

# Transcriptomic analysis and Novel gene pair-based signatures for Hepatitis B-related hepatocellular carcinoma

**Biao Song**

Shanghai Medical University: Fudan University

**Yulin Wang**

Shanghai Medical University: Fudan University

**Shaocong Mo** (✉ [msc245@foxmail.com](mailto:msc245@foxmail.com))

Shanghai Medical University: Fudan University <https://orcid.org/0000-0001-8556-1314>

---

## Research

**Keywords:** hepatitis B virus, hepatocellular carcinoma, gene pair, signature, diagnosis, prognosis

**Posted Date:** April 2nd, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-350837/v1>

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

[Read Full License](#)

---

# Abstract

**Background:** Hepatitis B virus (HBV)-related hepatocellular carcinoma (HCC) misses the opportunity for surgery because it is not detected early. The molecular mechanism of hepatitis B-related liver cancer needs further understanding, and effective diagnostic and prognostic models are in urgent need.

**Methods:** Expression profiles from the Cancer Genome Atlas (TCGA) Liver Hepatocellular Carcinoma (LIHC), GSE121248, GSE94660 GSE76724 from Gene Expression Omnibus (GEO) database were obtained. Differentially expressed genes (DEGs) between normal and tumor HBV-related HCC samples based on GSE121248 and GSE94660. Gene pairs are generated by comparing the expression levels of every two DEGs. A diagnostic signature of pairs of DEGs was built using cross-validation Lasso and Best Subset Selection regression. Hub genes and significant modules were screened by Cytoscape, and potential drugs were predicted by DGIdb. A prognostic signature was established and xCell and ssGSEA were utilized to reveal the cell composition and cancer hallmarks to get an elucidation for the risk.

**Results:** 457 DEGs were screened. A powerful diagnostic signature of 2 pairs of DEGs was built and validated in TCGA-LIHC and GEO datasets repeatedly with assured performance. 10 Hub genes were found and fostamatinib was predicted to have potential therapeutic effect on HBV-related HCC. A prognostic signature with good efficiency (Log-rank P value<0.05, AUC=93.1%) were established, with stromal score and several hallmarks related to the risk

**Conclusion:** Taken together, the study provided sight into the molecular mechanism as well as a novel strategy for the early diagnosis and prognosis for HBV-related HCC.

## 1. Introduction

Among all cancer-related death cases throughout the world, hepatocellular carcinoma (HCC) is the third-commonest cause, only after lung cancer and breast cancer[1–3]. Although many treatments for HCC have been introduced, including radiation therapy, ethanol ablation and transcatheter arterial chemoembolization, the overall survival rate of 5-year remains unsatisfactory, ranging from 40–75%[4].

Currently, the occurrence of liver cancer is considered to be the result of multiple factors, including genetic susceptibility, environmental factors and viral infection, etc[5]. One of the risk factors that has been universally accepted is Chronic hepatitis B virus infection, especially in China where HBsAg carrier ratio among patients aged 1–60 is around 7.18%[6, 7]. Therefore, identifying the specific genetic characteristics and variation of patients with chronic HBV induced HCC is a key task for us. However, few studies have proposed a comprehensive molecular mechanism of hepatitis B inducing HCC.

Additionally, early diagnosis and prognosis of HBV-related HCC was important, since in the early stages of HCC, liver transplantation or liver resection can be used to remove tumor tissue to obtain a better prognosis, while these methods are not suitable for advanced-stage of HCC because of metastasis[8–10]. However, early diagnosis and prognosis of HBV-related liver cancer is still difficult although

progressions had been made. Therefore, it is critical to find more effective biomarkers of HBV-related HCC for early diagnosis and prognosis.

With the help of high-throughput sequencing and methylation microarray technology, it has become more convenient and common to identify HCC-related differentially expressed genes (DEGs), which may help us to understand the specific mechanisms of HCC progress and provide an effective way to detect potential targets in diagnosing and outcome prediction for HCC in early stages. In this study, we applied bioinformatics approaches to study the microarray data, RNA-Seq data imported from GEO database as well as LIHC expression profile from TCGA database. Multiple bioinformatics analyses were performed, such as Gene Ontology (GO) analysis, Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis, PPI network construction and survival analysis. Drugs targeting on HBV-related HCC were predicted. Innovatively, A diagnostic signature composed of gene pairs and a prognostic signature composed of 4 gene pairs were established. And the cellular composition differences and cancer hallmarks were comprehensively analyzed.

Our aim was to identify hub genes and vital pathways in chronic HBV-related HCC and then to establish an both diagnostic and prognostic models for HBV-related HCC.

## 2. Materials And Methods

Workflow of the present study was depicted in the Fig. 1.

### 2.1 Data Sources

GSE121248, a microarray dataset containing gene expression information of 70 chronic HBV-related HCC and 37 adjacent normal tissues, was obtained from GEO database, which we only select the paired 37 tissues. GSE94660, RNA-sequencing of 21 HCC patients' tumor and normal paired tissue was also downloaded from GEO database in the form of normalized Reads Per Kilobase per Million mapped reads (RPKM). GSE76724, a microarray dataset containing gene expression information of 115 HCC (46% with HBV infection and 54% with cirrhosis) and 52 adjacent non-tumor tissues, was also obtained from GEO. The Cancer Genome Atlas (TCGA) RNA-seq data of LIHC in the form of Fragments Per Kilobase of exon model per Million mapped fragments (FPKM) was downloaded from UCSC Xena (<https://xenabrowser.net/datapages/>), which was immediately converted into Transcripts Per Kilobase of exon model per Million mapped reads (TPM). Detailed clinical data of TCGA-LIHC patients were downloaded by *TCGAbiolinks* package, patients who carried HBV but did not have other HCC risk factors such as HCV, alcohol or non-alcoholic fatty liver disease etc. were included. Survival data was accessed from UCSC Xena.

### 2.2 Identification of Differentially expressed genes (DEGs)

Differential gene expression analysis between normal and tumor tissues of GSE121248 and GSE94660 was performed using *limma* package, with a paired testing applied. After obtaining the degree of

difference of all genes, we used the cutoff criteria that genes satisfying  $|\log\text{Foldchange}| \geq 1$  and adjusted p-value  $< 0.05$  defined as the differential expression genes (DEGs), which were then intersected to get the common DEGs of two datasets.

### 2.3 Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis and single cell sGene Set Enrichment Analysis (ssGSEA).

GO ([www.geneontology.org](http://www.geneontology.org)) and KEGG ([www.genome.jp/](http://www.genome.jp/)) pathway enrichment analyses are two common methods for large-scale or systematic functional enrichment analyses[11, 12]. The David Database (<https://david.ncifcrf.gov/>) was applied to conduct GO enrichment and the *ClusterProfiler* package was used to conduct KEGG enrichment; the adjusted P value  $< 0.05$  was set and only the term whose the number of genes enriched greater than 10 were included in the results. The GO enrichment result was visualized by *GOpot*[13]. ssGSEA were conducted by GSVA package and visualized by heatmap, with gene sets obtained from MSigDb (<http://www.gsea-msigdb.org/>). Besides, xCell, a novel tool for detailed cellular composition analysis used in our study, were also based on ssGSEA[14].

### 2.4 Least absolute shrinkage and selection operator (LASSO) regression and Best subset selection regression

LASSO regression is a biased estimate for processing multicollinearity data, which constructs a penalty function to obtain more refined model and compresses regression coefficients to avoid overfitting. Here we first listed all the gene pairs obtained from DEGs to eliminate the batch effect: expression level of every two genes (A and B) were compared, 0 represents A lower than B and 1 represented A higher than B. Only the pairs whose comparison was consistent in more than 80% of the samples could be included in the variable selection[15]. 10 folds Cross validation Lasso regression was implemented on the gene pairs by *glmnet* package in R. The gene pairs with coefficients not equal to zero continued to have a Best subset selection regression by *bestglm* package to further decrease the variates[16]. Starting from the null model (M0), which has only the intercept term but no independent variables, different variates combinations are used for fitting, that is, from model M1 containing 1 variate, M2 containing 2 variates, to Mp containing p variates. Then the best model is selected (according to BIC) from the total of  $p + 1$  models. The variates configured by this best model are the filtered variates. Finally, gene pairs selected were included in a logistic regression. And we performed classification prediction using the *predict* function in the training set and validation set.

### 2.5 Construction of PPI network and identification of hub genes

For the purpose of exploring the protein interactions among DEGs, the PPI network was constructed with the help of STRING, a web-based tool (<http://www.string-db.org>). DEGs identified were imported into STRING and the interaction pairs between proteins were put out when their combined score  $> 0.7$ , which according to the website, had a high confidence. Afterwards, we used the Cytoscape software to conduct the visualization of the network. Furthermore, to obtain a clearer and more focused view of the PPI

network, MCODE was applied to select the significant modules, then degree of genes were sorted by Cytohubba, a plugin in Cytoscape, and the top 10 genes were considered as hub genes.

## 2.6 Drug-Gene interaction prediction

DGIdb ([https://www.dgidb.org/search\\_interactions](https://www.dgidb.org/search_interactions)) was used to predict the drug-gene interaction to provide potential therapies for HBV induced HCC[17]. Firstly, hub genes which had a significant prognostic value was input into the database. The drugs output should be the ones that had been approved and had a clear pharmacological effect (inhibitor or activator).

## 2.7 Kaplan–Meier analysis and Cox regression

The prognostic values of hub genes in chronic HBV-related HCC patients were evaluated in the K–M plotter (<http://kmplot.com/analysis/>). A total of 150 hepatitis virus positive patients were analyzed for overall survival. The univariate Cox regression and multivariate Cox regression for the gene pairs were conducted in R by *survival* and *survminer*. Gene pairs which had a significant prognostic value ( $P < 0.001$ ) firstly underwent Lasso-Cox regression and then backward stepwise regression was implemented to shrink the variates number. Then the patients' risk was calculated using predict function. A K-M analysis was done in R. At last, the ROC curve was drawn by *pROC* package.

## 2.8 Statistical analysis

Statistical analysis was completed by R 4.0.3 and other online tools. The identification of DEGs was based on the built-in *limma* package. Briefly, the genes were first tested by paired t test and then underwent a Bayesian test. Paired t test was to compare the expression level of hub genes in TCGA database. Cross-validation Lasso regression, Best subset selection regression and Logistic regression were used to establish diagnostic model. Log rank test were applied for K-M analysis. Univariate and multivariate Cox regression were used to select prognostic pairs and establish a signature.

## 3. Results

### 3.1 Identification of DEGs and function annotations

GSE121248, containing 37 chronic HBV-related HCC samples and matched 37 normal samples and GSE94660, which contained 21 HBV-HCC patients' non-neoplastic liver tissues and matched tumor tissues were selected to undergo differentially expressed analysis with threshold set to adjusted P value  $< 0.05$  and  $|\log FC| \geq 1.5$ , 817 DEGs were identified in GSE121248 and 1857 in GSE94660. After intersection, 457 genes were regarded as common DEGs in two sets (Fig. 2a). KEGG and GO pathway enrichment analysis towards DEGs were carried out by *clusterProfiler* and DAVID v6.8 respectively. KEGG pathway enrichment indicated that DEGs played a role in Chemical carcinogenesis, Cell cycle, Retinol metabolism and Fatty acid degradation, etc. (Fig. 2b). And for the biological process (BP) in GO enrichment, it was shown that DEGs mainly enriched in oxidation – reduction process, cell division (mainly upregulated DEGs) and metabolic process (mainly downregulated DEGs) (Table S1, Fig. 2c). Generally, we

demonstrated that the DEGs of HBC-related HCC patients mainly involved in two aspects of pathways, one is cell cycle and the other is metabolism.

### 3.2 A diagnostic signature composed of gene pairs for HBV-related HCC patients

To make the DEGs identified available for diagnostic usage, such as pathologies difficult to diagnose, we sorted to establish an effective model based on DEGs for suspicious HBV-HCC patients. In order to avoid the impact of batch effects when building and verifying the model, we adopt a novel approach that is the gene pairs to replace genes. After being intersected between two groups, a total of 25286 gene pairs were identified to construct the signature (With the criteria set to 0.2 between 0.8, all of these gene pairs were not consistent within samples). Cross-validation Lasso regression in GSE121248 (training cohort) identified 13 gene pairs with classification value based on the lambda.min selected (Fig. 3a, 3b, **Table S2**). Then the 13 gene pairs underwent best subset selection regression which shrink the variate number into 2, *RGS5*|*FAHD2A*, *CXCL14*|*SAMD5*. Next, logistic regression was applied to get the gene pairs' coefficients (Table 1). The diagnostic model was first to predict the classification in the training cohort itself, which presented an excellent performance of classification with sensitivity, specificity both equaled to 100% (Fig. 3c, 3d).

Table 1  
Gene pairs selected by best subset regression and logistic regression

Variates	Estimate
(Intercept)	24.09178
RGS5 FAHD2A	50.58907
CXCL14 SAMD5	-49.88468

Then, the signature was applied to predict the classification of the first validation cohort, GSE94660, which also showed a quite good efficiency (AUC = 90.5%, 95%CI = 81.87%-99.08%) (Fig. 3e, 3f). Then, in order to verify the signature again, we applied the model on the HCC patients with HBV infection in TCGA-LIHC cohort which contained 6 tumor and 6 normal samples. The result indicated that our signature indeed had a good diagnostic effect (ROC-AUC = 100%) (Fig. 3g, 3h).

To explore if our signature had a good performance on the HCC patients but not HBV-related HCC, we then applied the model on GSE76724, which showed that the performance of the model was slightly weakened, but still in a higher quality (ROC-AUC = 86.6%, 95%CI = 81.07%-92.06%) (Fig. 3i, 3j). In a word, the gene pair-based diagnostic model had a good performance both in training and validation cohort.

### 3.3 Construction of PPI network and identification of hub genes

STRING was applied to perform a PPI network of DEGs, and 308 nodes and 1943 edges were depicted in the network map. The network was then visualized by Cytoscape. MCODE was applied to identify crucial gene modules in the network. 4 modules with highest score were selected (Fig. 4a). Genes in each clusters were input into string again for functional enrichment, which were shown in Table 2 (Table 2).

Table 2  
KEGG and Reactome enrichment for 4 clusters identified by MCODE

Cluster	Nodes/edges	KEGG	Reactome
1	49/1160	hsa04110	HSA-1640170
		Cell cycle	Cell Cycle
2	11/48	hsa01100	HSA-211897
		Metabolic pathways	Cytochrome P450 - arranged by substrate type
3	8/28	hsa04512	HSA-8957275
		ECM-receptor interaction	Post-translational protein phosphorylation
4	8/23	hsa00140	HSA-193775
		Steroid hormone biosynthesis	Synthesis of bile acids and bile salts via 24-hydroxycholeste

Subsequently, we applied Cytohubba to sort out the 10 genes assessed with highest connectivity degree. The results revealed that the most significant gene was *CDK1* with connectivity degree = 69, followed by *CDC20* (degree = 68), *MAD2L 1*(degree = 64), *CCNB2* (degree = 63), *CCNA2* (degree = 63), *UBE2C* (degree = 62), *BUB1B* (degree = 39), *CCNB1* (degree = 60), *AURKA* (degree = 59) and *NDC80* (degree = 58) (Fig. 4b, **Table S3**). All these hub genes were relatively overexpressed in chronic HBV-related HCC. And we applied the TCGA-LIHC cohort, which we selected the 6 HBV-related tumor samples and 6 matched normal samples, to validate the hub gene expression levels. It was shown that all of the hub genes were proved to be upregulated in the tumor samples, only with *MAD2L 1* having a slightly large adjusted P value (Fig. 5, Table 3).

Table 3  
Validation of 10 hub genes in TCGA-  
LIHC 6 pairs of HBC-HCC tissues

Gene	logFC	Adj. PValue
<i>CDK1</i>	2.83	0.03
<i>CDC20</i>	3.67	0.03
<i>MAD2L1</i>	1.60	0.08
<i>CCNB2</i>	2.80	0.02
<i>CCNA2</i>	2.65	0.04
<i>UBE2C</i>	3.90	0.01
<i>BUB1B</i>	2.03	0.04
<i>CCNB1</i>	3.39	0.02
<i>AURKA</i>	2.49	0.02
<i>NDC80</i>	2.55	0.02

### 3.4 Survival analysis of 10 hub genes individually and conjointly

We applied the K–M plotter to examine the overall survival of the 10 potential hub genes. The tool set the best cutoff value automatically according to expression level of each input gene, two groups of patients were established to calculate their overall survival curves respectively. A total of 150 Hepatitis virus positive HCC patients were analyzed for overall survival. Data showed that the higher expression of 7 genes (*CDK1*, *CDC20*, *MAD2L1*, *CCNA2*, *CCNB1*, *AURKA*, *NDC80*) was related to disadvantageous survival impact of HCC samples (Log-rank  $P < 0.05$ ) (Fig. 6a-g). And the 7 genes underwent a gene-weighting conjoint overall survival analysis to look for their comprehensive effect on the overall survival in chronic HBV-related HCC patients, which also presented an adverse prognostic value (Upper quartile survival: 54.1 months in low expression cohort vs 13.8 months high expression cohort, log-rank  $P$  value = 0.01) (Fig. 6h).

Then, we used the DGIdb database to predict drugs targeting the 7 hub genes which was believed to have adverse prognostic effect. Based on the criteria described in Methods, only one drug, FOSTAMATINIB, was identified to inhibit both *CDK1* and *AURKA*.

### 3.5 A prognostic signature composed of gene pairs for HBV-related HCC patients

Moreover, prognostic signatures for HBC-HCC patients were also in urgent needs. Again, gene pairs were selected to construct the signature because of the repeatability. A total of 40063 gene pairs were



screened to be heterogeneous based on 442 of 457 DEGs in HBV-related TCGA-LIHC cohort (n = 95). With univariate Cox regression, 83 pairs exhibited significant prognostic value ( $P < 0.001$ ) (**Table S4**). After Lasso-Cox and stepwise regression, 4 gene pairs were left to compose the model (Fig. 7a).

To evaluate the prognostic effect of the signature, 95 patients were divided into two groups according to the risk calculated by the signature. K-M analysis revealed a significant prognostic efficiency of OS between two groups (Log-rank p value  $< 0.0001$ , ROC-AUC = 93.1% 95%CI = 87.8%-98.5% ) (Fig. 7b, 7c). Subsequently, a nomogram was depicted to predict the 1-year and 5-year overall survival based on the signature, with a assured performance shown in the 5-year calibration curve (Fig. 7d, 7e).

### 3.6 Elucidation of the risk predicted by the prognostic signature

To undermine the mechanisms that determine the higher risk of HBV-HCC patients, we firstly applied xCell to unveil the cellular landscape between high and low risk groups. In the heatmap in which all of the cells had been hierarchically clustered, we observed the first cluster in which the cells were downregulated in the high-risk group, including hematopoietic stem cells (HSCs), pericytes, megakaryocytes as well as a reduction of stromal score (Wilcoxon test, P value  $< 0.05$ ) (Fig. 8a).

Subsequently, hallmarks of cancer which had been widely recognized were downloaded from MSigDb, including angiogenesis, apoptosis, cell cycle, DNA repair, EMT, hypoxia, inflammatory & immunity and metabolism pathways. ssGSEA demonstrated multiple hallmarks responded to the risk. DNA Repair, G2M Checkpoint, Myc Targets, WNT/ $\beta$  - Catenin Signaling and E2F Targets were found to be upregulated in the high-risk groups, and Heme Metabolism, Bile Acid Metabolism, Fatty Acid Metabolism and Xenobiotic Metabolism were found to be downregulated. (Fig. 8b).

Furthermore, to determine whether the prognostic model could be an independent prognostic factor, we first included different demographic characteristics, such as age, pathologic stage, T of TMN staging system, which have been considered to have impact on the HCC patients (M and N staging were not included because a large proportion of missing value). Univariate Cox regression implicated that tumor stage, T staging and our signature had a prognostic impact on HBV-HCC patients, however, when it comes to multivariate Cox regression, only the model still had a significant prognostic impact rather than tumor stage and T staging ( $P < 0.001$ ) (Fig. 8c), hinting that the gene pairs-based signature was an independent prognostic factor for HBV-related HCC patients.

## 4. Discussion

In the present study, we focused on the HBV-related HCC because it is not only one of the common subtypes of liver cancer, but also a tricky public health problem. With the development of high-throughput sequencing and microarray technology, identifying DEGs and potent pathways in hepatocellular carcinoma is becoming a routine task, which provides an effective way to explore potential targets in diagnosing and prognosing HCC in the early stages[18, 19]. However, only a small part of studies focused

on the establishment of signatures for HBC-related HCC[20] which makes our study meaningful. Here, we excavated DEGs between HCC and adjacent normal tissues based on two sets of gene expression profiling data. 457 genes were identified as DEGs which underwent GO analysis, KEGG analysis. PPI network were constructed and 10 hub genes were selected to undergo survival analysis. Innovatively, A diagnostic signature composed of 2 gene pairs and a prognostic signature composed of 4 pairs were built with high efficiency. Moreover, high risk of HBV-related HCC was primarily elucidated based on cellular composition analysis and enrichment.

Firstly, combined KEGG and GO analysis, it was clearly implicated that the identified DEGs were significantly involved in both cell cycle and metabolism process. Interestingly, up-regulated genes were mainly enriched in pathways ways associated with cell cycle and down-regulated genes mainly involved in metabolism pathway. For the cell cycle, there have been some studies unveiled the relationship of Hepatitis B virus, mitosis and hepatocarcinogenesis. HBV is usually involved in the development of liver cancer through HBx, immune imbalance and viral DNA integration into the host genome[21, 22]. Higher HBV load may contribute chronic liver inflammation by mediating immune response of host cytotoxic T lymphocyte (CTL) to promote the destruction and regeneration of HBV infected hepatocytes, and then increase the opportunity of mitotic replication errors[23, 24]. The accumulation of genetic and epigenetic alterations on cell growth advantage ultimately promotes the hepatocarcinogenesis[25]. HBx protein is an oncoprotein encoded by HBV, which may act as a transcription activator of oncogene or host gene to participate in cell growth regulation, DNA repair and epigenetic modification[26]. It may also affect cell cycle, cell transformation and other signal transduction pathways, and then contribute to HCC development[27, 28]. HBx can inhibit the expression of Wnt/catenin, or activate the expression of HURP through p38 / MAPK pathway and SATB1 protein by interacting with p53, TGF  $\beta$ , Fas and TNF, leading to the accumulation of anti-apoptotic proteins, and eventually inhibit the apoptosis of hepatoma cells[22, 29–32]. On the other hand, genomic integration of HBV leads to host genome dysfunction, affecting cell proliferation, cell cycle progression, apoptosis, and even chromosome stability[33]. Studies have found that recurrent integration of TERT and MLL4 is often observed in HCC[34]; HBx-LINE integration, as a function of long non-coding RNA, has been proved to promote the carcinogenesis[35].

Secondly, for the phenomenon that downregulated genes mainly participated in the metabolism pathways, it has been found that HBV infection is widely involved in the metabolic process of liver cells, thus participating in the hepatocarcinogenesis[22, 36]. Studies found that HBV infection leads to up-regulation of glucose metabolism (such as gluconeogenesis, glucose aerobic oxidation, and pentose phosphate pathway) and lipid metabolism (such as fatty acids, phospholipids and cholesterol biosynthesis)[37, 38] in liver. Recent studies on the mechanism of metabolic recombination provide a new perspective for HBV induced hepatocarcinogenesis. Studies have shown that HBx overexpression can participate in the transcriptional activation of steroid regulatory element binding protein-1 and peroxisome proliferator activated receptor (PPAR  $\gamma$ ), leading to hepatocyte steatosis[39]. On the other hand, HBx also participate in the process of glucose metabolism recombination by regulating mitochondrial autophagy through BNIP3L, which can increase the stemness of hepatoma cells[40]. Some of the above studies are consistent with our study, but why most metabolic-related genes are down-

regulated deserves further study. Furthermore, for the oxidation-reduction process, it is recognized that the formation of ROS is a common feature of many cellular biological functions and an important target for cytotoxic effects in the progression of HCC[41]. Important ROS-related liver cancer growth signal transduction pathways include nuclear translocation of nuclear factor-Kb (NF-kB)[42]. PI3K/AKT/mTOR signaling cascade is also involved in oxidative stress of HCC. Its activation is regarded as a stimulator for cell growth and proliferation[43]. Unlike NF-kB, the increased level of ROS suppresses the phosphorylation of AKT and mTOR, thereby enhancing the apoptosis of HCC cells[44]. Recent studies have reported various genes and drugs targeting the PI3K/AKT/mTOR cascade. A combination of sorafenib and C2-ceramide therapy was thought to stimulate the production of ROS, activating caspases-dependent cell apoptosis via PI3K/AKT/mTOR pathway[45]. CYP3A5, a member of CYP450, can induce ROS accumulation, inhibiting AKT phosphorylation at Ser473[44]. CYP450 is considered as another major source of ROS production except from NADPH, especially in liver[41]. However, there are few studies associating CYP450 with the oxidation-reduction process of HCC patients. In this study, a total of 57 genes enriched in oxidation-reduction process, 16 of them belong to CYP450 family, indicating a potential target for the HCC therapy.

The diagnostic model, which was established after cross-validation Lasso regression and best subset selection regression, presented to have a good diagnostic performance both in the training and validation cohort. We thought highly of the diagnostic model constructed in the present study firstly because of the application of gene pairs, which was a good way to eliminate overall sequencing differences between datasets[46], and we used a quite serious variates selection pipeline to determine the gene pairs included in the model. Here, *RGS5*|*FAHD2A* and *CXCL14*|*SAMD5* are two gene pairs that composed the signature. It could be conceived that clinically, the expression levels of these two pairs of genes can be compared to identify whether patients with hepatitis B have already acquired HCC. This can solve the problem that different high-throughput methods may not be suitable for model coefficients but the pity is that the number of samples in our training set is not very large. Secondly, because all possible variate combinations have been traversed in the Best subset selection, the selected features can be optimal, and additionally, cross-validation Lasso avoided the overfitting. This model has shown very high efficacy in repeated validations, and we believe it can be a good method for assisting pathological diagnosis of HBV induced liver cancer.

As to the hub genes, including all hub genes were up-regulated in cancer tissues, and most of them are related to cell cycle and mitosis. *CDK1* was the one with highest degree. According to previous studies, *CDK1*, also called cell division cycle protein 2 (*CDC2*), is an important regulator of the cell cycle, which is required for the transition from the G2 phase into mitosis[47, 48]. For HCC, several studies have suggested that *CDK1* related to HCC cell proliferation[49] and regulated apoptin-related cell apoptosis[50]. Interestingly, FOSTAMATINIB was predicted to have therapeutic effect for HBV-related HCC patients. Fostamatinib, a FDA-proved tyrosine kinase (Syk) inhibitor, has been proved to be able to treat chronic immune thrombocytopenia and rheumatoid arthritis[51, 52]. Recent studies even hinted that it has effect on COVID-19[53]. The effect of fostamatinib on HCC was hardly reported and lacked large-scale clinical

trials[54]. Here, we further revealed the potential mechanism of fostamatinib treatment of HBV-reduced HCC, that is, targeting *CDK1* and *AURKA* and breaking the disease-causing PPI network.

Furthermore, a prognostic model composed 4 gene pairs was built for HBV-HCC patients, which presented an excellent performance (AUC = 93.1%). The advantages of pairs discussed above also made sense in the prognostic model. However, we hope to have a larger cohort to establish and validate models in the future, for there are still few high-quality datasets for hepatitis B-related liver cancer, especially those with clinical follow-up data.

With xCell, we primarily demonstrated the underlying mechanism related to the rising risk of death from HBV-related HCC, which mainly points to the loss of stromal cells in the tissues. It had been reported that the score of stromal and immune cells of tumor tissue have association with its clinical characteristics[55]. Tissue-based approaches were confirmed to be able to characterize the tumor infiltrates in clear cell renal cell carcinoma and primary glioblastoma multiforme, according to previous studies[56, 57]. A recent research showed that high-risk cohorts in HCC have higher immune, and stromal scores than that in low-risk cohorts, which was consistent with our results[58].

Further, result of ssGSEA hinted that the apart from WNT/ $\beta$ -Catenin Signaling, DNA Repair and G2M Checkpoint which had been fully discussed above, Myc and E2F targets also related with the higher risk of HBV-related HCC. And metabolism pathways exhibited down-regulation in the high risk group. For it was widely accepted that HBx and c-Myc were respectively oncogenic factors in the HBV virus and host hepatocyte, for critical interaction was confirmed between the two proteins[59]. HBx-mediated Myc stabilization greatly promoted the carcinogenic effect of the virus, on account that SCF (Skp2) ubiquitin E3 ligase could intervene Myc ubiquitination that led to Myc oncoprotein, upon which HBx has an effect of stabilization[60, 61]. At the same time, by activating PIK3CA/Akt/mTOR and c-Myb/COX-2 pathways, c-Myc drove apoptosis, for which E2F1 was found acting as a weakening factor[62, 63]. In RB1-altered tumors, the E2F pathway was found activated which was already reported in HCC. While E2F pathway could also activate in CCN-HCC without RB1 inactivation event which might be partly contributed by the ability of cyclin E/Cdk2 complexes to phosphorylate Rb[63]. Briefly, we demonstrated that some pathways not only involved in the occurrence of HCC but also involved in the subsequent progression of the tumor.

## 5. Conclusion

This study conducted a relatively comprehensive bioinformatics analysis of hepatitis B related HCC, which revealed that DEGs participate in various pathways but mostly in cell cycle and metabolism process in HBV-HCC. And analysis of these genes pave roads for future molecular mechanism studies and provide targets for the potential therapy of HBV-related HCC in clinic. Innovatively, a diagnostic model and a prognostic model based on gene pairs were established with a good performance, and the risk was related with stromal cells and different hallmarks. Together, this study sheds lights on the prevention and treatment of HBV-related hepatocellular carcinoma.

# Declarations

**Data Availability:** TCGA-LIHC profiles and clinical characteristics are gained from UCSC Xena database (<https://xenabrowser.net/datapages>). GSE121248 and GSE94660, GSE76724 are from Gene Expression Omnibus (GEO) (<http://www.ncbi.nlm.nih.gov/geo/>).

**Author Contributions:** B.S. completed the manuscript writing, Y.W. participated in the reviewing of literatures and discussion. S.M. was responsible for the data analysis. All of the author agreed to publish the manuscript.

**Funding:** This research received no external funding.

**Conflicts of interest:** The authors declare no conflict of interest.

**Ethics approval and consent to participate:** Not applicable.

**Consent for publication:** Not applicable.

# References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2018;68:394–424.
2. Thomas MB, Jaffe D, Choti MM, Belghiti J, Curley S, Fong Y, Gores G, Kerlan R, Merle P, O'Neil B, et al. Hepatocellular carcinoma: consensus recommendations of the National Cancer Institute Clinical Trials Planning Meeting. *J Clin Oncol.* 2010;28:3994–4005.
3. Ozen C, Yildiz G, Dagcan AT, Cevik D, Ors A, Keles U, Topel H, Ozturk M. Genetics and epigenetics of liver cancer. *N Biotechnol.* 2013;30:381–4.
4. Coskun M. Hepatocellular Carcinoma in the Cirrhotic Liver: Evaluation Using Computed Tomography and Magnetic Resonance Imaging. *Exp Clin Transplant.* 2017;15:36–44.
5. Niu J, Lin Y, Liu P, Yu Y, Su C, Wang X. Microarray analysis on the lncRNA expression profile in male hepatocellular carcinoma patients with chronic hepatitis B virus infection. *Oncotarget.* 2016;7:76169–80.
6. Nguyen VT, Law MG, Dore GJ. Hepatitis B-related hepatocellular carcinoma: epidemiological characteristics and disease burden. *J Viral Hepat.* 2009;16:453–63.
7. Cui Y, Jia J. Update on epidemiology of hepatitis B and C in China. *J Gastroenterol Hepatol.* 2013;28(Suppl 1):7–10.
8. Pelus LM, Fukuda S. Peripheral blood stem cell mobilization: the CXCR2 ligand GRObeta rapidly mobilizes hematopoietic stem cells with enhanced engraftment properties. *Exp Hematol.* 2006;34:1010–20.

9. Song X, Wang Z, Jin Y, Wang Y, Duan W. Loss of miR-532-5p in vitro promotes cell proliferation and metastasis by influencing CXCL2 expression in HCC. *Am J Transl Res.* 2015;7:2254–61.
10. Zhu Q, Sun Y, Zhou Q, He Q, Qian H. Identification of key genes and pathways by bioinformatics analysis with TCGA RNA sequencing data in hepatocellular carcinoma. *Mol Clin Oncol.* 2018;9:597–606.
11. Li MX, Jin LT, Wang TJ, Feng YJ, Pan CP, Zhao DM, Shao J. Identification of potential core genes in triple negative breast cancer using bioinformatics analysis. *Onco Targets Ther.* 2018;11:4105–12.
12. Long T, Liu Z, Zhou X, Yu S, Tian H, Bao Y. Identification of differentially expressed genes and enriched pathways in lung cancer using bioinformatics analysis. *Mol Med Rep.* 2019;19:2029–40.
13. Walter W, Sanchez-Cabo F, Ricote M. GOpilot: an R package for visually combining expression data with functional analysis. *Bioinformatics.* 2015;31:2912–4.
14. Aran D, Hu Z, Butte AJ. xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol.* 2017;18:220.
15. Nie H, Bu F, Xu J, Li T, Huang J. 29 immune-related genes pairs signature predict the prognosis of cervical cancer patients. *Sci Rep.* 2020;10:14152.
16. Zhang Z. Variable selection with stepwise and best subset approaches. *Ann Transl Med.* 2016;4:136.
17. Cotto KC, Wagner AH, Feng YY, Kiwala S, Coffman AC, Spies G, Wollam A, Spies NC, Griffith OL, Griffith M. DGIdb 3.0: a redesign and expansion of the drug-gene interaction database. *Nucleic Acids Res.* 2018;46:D1068–73.
18. Ye C, Tao R, Cao Q, Zhu D, Wang Y, Wang J, Lu J, Chen E, Li L. Whole-genome DNA methylation and hydroxymethylation profiling for HBV-related hepatocellular carcinoma. *Int J Oncol.* 2016;49:589–602.
19. Liu A, Wu Q, Peng D, Ares I, Anadon A, Lopez-Torres B, Martinez-Larranaga MR, Wang X, Martinez MA. A novel strategy for the diagnosis, prognosis, treatment, and chemoresistance of hepatocellular carcinoma: DNA methylation. *Med Res Rev.* 2020;40:1973–2018.
20. Pezzuto F, Buonaguro L, Buonaguro FM, Tornesello ML. **The Role of Circulating Free DNA and MicroRNA in Non-Invasive Diagnosis of HBV- and HCV-Related Hepatocellular Carcinoma.** *Int J Mol Sci* 2018, 19.
21. Forner A, Reig M, Bruix J. Hepatocellular carcinoma. *Lancet.* 2018;391:1301–14.
22. Levrero M, Zucman-Rossi J. Mechanisms of HBV-induced hepatocellular carcinoma. *J Hepatol.* 2016;64:84-s101.
23. Schuch A, Salimi Alizei E, Heim K, Wieland D, Kiraithe MM, Kemming J, Llewellyn-Lacey S, Sogukpinar Ö, Ni Y, Urban S, et al. Phenotypic and functional differences of HBV core-specific versus HBV polymerase-specific CD8 + T cells in chronically HBV-infected patients with low viral load. *Gut.* 2019;68:905–15.
24. Lee PC, Chao Y, Chen MH, Lan KH, Lee IC, Hou MC, Huang YH. **Risk of HBV reactivation in patients with immune checkpoint inhibitor-treated unresectable hepatocellular carcinoma.** *J Immunother*

25. Lee SM, Kim-Ha J, Choi WY, Lee J, Kim D, Lee J, Choi E, Kim YJ. Interplay of genetic and epigenetic alterations in hepatocellular carcinoma. *Epigenomics*. 2016;8:993–1005.
26. Slagle BL, Bouchard MJ. Role of HBx in hepatitis B virus persistence and its therapeutic implications. *Curr Opin Virol*. 2018;30:32–8.
27. Chaturvedi VK, Singh A, Dubey SK, Hetta HF, John J, Singh MP. Molecular mechanistic insight of hepatitis B virus mediated hepatocellular carcinoma. *Microb Pathog*. 2019;128:184–94.
28. Liu S, Koh SS, Lee CG. **Hepatitis B Virus X Protein and Hepatocarcinogenesis**. *Int J Mol Sci* 2016, 17.
29. Cheng ST, Ren JH, Cai XF, Jiang H, Chen J. HBx-elevated SIRT2 promotes HBV replication and hepatocarcinogenesis. *Biochem Biophys Res Commun*. 2018;496:904–10.
30. Chen SL, Zhang CZ, Liu LL, Lu SX, Pan YH, Wang CH, He YF, Lin CS, Yang X, Xie D, Yun JP. A GYS2/p53 Negative Feedback Loop Restricts Tumor Growth in HBV-Related Hepatocellular Carcinoma. *Cancer Res*. 2019;79:534–45.
31. Cao P, Yang A, Wang R, Xia X, Zhai Y, Li Y, Yang F, Cui Y, Xie W, Liu Y, et al. Germline Duplication of SNORA18L5 Increases Risk for HBV-related Hepatocellular Carcinoma by Altering Localization of Ribosomal Proteins and Decreasing Levels of p53. *Gastroenterology*. 2018;155:542–56.
32. Wungu CDK, Amin M, Ruslan SEN, Purwono PB, Kholili U, Maimunah U, Setiawan PB, Lusida MI, Soetjipto S, Handajani R. Association between host TNF- $\alpha$ , TGF- $\beta$ 1, p53 polymorphisms, HBV X gene mutation, HBV viral load and the progression of HBV-associated chronic liver disease in Indonesian patients. *Biomed Rep*. 2019;11:145–53.
33. Tu T, Budzinska MA, Shackel NA, Urban S. **HBV DNA Integration: Molecular Mechanisms and Clinical Implications**. *Viruses* 2017, 9.
34. An P, Xu J, Yu Y, Winkler CA. Host and Viral Genetic Variation in HBV-Related Hepatocellular Carcinoma. *Front Genet*. 2018;9:261.
35. Lau CC, Sun T, Ching AK, He M, Li JW, Wong AM, Co NN, Chan AW, Li PS, Lung RW, et al. Viral-human chimeric transcript predisposes risk to liver cancer development and progression. *Cancer Cell*. 2014;25:335–49.
36. Torresi J, Tran BM, Christiansen D, Earnest-Silveira L, Schwab RHM, Vincan E. HBV-related hepatocarcinogenesis: the role of signalling pathways and innovative ex vivo research models. *BMC Cancer*. 2019;19:707.
37. Masson JJ, Billings HW, Palmer CS. Metabolic reprogramming during hepatitis B disease progression offers novel diagnostic and therapeutic opportunities. *Antivir Chem Chemother*. 2017;25:53–7.
38. Lamontagne RJ, Casciano JC, Bouchard MJ. A broad investigation of the HBV-mediated changes to primary hepatocyte physiology reveals HBV significantly alters metabolic pathways. *Metabolism*. 2018;83:50–9.

39. Kim KH, Shin HJ, Kim K, Choi HM, Rhee SH, Moon HB, Kim HH, Yang US, Yu DY, Cheong J. Hepatitis B virus X protein induces hepatic steatosis via transcriptional activation of SREBP1 and PPARgamma. *Gastroenterology*. 2007;132:1955–67.
40. Chen YY, Wang WH, Che L, Lan Y, Zhang LY, Zhan DL, Huang ZY, Lin ZN, Lin YC: **BNIP3L-Dependent Mitophagy Promotes HBx-Induced Cancer Stemness of Hepatocellular Carcinoma Cells via Glycolysis Metabolism Reprogramming**. *Cancers (Basel)* 2020, **12**.
41. Tien Kuo M, Savaraj N. Roles of reactive oxygen species in hepatocarcinogenesis and drug resistance gene expression in liver cancers. *Mol Carcinog*. 2006;45:701–9.
42. Wang F, Yang JL, Yu KK, Xu M, Xu YZ, Chen L, Lu YM, Fang HS, Wang XY, Hu ZQ, et al. Activation of the NF-kappaB pathway as a mechanism of alcohol enhanced progression and metastasis of human hepatocellular carcinoma. *Mol Cancer*. 2015;14:10.
43. Calvisi DF, Wang C, Ho C, Ladu S, Lee SA, Mattu S, Destefanis G, Delogu S, Zimmermann A, Ericsson J, et al. Increased lipogenesis, induced by AKT-mTORC1-RPS6 signaling, promotes development of human hepatocellular carcinoma. *Gastroenterology*. 2011;140:1071–83.
44. Jiang F, Chen L, Yang YC, Wang XM, Wang RY, Li L, Wen W, Chang YX, Chen CY, Tang J, et al. CYP3A5 Functions as a Tumor Suppressor in Hepatocellular Carcinoma by Regulating mTORC2/Akt Signaling. *Cancer Res*. 2015;75:1470–81.
45. Jiang S, Wang Q, Feng M, Li J, Guan Z, An D, Dong M, Peng Y, Kuerban K, Ye L. C2-ceramide enhances sorafenib-induced caspase-dependent apoptosis via PI3K/AKT/mTOR and Erk signaling pathways in HCC cells. *Appl Microbiol Biotechnol*. 2017;101:1535–46.
46. Zhong J, Huang Q, Wang Y, Gao H, Jia H, Fan J, Liang H. Distinguishing Kawasaki Disease from Febrile Infectious Disease Using Gene Pair Signatures. *Biomed Res Int*. 2020;2020:6539398.
47. Malumbres M, Barbacid M. Cell cycle, CDKs and cancer: a changing paradigm. *Nat Rev Cancer*. 2009;9:153–66.
48. Malumbres M. Cyclin-dependent kinases. *Genome Biol*. 2014;15:122.
49. Zhou J, Han S, Qian W, Gu Y, Li X, Yang K. Metformin induces miR-378 to downregulate the CDK1, leading to suppression of cell proliferation in hepatocellular carcinoma. *Onco Targets Ther*. 2018;11:4451–9.
50. Zhao J, Han SX, Ma JL, Ying X, Liu P, Li J, Wang L, Zhang Y, Ma J, Zhang L, Zhu Q. The role of CDK1 in apoptin-induced apoptosis in hepatocellular carcinoma cells. *Oncol Rep*. 2013;30:253–9.
51. Bussel J, Arnold DM, Grossbard E, Mayer J, Trelinski J, Homenda W, Hellmann A, Windyga J, Sivcheva L, Khalafallah AA, et al. Fostamatinib for the treatment of adult persistent and chronic immune thrombocytopenia: Results of two phase 3, randomized, placebo-controlled trials. *Am J Hematol*. 2018;93:921–30.
52. Tanaka Y, Millson D, Iwata S, Nakayamada S. **Safety and efficacy of fostamatinib in rheumatoid arthritis patients with an inadequate response to methotrexate in phase II OSKIRA-ASIA-1 and OSKIRA-ASIA-1X study**. *Rheumatology (Oxford)* 2020.



53. Strich JR, Ramos-Benitez MJ, Randazzo D, Stein SR, Babyak A, Davey RT, Suffredini AF, Childs RW, Chertow DS. **Fostamatinib Inhibits Neutrophils Extracellular Traps Induced by COVID-19 Patient Plasma: A Potential Therapeutic.** *J Infect Dis* 2020.
54. Regan-Fendt K, Li D, Reyes R, Yu L, Wani NA, Hu P, Jacob ST, Ghoshal K, Payne PRO, Motiwala T: **Transcriptomics-Based Drug Repurposing Approach Identifies Novel Drugs against Sorafenib-Resistant Hepatocellular Carcinoma.** *Cancers (Basel)* 2020, **12**.
55. Yoshihara K, Shahmoradgoli M, Martinez E, Vegesna R, Kim H, Torres-Garcia W, Trevino V, Shen H, Laird PW, Levine DA, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun.* 2013;4:2612.
56. Senbabaoglu Y, Gejman RS, Winer AG, Liu M, Van Allen EM, de Velasco G, Miao D, Ostrovnaya I, Drill E, Luna A, et al. Tumor immune microenvironment characterization in clear cell renal cell carcinoma identifies prognostic and immunotherapeutically relevant messenger RNA signatures. *Genome Biol.* 2016;17:231.
57. Zhao Z, Zhang KN, Chai RC, Wang KY, Huang RY, Li GZ, Wang YZ, Chen J, Jiang T. ADAMTSL4, a Secreted Glycoprotein, Is a Novel Immune-Related Biomarker for Primary Glioblastoma Multiforme. *Dis Markers.* 2019;2019:1802620.
58. Liu J, Li W, Zhao H. CFHR3 is a potential novel biomarker for hepatocellular carcinoma. *J Cell Biochem.* 2020;121:2970–80.
59. Tsuchiya H, Amisaki M, Takenaga A, Honjo S, Fujiwara Y, Shiota G. **HBx and c-MYC Cooperate to Induce URI1 Expression in HBV-Related Hepatocellular Carcinoma.** *Int J Mol Sci* 2019, **20**.
60. Kalra N, Kumar V. The X protein of hepatitis B virus binds to the F box protein Skp2 and inhibits the ubiquitination and proteasomal degradation of c-Myc. *FEBS Lett.* 2006;580:431–6.
61. Lee S, Kim W, Ko C, Ryu WS. Hepatitis B virus X protein enhances Myc stability by inhibiting SCF(Skp2) ubiquitin E3 ligase-mediated Myc ubiquitination and contributes to oncogenesis. *Oncogene.* 2016;35:1857–67.
62. Ladu S, Calvisi DF, Conner EA, Farina M, Factor VM, Thorgeirsson SS. E2F1 inhibits c-Myc-driven apoptosis via PIK3CA/Akt/mTOR and COX-2 in a mouse model of human liver cancer. *Gastroenterology.* 2008;135:1322–32.
63. Conner EA, Lemmer ER, Omori M, Wirth PJ, Factor VM, Thorgeirsson SS. Dual functions of E2F-1 in a transgenic mouse model of liver carcinogenesis. *Oncogene.* 2000;19:5054–62.

## Figures

Figure 1

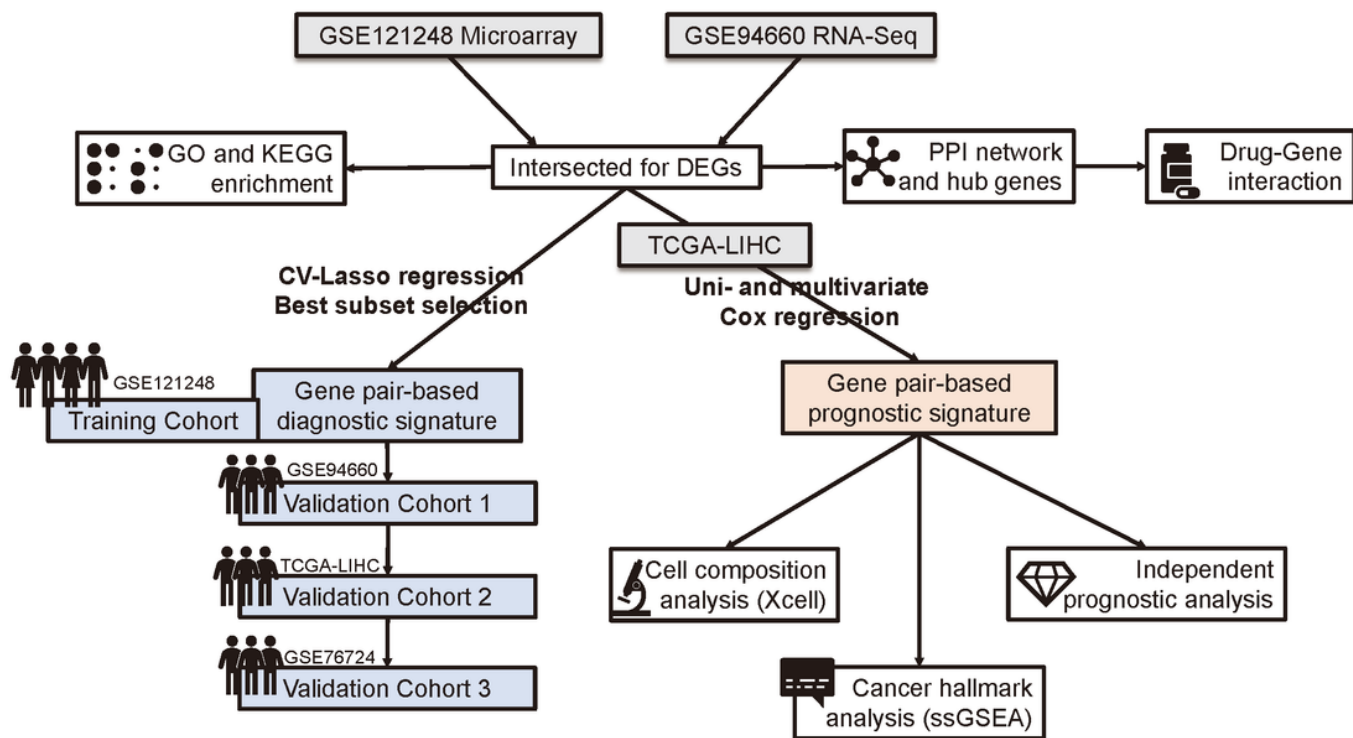


Figure 1

Workflow of our study. DEGs: differentially expressed genes, CV: cross validation

Figure 2

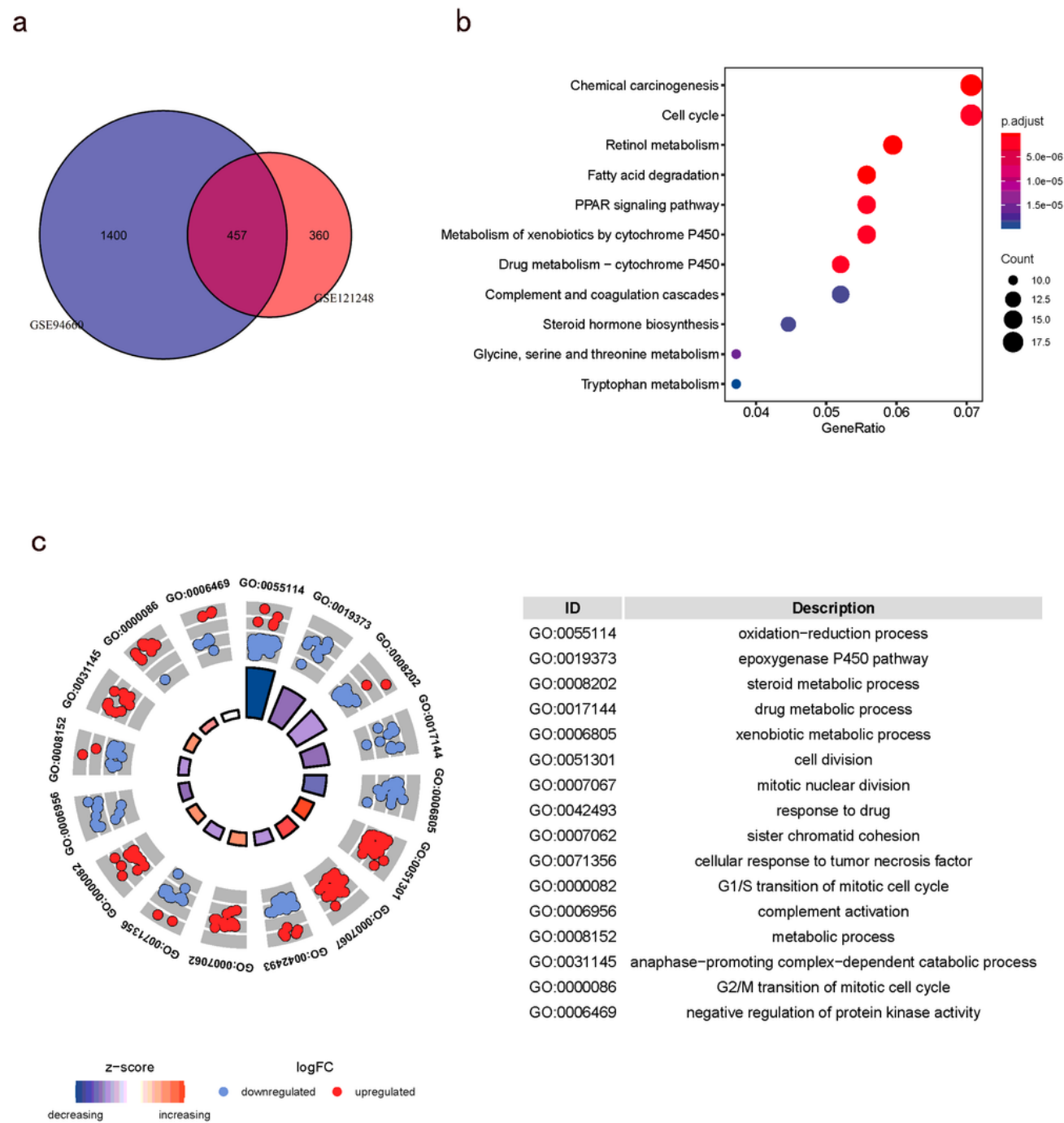


Figure 2

Identification of DEGs and function annotations. (a) DEGs were intersected between GSE121248 and GSE94660. (b) KEGG pathway enrichment for 457 DEGs. (c) Biological process of GO enrichment for 457 DEGs.

Figure 3

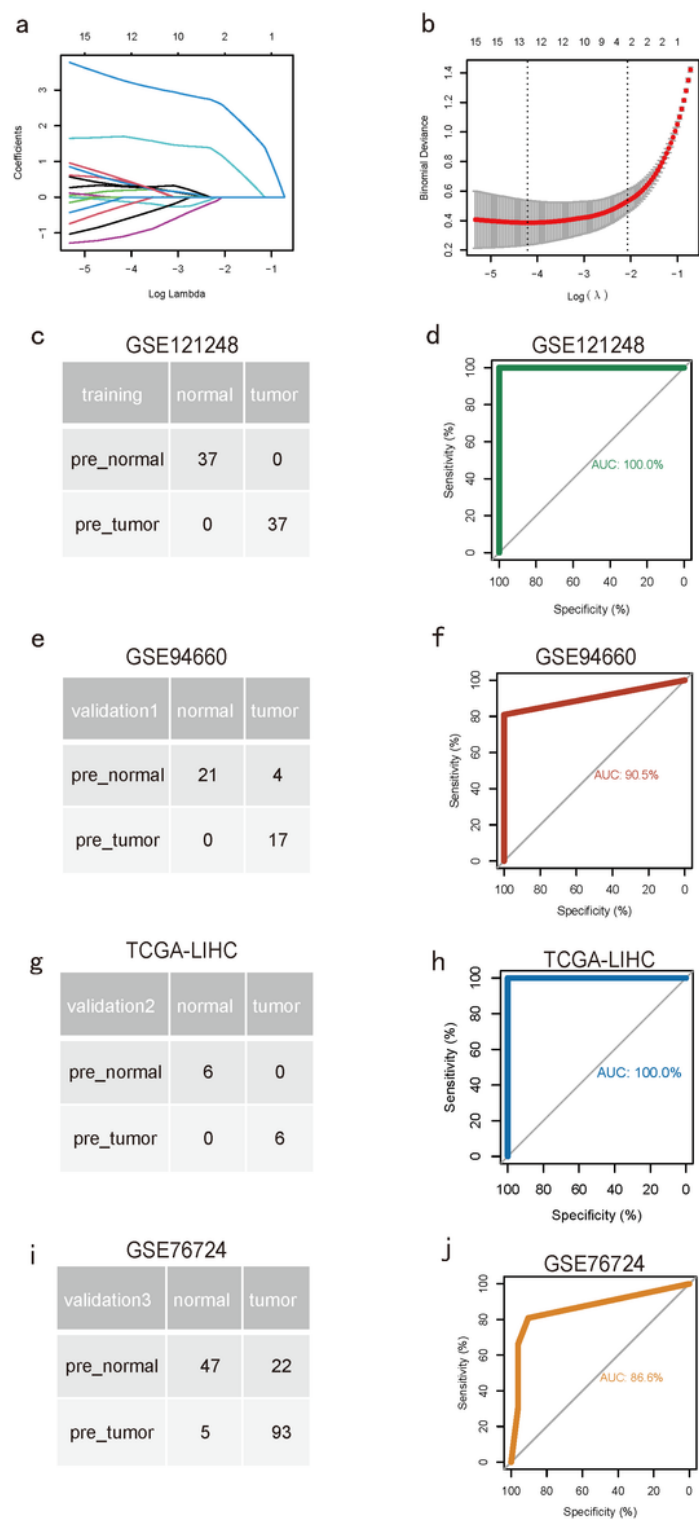
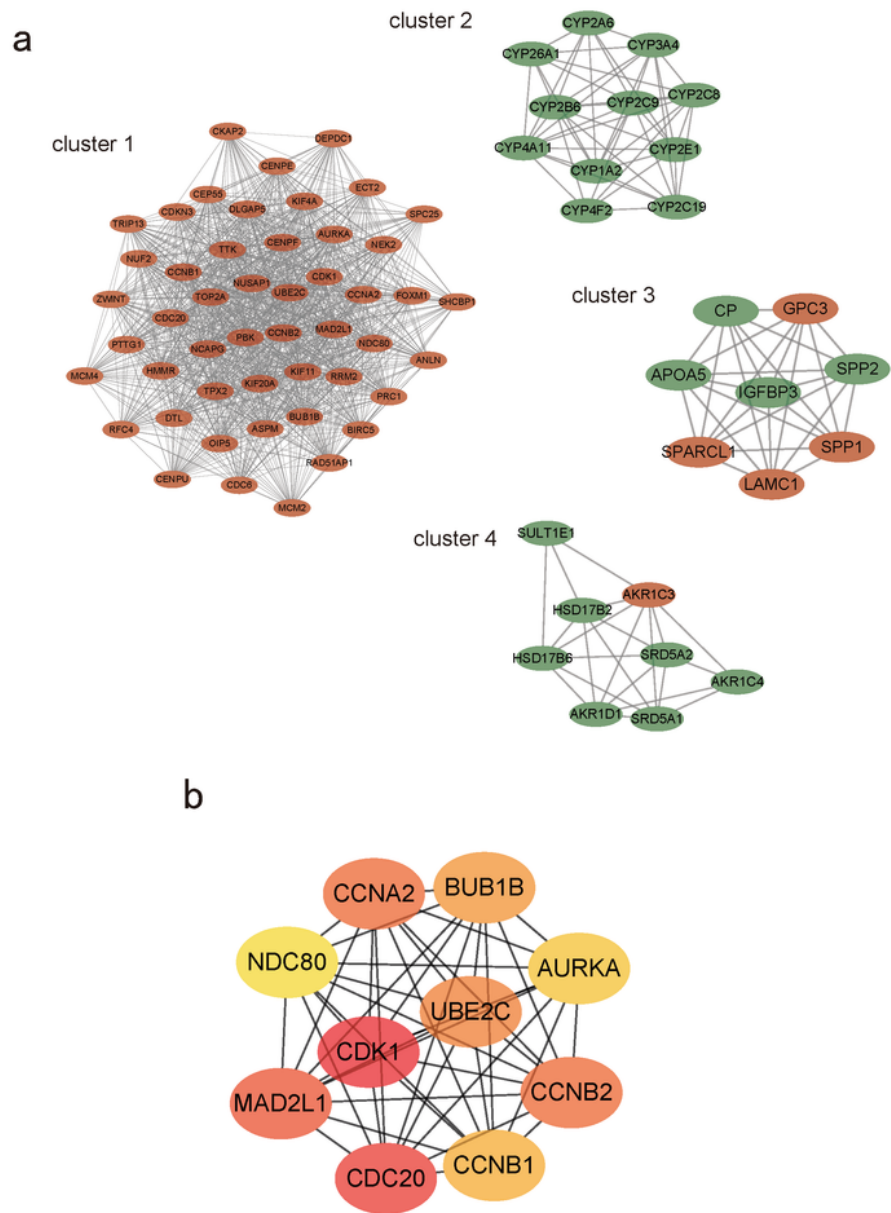


Figure 3

A diagnostic signature composed of gene pairs for HBV-related HCC patients. (a) The change trajectory of each independent variable coefficient during process of variate selection in Lasso regression. Ordinate is the coefficient and the upper abscissa presents the number of non-zero coefficients in the model. (b) The result of cross-validation. Red dot represents the target parameter corresponding to each lambda. (c) Confusion matrix in the training cohort. (d) ROC curve and AUC in the training cohort. (e) Confusion

matrix in the validation cohort 1. (f) ROC curve and AUC in the validation cohort 1 (g) Confusion matrix in the validation cohort 2. (h) ROC curve and AUC in the validation cohort 2. (i) Confusion matrix in the validation cohort 3. (j) ROC curve and AUC in the validation cohort 3.

Figure 4



**Figure 4**

Construction of PPI network and identification of hub genes. (a) Crucial clusters identified by MCODE. (b) PPI network for 10 hub genes. Color shade represents the degree.

Figure 5

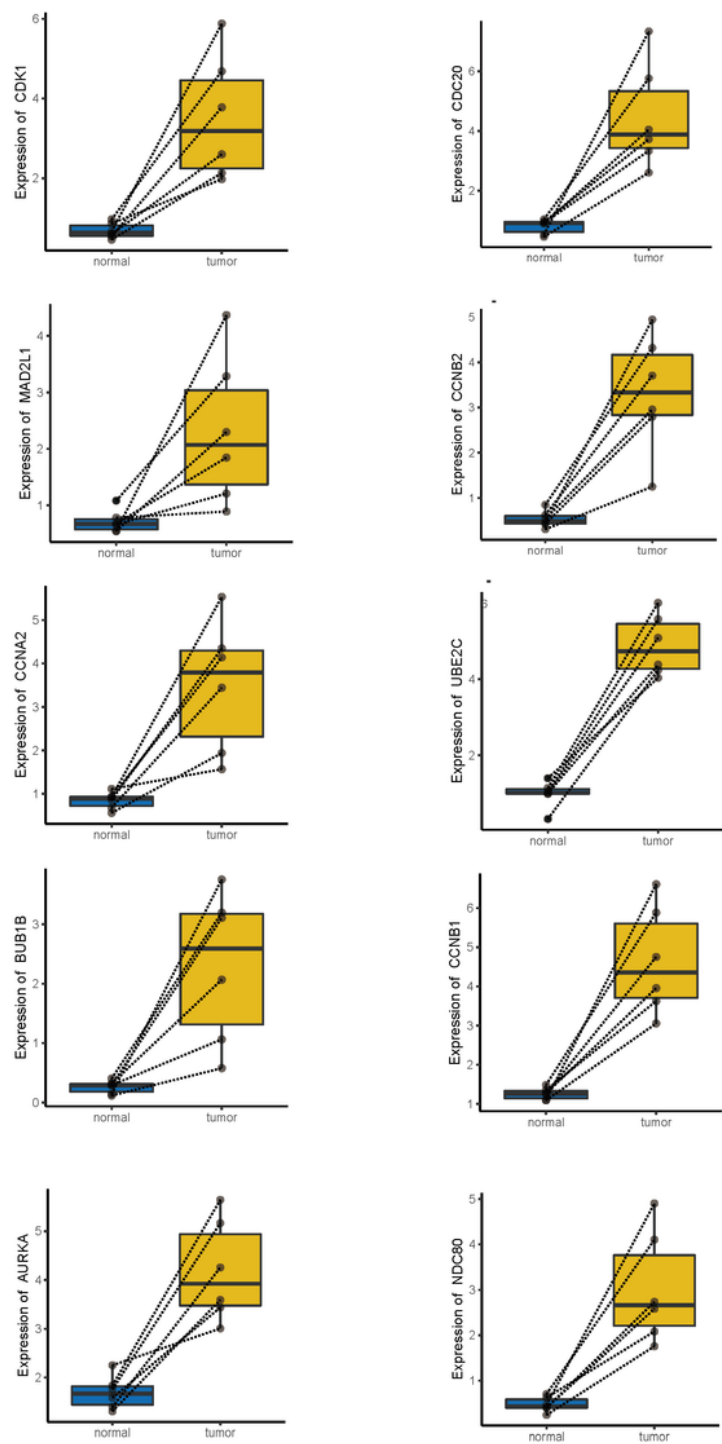


Figure 5

Validation of 10 differentially expressed hub genes in TCGA-LIHC cohort.

Figure 6

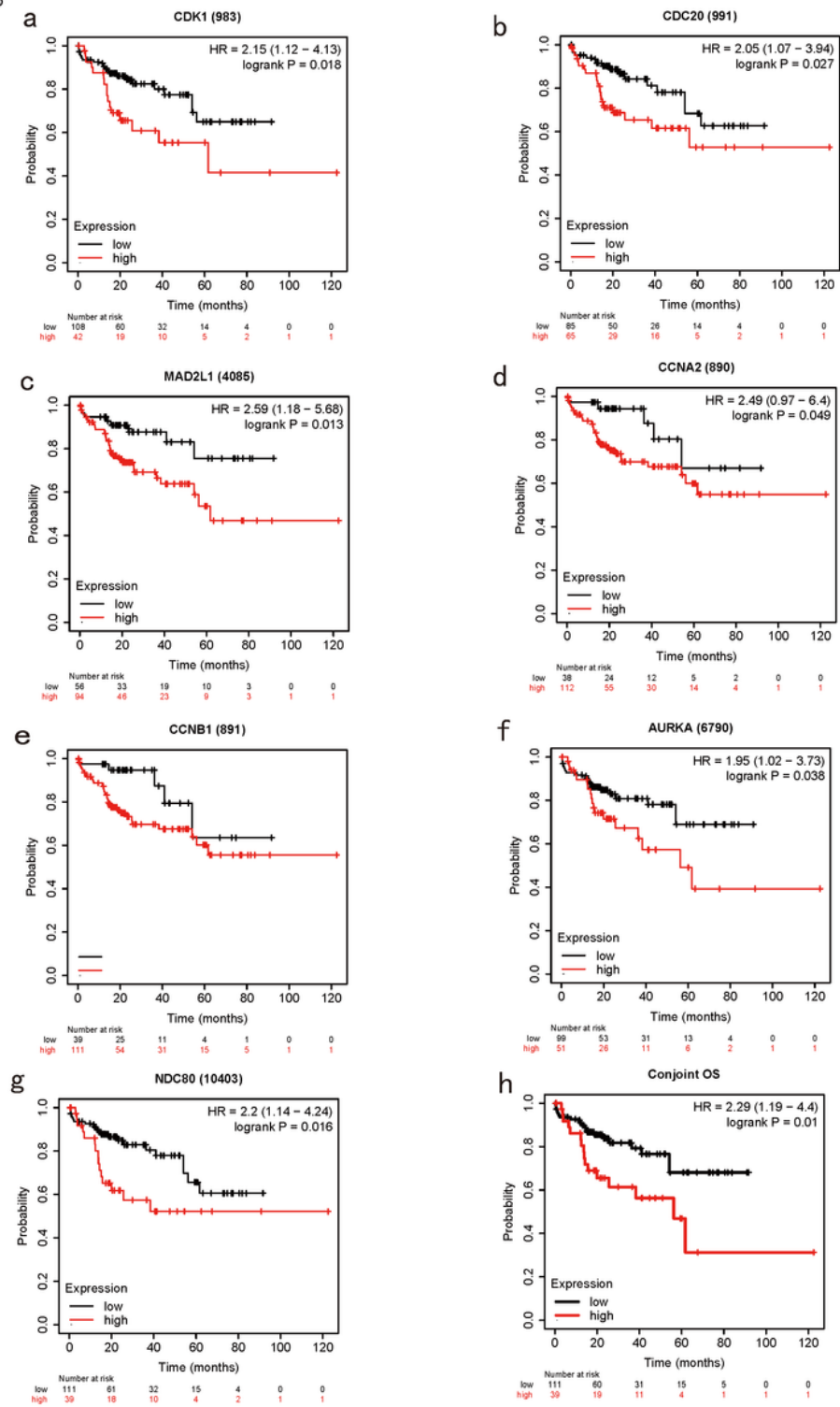


Figure 6

Survival analysis of 10 hub genes individually and conjointly. (a-g) Kaplan-Meier analysis screened out 7 hub genes with survival influence. Red curve: high expression group, black curve: low expression group. (h) The combined prognostic value of 7 hub genes. Mean expression of selected genes were used with gene weights all set to 1.

Figure 7

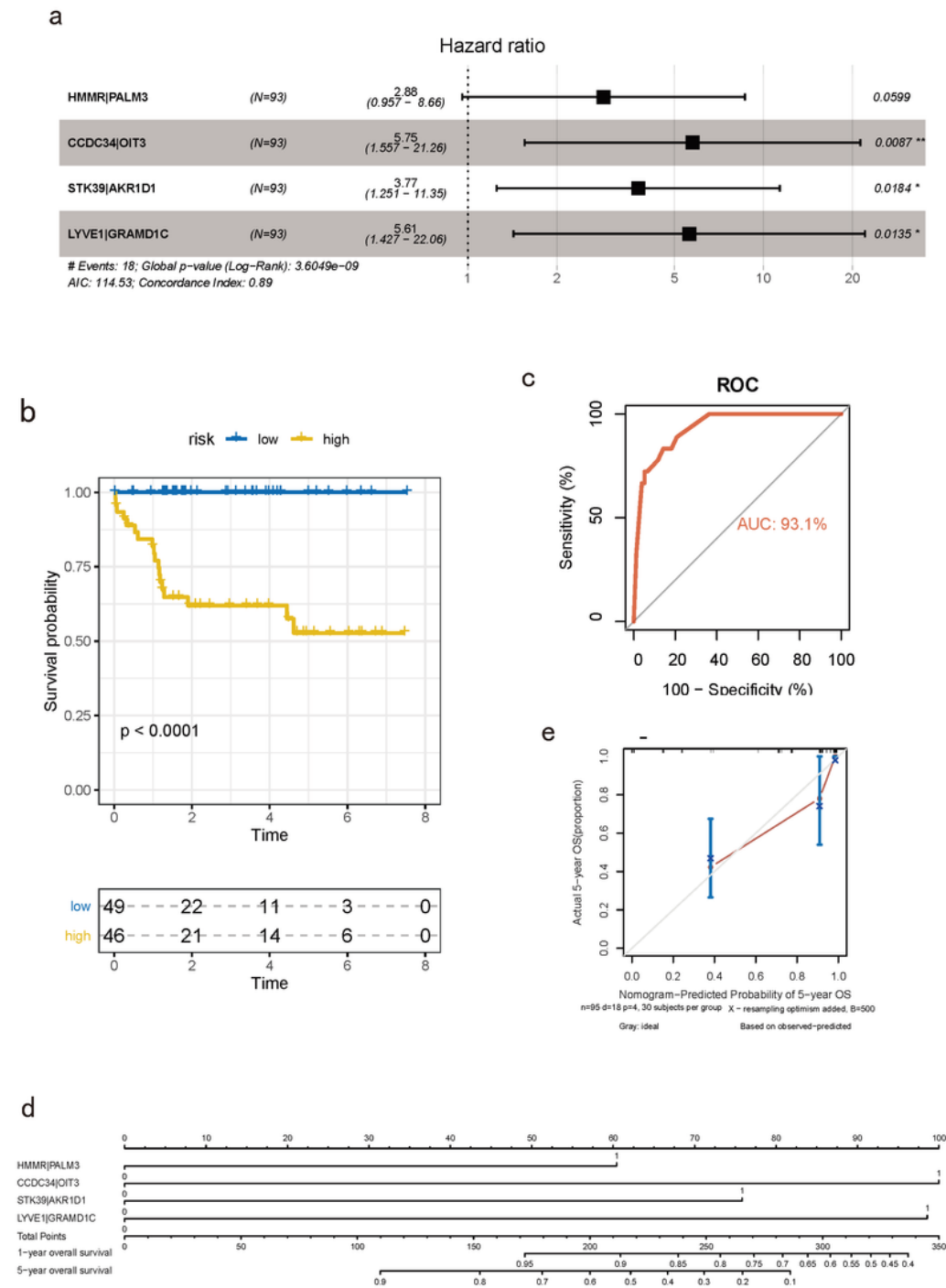
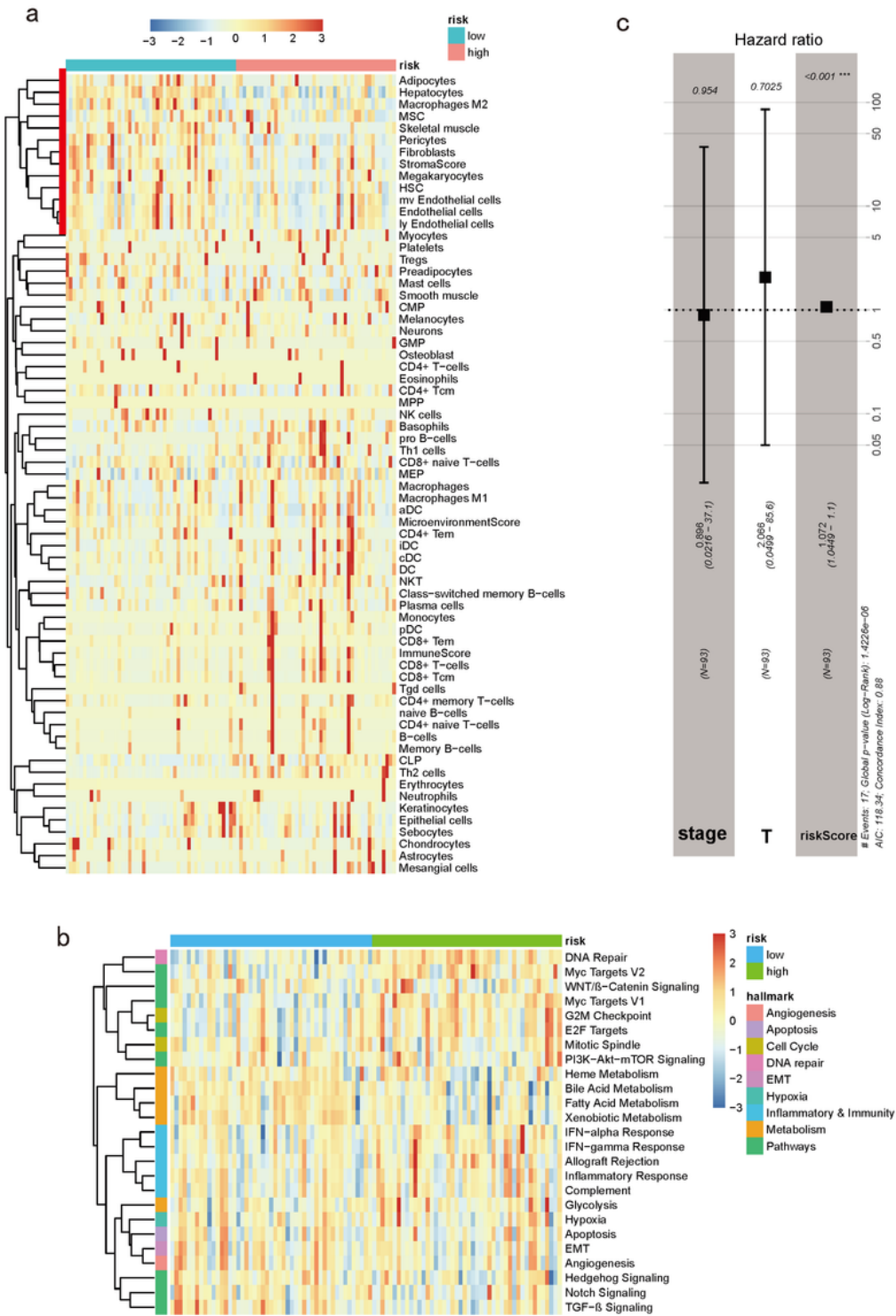


Figure 7

A prognostic signature composed of gene pairs for HBV-related HCC patients. (a) Forest plot of the gene pairs in the multivariate Cox model. (b) Kaplan-Meier analysis for the high and low risk groups predicted by the prognostic signature. (c) ROC curve and AUC for the prognostic model. (d) Nomogram for the prognostic model. (e) 5-year overall survival calibration curve for the model.



Figure 8



**Figure 8**

Elucidation of the risk predicted by the prognostic signature. (a) Cell composition analysis by xCell tool (b) Cancer hallmark analysis by ssGSEA (c) Multivariate Cox regression for tumor stage, T staging and our signature on 93 patients.

Supplementary Files

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

- [ShaocongSupplementarymaterialsWSJO.pdf](#)