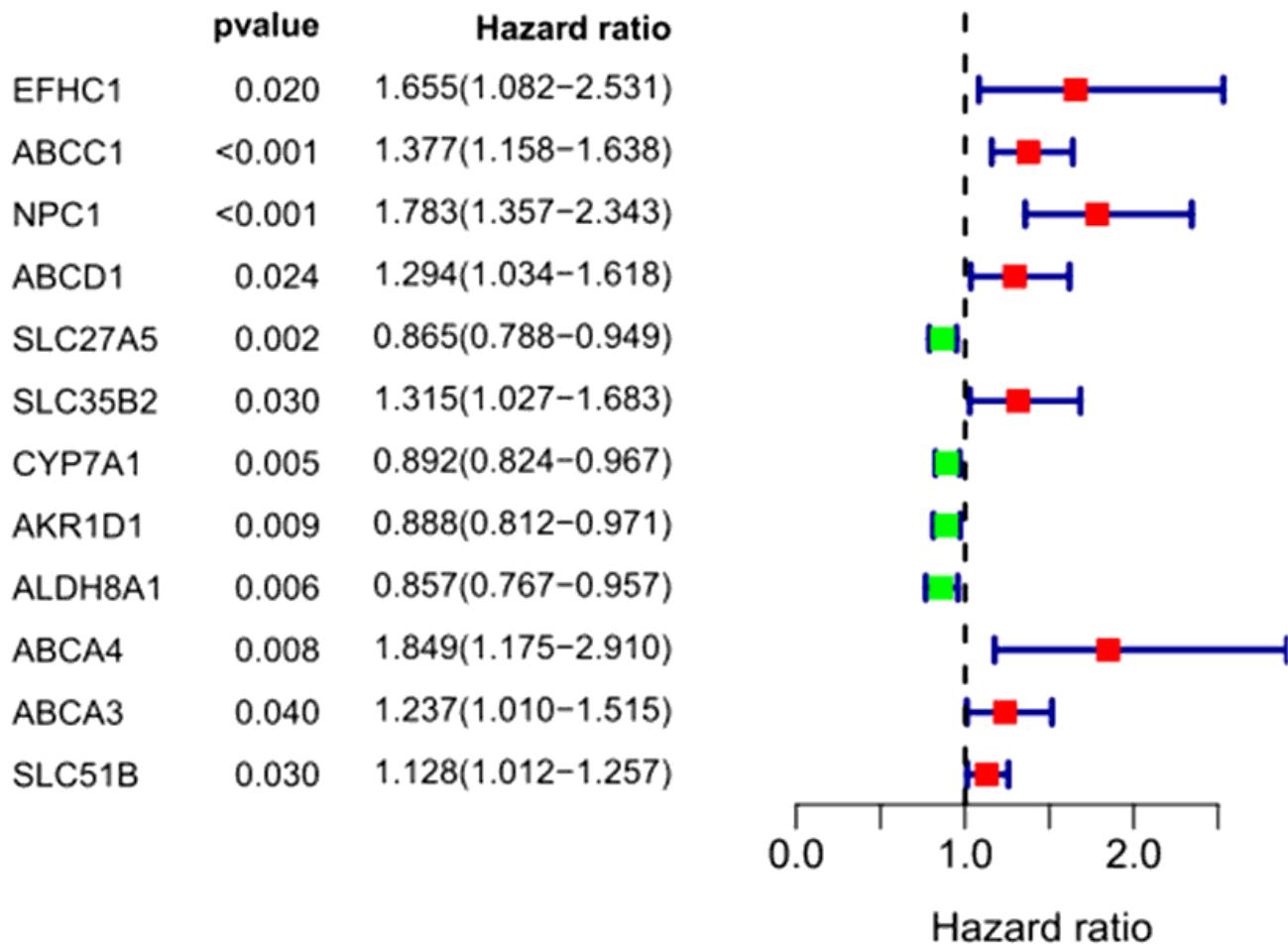
**Figure 1**

Bile acid-related genes with differential expression in tumor and non-tumor samples Twenty-six genes were identified with differential expression (fold-change  $> |2|$ ) (A). Among these genes, 19 genes were up-regulated in tumor samples and 7 genes were up-regulated in non-tumor samples ( $P < 0.05$ ) (B and C).



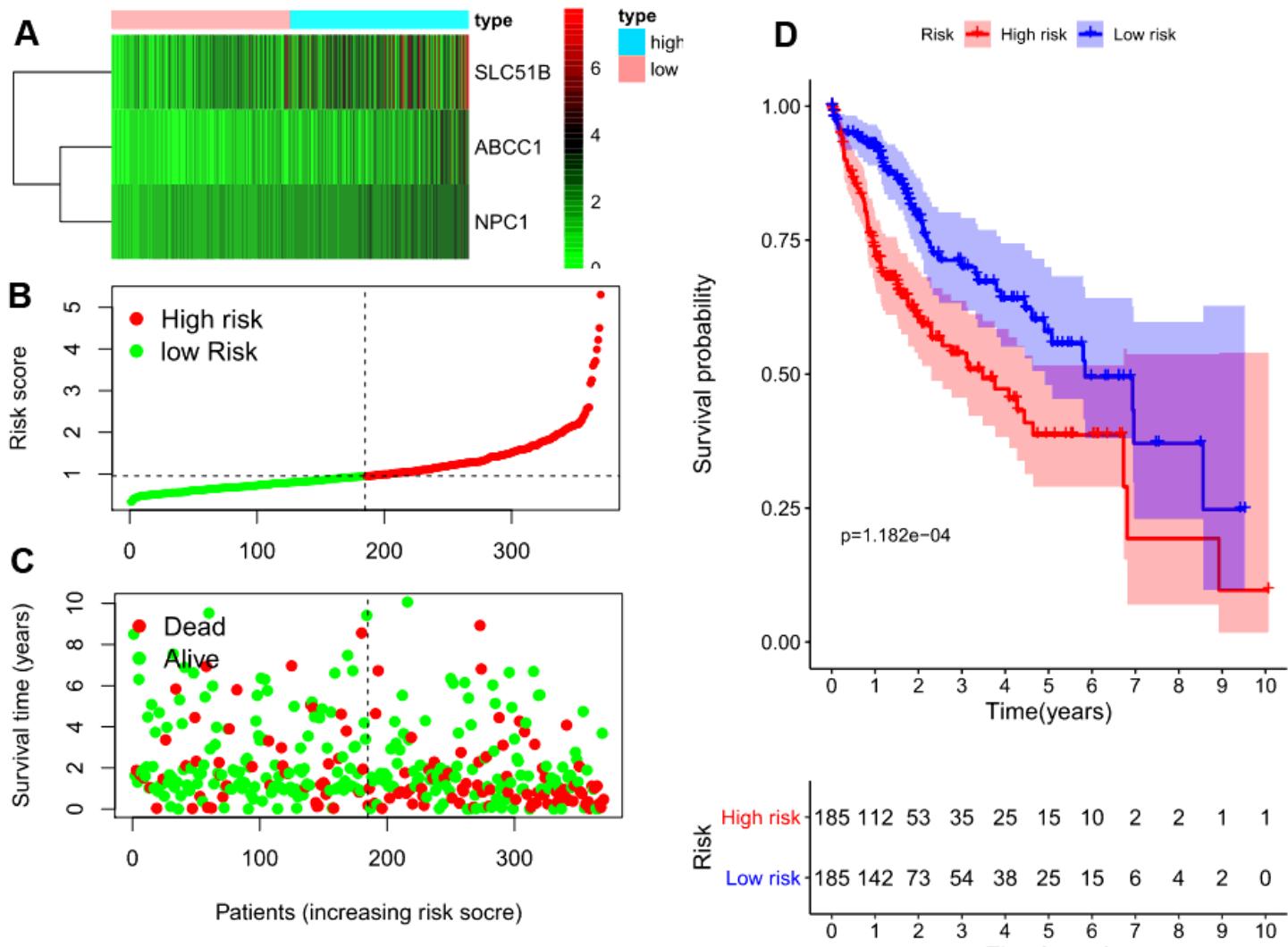
**Figure 2**

Annotation and enrichment of the 26 identified genes GO analysis showed that genes were involved in metabolism of sterol, monocarboxylic acid, fatty acid, cholesterol and bile acid, and transport of lipid, organic anion and bile acid (A); KEGG analysis showed that most of the genes contributed to ABC transporter family, and were involved in the PPAR signaling pathway and the synthesis of bile acid, cholesterol and fatty acids (B,C).



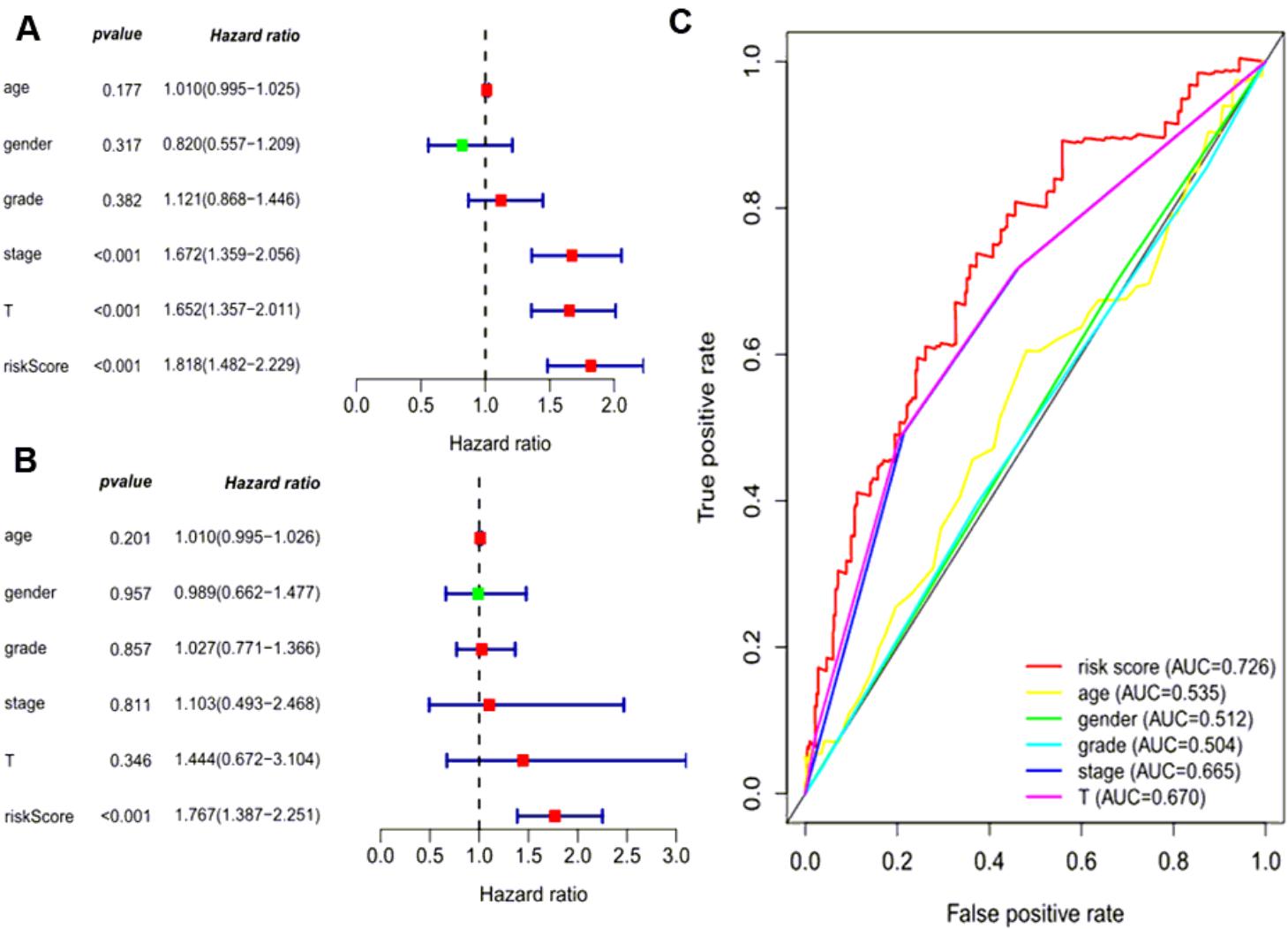
**Figure 3**

Identification of survival-related genes related to bile acid metabolism. Twelve genes were identified to have relationships with the prognosis of HCC patients. Among these genes, SLC27A5, CYP7A1, AKR1D1 and ALDH8A1 showed benefits to the patients, whereas EFHC1, ABCC1, NPC1, ABCD1, SLC35B2, ABCA4, ABCA3 and SLC51B showed increased risk of poor prognosis of the patients.



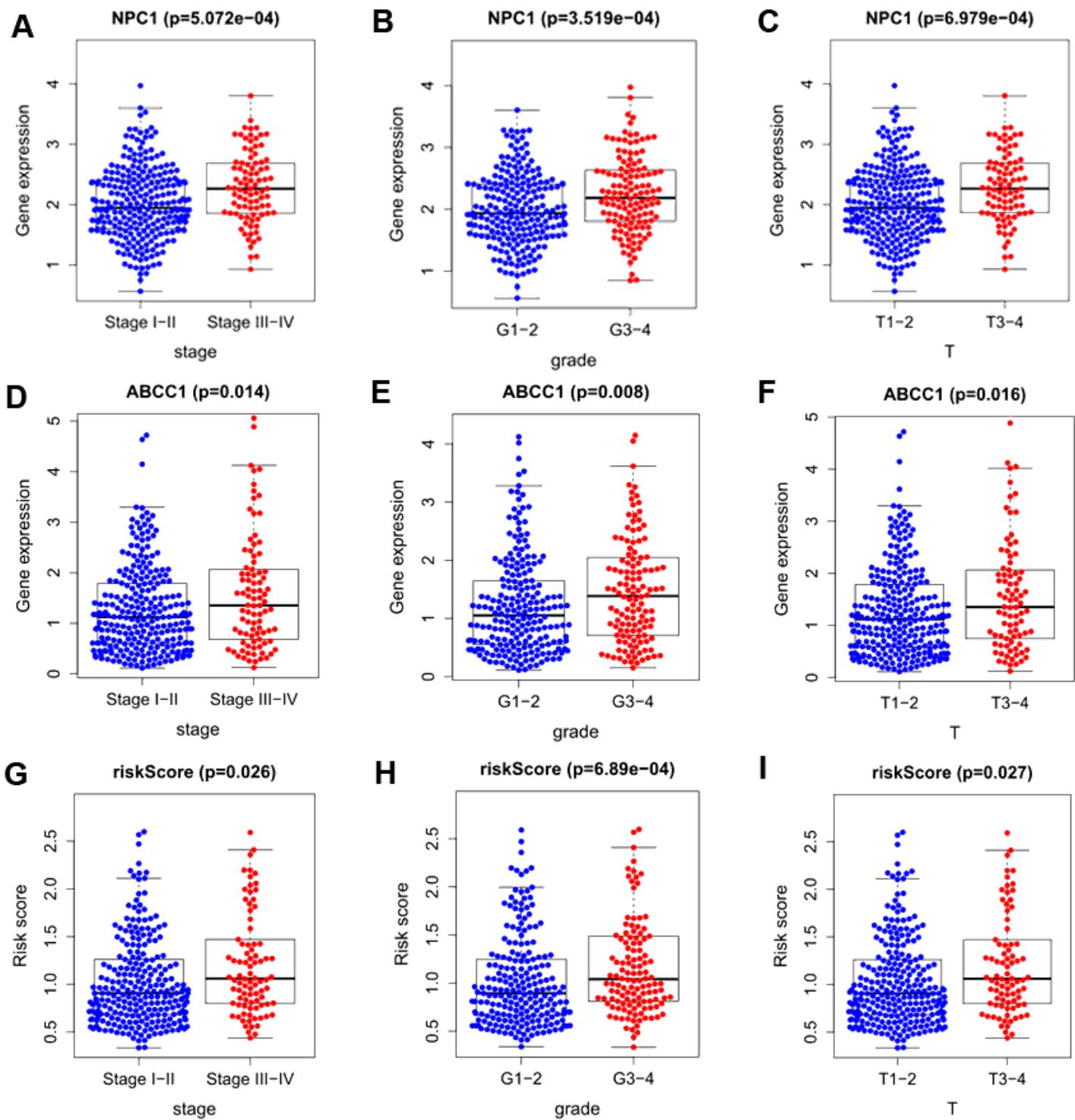
**Figure 4**

The three genes used for construction of a prediction model for the prognosis of HCC patients. The expression of NPC1, ABCC1 and SLC51B in tumor and non-tumor samples (A). The distribution of HCC patients with different risk scores (B). The status of HCC patients with increasing risk score (C). The survival curve of the two groups of HCC patients with high and low risk score.



**Figure 5**

Risk score and clinical characters on the prognosis of HCC patients. Univariate and multivariate Cox regression analysis of risk score, age, gender, grade, stage and T (TNM) for the prognosis of HCC patients ( $P < 0.001$ , Hazard ratio = 1.568, 95% CI: 1.256-1.956) (A, B). ROC curve of risk score, age, gender, grade, stage and T (TNM) displayed risk score was more reliable for prediction of the survival of HCC patients ( $AUC = 0.710$ ) (C).



**Figure 6**

Correlation between model genes and clinical characters. **NPC1** had different expression levels in (A) different stages (stage III-IV was higher,  $P < 0.001$ ), (B) grade (grade 3-4 was higher,  $P < 0.001$ ) and (C) TNM (T3-4 was higher,  $P < 0.001$ ). **ABCA4** in high (D) grade, (E) stage and (F) T had higher expression ( $P = 0.014$ ,  $P = 0.008$  and  $P = 0.016$ , respectively). The patients with higher risk scores had advanced pathological staging, including (G) stage, (H) grade and (I) T ( $P = 0.026$ ,  $P < 0.001$ , and  $P = 0.027$ , respectively).

# Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementaryfile.docx](#)
- [Fig.S1.tif](#)
- [Fig.S2.tif](#)
- [Fig.S3IHC.tif](#)