

Screening of Early Diagnostic Biomarkers and Prognostic Biomarkers for Liver Cancer Based on GEO and TCGA Databases and Studies on Pathways and Biological Functions Affecting the Survival Time of Liver Cancer

Shiyong Gao (✉ sygao2002@163.com)

Harbin University of Commerce - Harbin <https://orcid.org/0000-0002-8499-9485>

Jian Gang

Harbin University of Commerce - South Campus

Huixin Tan

Fourth Affiliated Hospital of Harbin Medical University

Research

Keywords: bioinformatics, biomarker, liver cancer, diagnostic biomarker, prognostic biomarker

Posted Date: October 8th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-86072/v1>

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

[Read Full License](#)

Abstract

Background

Liver cancer is the sixth most common diagnosed cancer, and the fourth most common cause of cancer death in the world. Currently, the most commonly used diagnostic marker in clinical practice is alpha-fetoprotein(AFP), but its diagnostic accuracy is low. At the same time, the prognosis of liver cancer patients is of great significance to the determination of the diagnosis and treatment of liver cancer and the improvement of the quality of life of the patients. Therefore, the purpose of this paper is to find new diagnostic markers and prognostic markers of liver cancer and to explore the pathways and biological functions related to the prognosis of liver cancer.

Methods

We firstly obtained the GSE25097 dataset and the the Cancer Genome Atlas(TCGA) datasets respectively and then analysed the differentially expressed genes(DEGs). To study the potential biological function of DEGs, we conducted enrichment analysis of GO biological functions and the Reactome pathway with R language. We used protein-protein interaction network analysis to identify the relationship among these common DEGs(Common DEGs), and further screened out the hub genes among these Common DEGs. We used ROC curve analysis to screen the hub genes to determine the genes that could be used as diagnostic markers of liver cancer. Kaplan-Meier analysis and Cox proportional hazards model were used to screen genes associated with prognosis of liver cancer, and further single-gene GSEA(gene set enrichment analysis) was performed on the prognosis genes to explore the mechanism affecting the survival and prognosis of liver cancer patients.

Results

We obtained 790 DEGs and 2162 DEGs respectively from the GSE25097 and TCGA LIHC data sets, and get 102 Common DEGs by overlapping the two DEGs. We screened 22 hub genes from 102 Common DEGs. We used ROC curve and survival curve to analyze these 22 Hub genes, and found that there were 16 genes with the AUC > 90%, among which the expression levels of ESR1,SPP1 and FOSB genes were closely related to the survival time of liver cancer patients. We revealed all the related pathways of ESR1, FOSB and SSP1 genes by using single-gene GSEA analysis, and found three common pathways of ESR1, FOBS and SPP1 genes, seven common pathways of ESR1 and SPP1 genes, and four common pathways of ESR1 and FOSB genes.

Conclusion

we found that ten genes with high expression in the liver cancer, including SPP1, AURKA, NUSAP1, TOP2A, UBE2C, AFP, GMNN, PTTG1, RRM2 and SPARCL1, and six genes with low expression, including CXCL12, FOS, DCN, SOCS3, FOSB and PCK1, can be used as markers of liver cancer diagnosis, among which FOBS and SPP1 genes can also be used as prognostic markers of liver cancer. Activation of the cell cycle-related pathway, PANCREAS BETA CELLS pathway, and the estrogen signalling pathway in

LIVER CANCER patients, while inhibition of the HALLMARK HEME METABOLISM pathway, HALLMARK COAGULATION pathway, and the fat metabolism pathway may promote the prognosis in LIVER CANCER patients.

Background

According to the Global Cancer Statistics 2018 report, liver Cancer became the sixth most common diagnosed cancer and the fourth leading cause of cancer death in the world in 2018[1]. The highest incidence (mortality) of liver cancer is in East Asia, accounting for 35.5% of the global total. The main risk factors for liver cancer are chronic hepatitis B virus (HBV)[2-4], hepatitis C virus (HCV)[5-7], aflatoxin-contaminated food[8], heavy alcohol consumption[6, 9, 10], obesity[11], smoking[12] and type 2 diabetes[13, 14]. According to statistics, the risk factors of liver cancer formation are different in 53 countries and regions in the world. In most high-risk areas such as China and East Africa, chronic HBV infection and aflatoxin exposure are the main determinants of liver cancer. In contrast, HCV infection is the leading cause of liver cancer in Japan and Egypt [15, 16]. Mongolia has the highest incidence of liver cancer in the world, HBV and HCV virus infection, HBV co-infection with HCV or hepatitis virus, and alcoholism are all the reasons for the high incidence in Mongolia[17]. For low-risk liver cancer areas, the increase in obesity rates is the cause of the increase in liver cancer incidence.

The internationally recognised TNM cancer staging method divides cancers into stage I, stage II, stage III and stage IV[18]. People divide cancer into early, middle and late stages, if corresponding to TNM stage, phase I is early-stage, phase II and III are middle-stages, and phase IV is late-stage. Most cancers are diagnosed at the late stage, especially liver cancer. According to Traditional Chinese medicine, the liver is the "Officer Of General"[19], playing a general role in the body and allocating the functions of various parts of the body. Liver as a "general", must have a strong ability to resist pressure, the daily task is heavy, with a variety of pressure, minor illness and pain must not be shown. Modern medical research has shown that there is no pain sensation in the liver, and even if the liver disease had happened, the body does not feel it. Therefore, no matter from the perspective of traditional Chinese medicine or modern medicine, the clinical manifestation of liver disease is very slight, Therefore, in clinical practice, most patients with liver cancer are diagnosed at a late stage[20-23]. The cure rate of early-stage liver cancer is very optimistic, so if we can make a diagnosis in stage III, stage II or even stage I, then the treatment of tumour will not be as desperate as it is now.

Alpha-fetoprotein(AFP) is currently the only clinically used biomarker for early diagnosis of liver cancer. AFP was discovered more than 50 years ago and is a less accurate diagnostic biomarker for liver cancer. 32% and 59% of liver cancer patients have normal AFP levels[24]. Therefore, finding new diagnostic biomarkers of liver cancer is of great significance for the accurate diagnosis of liver cancer. For cancer patients, the prognosis and survival time of cancer is of great significance for improving the quality of life of patients, as well as the diagnosis and treatment scheme adopted. Currently, the therapeutic indications for the treatment of liver cancer are more concerned with tumour size and number of nodules and less concerned with its aggressiveness[25]. Compared with a small and aggressive liver cancer node, patients

with multiple large but not aggressive liver cancer nodules may have a better prognosis, hence the current prognostic criteria are not accurate. Therefore, if we can find the genes related to the prognosis of liver cancer, it is of great significance for both treatment and improvement of patients' quality of life. In this paper, the data of liver cancer patients in TCGA and GEO databases were taken as objects to search for diagnostic biomarkers and prognostic biomarkers of liver cancer through data mining, to improve the accuracy of early diagnosis of liver cancer, achieve early detection and treatment, and reduce mortality. At the same time, through the accurate judgment of the prognosis of liver cancer patients, adjuvant treatment to determine the plan.

Methods

Microarray data

Liver Cancer dataset was obtained from TCGA(<https://portal.gdc.cancer.gov/>), including 50 normal liver tissue samples, 371 samples of liver cancer, coupled with the clinical data. Another gene expression profiling datasets (GSE25097), including information of 243 normal samples and 268 tumour samples, were downloaded from Gene Expression Omnibus(GEO, <https://www.ncbi.nlm.nih.gov/geo/database>), and measured in this array(Platform: GPL10687 Rosetta/Merck Human RSTA Affymetrix 1.0 microarray, Custom CDF).

Data processing

The original microarray data of GSE25097 dataset and TCGA LIHC dataset were respectively analysed with R language to screen the differentially expressed genes (DEGs), and $\text{adj.p-value} < 0.05$ and $|\log\text{FC}| > 2$ were used as the cut-off criteria. We used The Draw Venn Diagram online tool(<http://bioinformatics.psb.ugent.be/webtools/Venn/>) to calculate the intersection of two differentially DEGs derived from two different datasets, which was common differentially expressed genes(the Common DEGs).

Volcanic maps and heat maps of DEGs obtained from GEO and TCGA databases

We used packet pheatmap, packet ggplot2 and other R packets to draw heat maps and volcanic maps of DEGs.

Gene Ontology and Reactome pathway analysis

GO analysis of the obtained DEGs was carried out using the package clusterProfiler. The package ReactomePA was used for enrichment analysis of the obtained DEGs in the Reactome pathway. $P < 0.05$ was considered as statistically significant.

Protein-protein interactions network

The Retrieval of Interacting Genes/Protein (STRING) is an online protein interaction tool (<https://string-db.org/>) that can integrate known protein-protein correlation data to build upstream and downstream relationships between proteins[26]. We put the Common DEGs into STRING software to build and visualise the protein-protein interaction (PPI) network. Also, we applied the cytoHubba in Cytoscape software (Cytoscape_v3.6.1) to screen hub genes. Top 22 genes with connection degree > 5 were selected as hub gene.

Draw the ROC curve of hub gene

Using the package pROC, Receiver Operating Characteristic (ROC) curve analysis is performed on 22 hub genes. AUC > 90% is set as a cutoff value to determine the diagnostic significance of hub genes.

Survival and statistical analysis

For survival analysis, gene expression values were divided into low and high expression groups by using the R language. The hazard ratio (HR) was determined via a Cox regression model, and survival curves were plotted from Kaplan-Meier estimations. $P < 0.05$ was considered to indicate a statistically significant difference.

Hub gene expression

The package ggpubr was used to draw boxplot to observe the distribution of hub genes in liver cancer tissue and normal liver tissue.

Gene set enrichment analysis

Gene set enrichment analysis (GSEA) is a computational method that assesses whether a priori defined a set of genes shows statistically significant, concordant differences between two biological states[27]. To investigate the role of ESR1, SPP1 and MYH11 gene in liver cancer, we used the package clusterProfiler to conduct single-gene GSEA analysis. $p\text{-value} < 0.05$ and $p\text{-adjust} < 0.05$ were regarded as the cut-off criteria.

Results

Identification of DEGs

The GSE25097 dataset was processed with R, DEGs with $\text{adj.}p\text{ value} < 0.05$ and $|\log\text{FC}| > 2$, including 790 genes were screened for further investigation (Figure 1, Supplement table 1). The TCGA LIHC dataset was analysed with R x64 3.6.1, using the package DEGseq2, $\text{adj.}p\text{ value} < 0.05$ and $|\text{FC}| > 2$ were regarded as the cut-off criteria. We got 2162 genes that met the standards (Figure 1, Supplement table 2). To confirm the reliability of DEGs in liver cancer, we obtained Common DEGs of the two datasets, including 102 genes (Figure 1, Table 1). The volcano map (Figure 2A, Figure 2C) and heat map (Figure 2B, Figure 2D)

were drawn based on the differential genes obtained from data set GSE25097 and TCGA LIHC, respectively..

GO and Reactome pathway analysis of the DEGs

We used GO analysis and Reactome Pathway analysis to conduct enrichment analysis of 102 Common DEGs. GO analysis includes biological process(BP) analysis, cellular component(CC) analysis and molecular function(MF) analysis(Figure 3a). BP analysis showed that liver cancer had changes in hormone metabolism(Cellular hormone metabolic process, Hormone metabolic process), cell reaction to copper, cadmium ions and inorganic substances, and detoxification function(Cellular response to cadmium ion, Cellular response to metal ion, Cellular response to inorganic substance, Cellular response to copper ion, Detoxification of copper ion, Detoxification). CC analysis showed that the Collagen trimer and Collagen-containing extracellular matrix of liver cancer cells was changed. Moreover, the MF analysis showed that the patients with liver cancer had an abnormal expression of oxidoreductase activity and molecular binding function(Glycosaminoglycan binding, Cytokine receptor binding, iron ion binding, extracellular matrix binding, carbohydrate binding). The results showed that the changes of cellular collagen were observed at the cellular level, the changes of hormone metabolism, reaction to metal ions and detoxification were observed at the biological function, and the changes of molecular binding and oxidoreductase activity were observed at the molecular level.

Through Reactome enrichment analysis(Figure 3B), we found that liver cancer changes in biological oxidations, reactions and conjugation ability to metal ions(phase II-conjugation of compounds, metallothioneins bind metals, response to metal ions), and also affects growth hormone receptor signalling.

Comparing the results of the two enrichment analyses, we found that the information obtained by the two enrichment analyses was consistent, that is, the two analyses were enriched with changes in hormone metabolism, biological oxidation, cell reaction to metal ions and other aspects in patients with liver cancer.

PPI network analysis and screening for hub Genes

102 the DEGs were input into STRING to build PPI network (Figure 4A). Then the PPI network diagram was imported into Cytoscape(3.2.1). CytoHubba of app plug-in was used to calculate the Degree Value and other parameter values(Supplement table 3). Genes whose Degree Value is greater than or equal to 5 are taken as Hub genes, and a total of 22 Hub genes were obtained(Table 2). See Figure 4B for the relationship between 22 Hub genes.

Expression of Hub genes in patients with liver cancer

The expression of 22 Hub genes in liver cancer and normal liver tissues was analysed, and it was found that SPP1, AURKA, NQO1, NUSAP1, TOP2A, UBE2C, AFP, GMNN, PTTG1, RRM2, UBE2T, GPC3, SPARCL1

etc.(Figure 5A) 13 genes were highly expressed in liver cancer tissues. However, ESR1, CXCL12, FOS, DCN, EGR1, SOCS3, CYP1A2, FOSB, PCK1 *etc.* 9 genes were low expressed in liver cancer tissues(Figure 5B).

ROC curve analysis of Hub genes

ROC curve analysis was performed on 22 Hub genes using the package pROC. AUC > 90% was taken as the cutoff value, and it was found that 16 of the 22 Hub genes with AUC > 90% were SPP1, AURKA, CXCL12, FOS, NUSAP1, TOP2A, UBE2C, AFP, DCN, GMNN, PTTG1, RRM2, SOCS3, FOSB, PCK1, SPARCL1 respectively. The expression level of these genes has high accuracy in distinguishing normal tissue from liver cancer tissue, and could be a potential "tumour biomarker". At the same time, it can be used as a biomarker for the diagnosis of liver cancer, which has important significance for the accurate diagnosis of liver cancer(Figure 6).

The survival curve of Hub genes

Survival curves were plotted from Kaplan-Meier estimations(Figure 7), The Cox regression model was used to calculate Hazard Ratio(HR) of Hub genes for liver cancer patients. The results showed that among these Hub genes, the expression levels of ESR1, SPP1 and FOSB genes were closely related to the survival time of liver cancer patients, with statistically significant differences($p < 0.05$). HR values were 0.88, 1.1 and 0.88, respectively, that is, ESR1 and FOSB were low-risk factors, while SPP1 was a high-risk factor.

GSEA revealed the biological function that affects the survival time of liver cancer

Single-gene GSEA was used to investigate biological pathways and biological functions related to survival time(Figure 8). Figure 8A shows all the related pathways of ESR1, FOSB and SPP1 genes. Figure 8B shows the commonly related pathways of ESR1, FOSB and SPP1 genes. Figure 8B a1, b1 and c1 are the three common pathways of ESR1, FOSB and SPP1 genes. Figure 8B a2 and c2 are the seven common pathways of ESR1 and SPP1 genes, and Figure 8B a3 and b3 are the four common pathways of ESR1 and FOSB genes.

The three common pathways enriched by ESR1, FOSB and SPP1 genes are HALLMARK MYC TARGETS V1, HALLMARK G2M CHECKPOINT and HALLMARK E2F TARGETS pathways(Figure 8Ba, b, c). According to the information in Figure 8B, we can find that high expression of ESR1 and FOSB can activate these three pathways, while high expression of SPP1 can inhibit these three pathways. However, in liver cancer tissues, ESR1 and FOSB genes were low in expression, while SPP1 genes were high in expression (see Figure 5). Therefore, changes in the expression levels of ESR1, FOSB and SPP1 genes in liver cancer inhibited all three pathways.

Seven common pathways were obtained by enrichment analysis of ESR1 and SPP1 genes. They are HALLMARK PANCREAS BETA CELLS, HALLMARK ESTROGEN RESPONSE LATE, HALLMARK ADIPOGENESIS, HALLMARK FATTY ACID METABOLISM, HALLMARK BILE ACID METABOLISM, HALLMARK XENOBIOTIC METABOLISM and HALLMARK PEROXISOME pathway, the high expression of

ESR1 gene can activate the HALLMARK PANCREAS BETA CELLS and HALLMARK ESTROGEN RESPONSE LATE pathways. Five pathways, namely, HALLMARK ADIPOGENESIS, HALLMARK FATTY ACID METABOLISM, HALLMARK BILE ACID METABOLISM, HALLMARK XENOBIOTIC METABOLISM, and HALLMARK PEROXISOME, were inhibited, while SPP1 gene was just opposite to ESR1 gene (Figure 8Ba2, c2). In liver cancer, the ESR1 gene is a low expression, while the SPP1 gene is a high expression (see Figure 5). Therefore, changes in ESR1 and SPP1 gene expression in liver cancer activated the HALLMARK ADIPOGENESIS, HALLMARK FATTY ACID METABOLISM, HALLMARK BILE ACID METABOLISM, HALLMARK XENOBIOTIC METABOLISM, and HALLMARK PEROXISOME pathway. While both the HALLMARK PANCREAS BETA CELLS and HALLMARK ESTROGEN RESPONSE LATE pathway were suppressed.

The four common pathways enriched by ESR1 and FOSB genes are respectively HALLMARK MYC TARGETS V2, HALLMARK HEME METABOLISM, HALLMARK COAGULATION and HALLMARK UV RESPONSE DN pathways. High expression of ESR1 and FOSB can activate the HALLMARK MYC TARGETS V2 pathway and inhibit the three pathways, including HALLMARK HEME METABOLISM, HALLMARK COAGULATION and HALLMARK UV RESPONSE_DN(Figure 8B a3, b3). However, in liver cancer, both ESR1 and FOBS genes were low expressed(see Figure 5). Therefore, changes in the expression levels of ESR1 and FOBS genes in liver cancer inhibited HALLMARK MYC TARGETS V2 pathway, while HALLMARK HEME METABOLISM, HALLMARK COAGULATION, HALLMARK UV RESPONSE_DN three pathways were activated.

Discussion

The incidence and mortality of liver cancer are both high, especially in East Asia, where the incidence(mortality) accounts for more than 1/3 of the world[17]. The most patients with liver cancer do not seek medical treatment until they have symptoms in the late stage of the liver cancer, so the early diagnosis of liver cancer is of great significance for the treatment of liver cancer. At present, alpha-fetoprotein(AFP) is a diagnostic biomarker used in the clinical diagnosis of liver cancer. AFP was discovered 50 years ago as a diagnostic biomarker of liver cancer, and there are problems of inaccurate diagnosis. According to investigations, 32% to 59% of liver cancer patients have normal AFP levels[24]. Therefore, we need to find new and more accurate biomarkers for liver cancer diagnosis. Also, the prognosis of cancer patients is of great significance to the quality of life and treatment of patients. Therefore, the search for prognostic biomarkers is also of great significance for tumour patients. In order to achieve this goal, this paper uses data mining analysis to find diagnostic biomarkers and prognostic biomarkers of liver cancer.

First, we obtained liver cancer data sets from the TCGA database, including 50 normal liver tissue samples and 371 liver cancer samples. The GSE25097 dataset was obtained from the GEO database, consisting of 243 non-tumour tissue samples and 268 liver cancer samples. After DEGs analysis, 102 Common DEGs were obtained from TCGA and GSE25097 data sets, Then GO analysis and Reactome Pathway analysis were used to conduct enrichment analysis on 102 Common DEGs, The results showed

that liver cancer showed changes in collagen at the cellular level, changes in hormone metabolism and reaction to metal ions at the biological function, and abnormalities in molecular binding and oxidoreductase activity at the molecular level (Figure 3)

PPI network was constructed for 102 Common DEGs to find the correlation between genes, and 22 Hub genes were screened from 102 Common DEGs based on Degree value (Table 2). ROC curve is a curve reflecting the relationship between sensitivity and specificity, which is of great significance for accurate diagnosis of diseases [28]. We used ROC curve to analyse 22 Hub genes with AUC greater than 90% as the threshold, and obtained 16 Hub genes, namely, SPP1, AURKA, CXCL12, FOS, NUSAP1, TOP2A, UBE2C, AFP, DCN, GMNN, PTTG1, RRM2, SOCS3, FOSB, PCK1 and SPARCL1. The expression levels of the 16 Hub genes in liver cancer can accurately distinguish normal liver tissue from liver cancer, so the 16 genes can be used as diagnostic biomarkers of liver cancer for the early diagnosis of liver cancer, including AFP, which is currently used in clinical practice. At the same time, we observed the effect of the 22 Hub genes on the survival time of liver cancer patients and calculated the risk coefficient, and found that the expression levels of ESR1, SPP1 and FOSB genes in the 22 Hub genes had a significant impact on the survival time of liver cancer patients ($p < 0.05$), with HR values of 0.88, 1.1 and 0.88, respectively, indicating that ESR1 and FOSB are low-risk genes, while SPP1 is high-risk genes. However, the AUC value of ESR1 is 68.7% (Figure 6a), which showed that the accurate diagnosis rate of ESR1 gene is low, not suitable for use as diagnostic biomarkers. As a result, only the FOSB gene and SPP1 gene are suitable for use as prognostic biomarkers of liver cancer, the FOSB is a low-risk gene, while the SPP1 is a high-risk gene. In other words, the survival rate of liver cancer patients with high expression of FOSB is higher than that of patients with low expression. In comparison, the survival rate of patients with high expression of SPP1 is lower than that of patients with low expression.

Finally, single-gene GSEA analysis was performed on the three prognostic genes, ESR1, SPP1 and FOSB, that affect the survival time of liver cancer patients (Figure 8), to explore the mechanism affecting the prognosis of liver cancer patients. Through analysis, we found that there were three pathways closely related to ESR1, FOSB and SPP1 genes (Figure 8B a1, b1, c1), seven pathways closely related to ESR1 and SPP1 genes (Figure 8B a2, c2), and four pathways closely related to ESR1 and FOSB genes (Figure 8B a3, b3).

The three common pathways related to ESR1, FOSB, and SPP1 genes are respectively HALLMARK MYC TARGETS V1, HALLMARK G2/M CHECKPOINT, and HALLMARK E2F TARGETS. Among them, high expression of ESR1 and FOSB genes can activate these three pathways, while high expression of SPP1 gene inhibits these three pathways (Figure 8a1, b1, c1). At the same time, since ESR1 and FOSB genes are low-risk factors, high expression of ESR1 and FOSB genes can activate these three pathways, SPP1 gene is a high-risk factor high expression of SPP1 can inhibit these three pathways (Figure 8 a, b, t). Hence activation of these three pathways is conducive to improving the survival time of liver cancer patients. MYC TARGETS V1 pathway is a new anticancer target [29-31], which is closely related to cell proliferation, differentiation and cell cycle. In contrast, the G2/M CHECKPOINT pathway [32] and HALLMARK E2F TARGETS pathway are all closely related to the cell cycle [33], that is, the three pathways are cell cycle-

related pathways. That is to say, patients with liver cancer whose cell cycle pathway is activated have a better prognosis.

The seven common pathways related to ESR1 and SPP1 genes are respectively HALLMARK PANCREAS BETA CELLS, HALLMARK ESTROGEN RESPONSE LATE, HALLMARK ADIPOGENESIS, HALLMARK FATTY ACID METABOLISM, HALLMARK BILE ACID METABOLISM, HALLMARK XENOBIOTIC METABOLISM and HALLMARK PEROXISOME. Among them, the high ESR1 gene expression can activate the HALLMARK PANCREAS BETA CELLS and HALLMARK ESTROGEN RESPONSE LATE pathways, Inhibiting the five pathways of HALLMARK ADIPOGENESIS, HALLMARK FATTY ACID METABOLISM, HALLMARK BILE ACID METABOLISM, HALLMARK XENOBIOTIC METABOLISM, HALLMARK PEROXISOME. In contrast, SPP1 gene was just opposite to ESR1 gene(Figure 8 a2, c2). Similarly, the ESR1 gene is low-risk factors, SPP1 genes are high-risk factors, so the liver cancer patients that HALLMARK PANCREAS BETA CELLS and HALLMARK ESTROGEN RESPONSE LATE pathway activated, and The HALLMARK ADIPOGENESIS, HALLMARK FATTY ACID METABOLISM, HALLMARK BILE ACID METABOLISM, HALLMARK XENOBIOTIC METABOLISM and HALLMARK PEROXISOME pathway inhibited have a better prognosis. By analysing these pathways, we find that these seven pathways can be divided into four aspects in terms of function:

1. The prognosis of liver cancer patients that HALLMARK PANCREAS BETA CELLS pathway activated is better than that of liver cancer patients with this pathway inhibited. HALLMARK PANCREAS BETA CELLS pathway is restrained and Islet cell dysfunction are the important cause of type 2 diabetes, this also means that patients with liver cancer complicated with type 2 diabetes have a poor prognosis. Patients with type 2 diabetes are also at high-risk population of liver cancer. This conclusion is consistent with the conclusion of the epidemiological investigation of liver cancer[17].
2. The prognosis of liver cancer patients that HALLMARK ESTROGEN RESPONSE LATE pathway activated is better. Clinically, "Palmar Erythema" and "spider nevus" appear in the palms of some patients with cancer[34] and severe liver dysfunction[35]. These manifestations are caused by the decreased metabolism of estrogen in the liver, resulting in excessive estrogen[36] in the blood and stimulation of capillary arterial congestion and dilation. In other words, the presence of "Palmar Erythema" and "spider arachnoid" is a manifestation of inhibition of estrogen pathway, and the prognosis of liver cancer patients with " Palmar Erythema " and " spider nevus " is poor. Also, in clinical practice, some male liver cancer patients due to the inhibition of estrogen metabolism lead to the increase of estrogen level in the blood, resulting in their breast development, the prognosis of this kind of liver cancer patients is not good[37].
3. The prognosis is better in patients with liver cancer whose fat metabolism-related pathways(HALLMARK ADIPOGENESIS, HALLMARK FATTY ACID METABOLISM, HALLMARK BILE ACID METABOLISM, HALLMARK PEROXISOME) are inhibited. Epidemiological investigation shows that obesity is one of the important factors causing liver cancer, so for the prognosis of liver cancer patients, the prognosis of patients with fat metabolism-related pathways being inhibited is better.
4. Patients whose HALLMARK XENOBIOTIC METABOLISM is inhibited have a good prognosis.

Four common pathways related to ESR1 and FOSB genes are activation of HALLMARK MYC TARGETS V2, inhibition of HALLMARK HEME METABOLISM, HALLMARK COAGULATION and HALLMARK UV RESPONSE DN pathways, respectively. Both ESR1 and FOSB genes were low-risk factors, so the patients

whose HALLMARK MYC TARGETS V2 pathway was activated, and the HALLMARK HEME METABOLISM, HALLMARK COAGULATION and the HALLMARK UV RESPONSE DN pathway were suppressed had a better prognosis. HALLMARK E2F TARGETS V2 pathway is closely related to the cell cycle, that is to say, the prognosis of liver cancer patients with activated cell cycle pathway is better, which is consistent with the conclusion we got before. Also, HALLMARK HEME METABOLISM pathway regulates HEME METABOLISM, and the main product of HEME METABOLISM is bile pigment, which includes many compounds such as bilirubin, biliverdin, bilinogen and choline. Under normal circumstances, bile pigment is mainly excreted with bile. Bilirubin is the main pigment in bile, which is orange-yellow. The metabolic disorder of bilirubin is closely related to clinical hepatobiliary diseases. If the HALLMARK HEME METABOLISM pathway is activated, the heme will be massively metabolised into bilirubin, resulting in excessively high concentration in plasma and diffused into tissues, resulting in jaundice (easily seen in sclera, skin, *etc.*). According to the conclusion of data analysis in this paper, patients with inhibited HALLMARK METABOLISM pathway have a good prognosis. In contrast, those with activated HALLMARK METABOLISM pathway have a poor prognosis. After activated HALLMARK METABOLISM pathway, patients will have jaundice symptoms, so liver cancer patients with jaundice have a poor prognosis. Patients with suppressed HALLMARK COAGULATION pathway have a good prognosis, The HALLMARK COAGULATION pathway mainly regulates the COAGULATION function. Abnormal COAGULATION function in liver cancer patients is a common clinical symptom, mainly related to the lack of COAGULATION factor, thrombocytopenia and increased vascular permeability. The results of data analysis in this paper show that the prognosis of patients with inhibited blood clotting function is better than that of patients with this function activated.

Through analysis, we found that the prognosis of liver cancer patients is mainly related to the following functions: 1. It is closely related to the regulation of the cell cycle, and the patients with activated cell cycle have a good prognosis. 2. Liver cancer patients with activated HALLMARK PANCREAS BETA CELLS pathway have a good prognosis, while liver cancer patients with type 2 diabetes have a poor prognosis. 3. Patients with activated hepatocellular estrogen pathway have a good prognosis, and those with "liver palm", "spider nevus" and abnormal breast development have a poor prognosis. 4. Liver cancer patients whose fat metabolism-related pathways are inhibited have a good prognosis. 5. Liver cancer patients whose HALLMARK XENOBIOTIC METABOLISM pathway is inhibited have a good prognosis. 6. The prognosis of liver cancer patients is good if HALLMARK HEME METABOLISAM pathway is inhibited, and poor if liver cancer has "jaundice". 7. Liver cancer patients whose HALLMARK COAGULATION pathway is inhibited have a good prognosis.

Conclusions

SPP1, AURKA, NUSAP1, TOP2A, UBE2C, AFP, GMNN, PTTG1, RRM2, SPARCL1 genes that are highly expressed in liver cancer tissues, CXCL12, FOS, DCN, SOCS3, FOSB, AND PCK1 genes that are low expressed in liver cancer tissues can be used as biomarkers of a liver cancer diagnosis. Among them, FOBS and SPP1 genes can also be used as prognostic biomarkers of liver cancer. FOBS is low-expressed in liver cancer, with HR value of 0.88, which is a low-risk gene, while SPP1 is high-expressed in liver

cancer, with HR value of 1.1, which is a high-risk gene. Activation of the cell cycle-related pathway, PANCREAS BETA CELLS pathway, and the estrogen signalling pathway in liver cancer patients, while inhibition of the HALLMARK HEME METABOLISM pathway, HALLMARK COAGULATION pathway, and the fat metabolism pathway may improve the prognosis of liver cancer patients.

Abbreviations

AFP:alpha-fetoprotein; BP:biological process; CC:cellular component; Common DEGs:the common differentially expressed genes; DEGs:the differentially expressed genes; GEO:Gene Expression Omnibus; GSEA:Gene set enrichment analysis; HBV:hepatitis B virus; HCV:hepatitis C virus; HR:The hazard ratio; MF:molecular function; ROC:Receiver Operating Characteristic curve; STRING:The Retrieval of Interacting Genes/Protein; TCGA:the cancer genome atlas;

Declarations

Consent for publication

Not applicable

Availability of data and materials

The datasets generated and/or analyzed during the current study are available in the [<https://www.ncbi.nlm.nih.gov/geo/database>] and <https://portal.gdc.cancer.gov/>]repositories.

Conflicts of interests

All authors declare that they have no conflict of interests.

Funding

This study was supported by grants from Harbin Applied Technology Research and Development Project(Grant/Award Number:2017RAXXJ085)

Authors` contributions

1. G. and H. T. designed the study. S. G. and J. G. carried out data acquisition and analysis. S. G. and H. T. wrote the manuscript. S. G. and J. G. contributed to preparing and making figures and tables. All authors read and approved the final manuscript.

Acknowledgements

Not applicable

Ethical statement

We certify that this manuscript is original and has not been published and will not be submitted elsewhere for publication while being considered by JCP. And the study is not split up into several parts to increase the quantity of submissions and submitted to various

journals or to one journal over time. No data have been fabricated or manipulated (including images) to support our conclusions. No data, text, or theories by others are presented as if they were our own. The submission has been received explicitly from all co-authors. And authors whose names appear on the submission have contributed sufficiently to the scientific work and therefore share collective responsibility and accountability for the results.

Author details

[1]Drug Engineering and Technology Research Center, Harbin University of Commerce, Harbin 150076, Heilongjiang, China. ²Heilongjiang Provincial Key Laboratory of Tumor Prevention and Antitumor Drugs, Harbin 150076, Heilongjiang, China. ³Department of pharmacy, The Fourth Affiliated Hospital of Harbin Medicine University, Harbin 150001, Heilongjiang, China.

References

1. Feng J, Li J, Wu L, Yu Q, Ji J, Wu J, et al. Emerging roles and the regulation of aerobic glycolysis in hepatocellular carcinoma. *J Exp Clin Cancer Res.* 2020;39:126.
2. Piñero F, Tanno M, Aballay Soteras G, Tisi Baña M, Dirchwolf M, Fassio E, et al. Argentinian clinical practice guideline for surveillance, diagnosis, staging and treatment of hepatocellular carcinoma. *Annals of hepatology.* 2020;19:546-69.
3. Feld JJ, Krassenburg LAP. What Comes First: Treatment of Viral Hepatitis or Liver Cancer? *Digestive diseases and sciences.* 2019;64:1041-9.
4. Ringelhan M, O'Connor T, Protzer U, Heikenwalder M. The direct and indirect roles of HBV in liver cancer: prospective markers for HCC screening and potential therapeutic targets. *The Journal of pathology.* 2015;235:355-67.
5. Pham C, Sin MK. Use of Electronic Health Records at Federally Qualified Health Centers: a Potent Tool to Increase Viral Hepatitis Screening and Address the Climbing Incidence of Liver Cancer. *Journal of cancer education : the official journal of the American Association for Cancer Education.* 2020.
6. Szabo G, Saha B, Bukong TN. Alcohol and HCV: implications for liver cancer. *Advances in experimental medicine and biology.* 2015;815:197-216.
7. Yao M, Yang JL, Wang L, Yao DF. [Carcinoembryonic type specific markers and liver cancer immunotherapy]. *Zhonghua gan zang bing za zhi = Zhonghua ganzangbing zazhi = Chinese journal of hepatology.* 2020;28:466-70.
8. Henry SH, Bosch FX, Bowers JC. Aflatoxin, hepatitis and worldwide liver cancer risks. *Advances in experimental medicine and biology.* 2002;504:229-33.

9. Grewal P, Viswanathan VA. Liver cancer and alcohol. *Clinics in liver disease*. 2012;16(4):839-50.
10. Turati F, Galeone C, Rota M, Pelucchi C, Negri E, Bagnardi V, et al. Alcohol and liver cancer: a systematic review and meta-analysis of prospective studies. *Annals of oncology : official journal of the European Society for Medical Oncology*. 2014;25:1526-35.
11. Saitta C, Pollicino T, Raimondo G. Obesity and liver cancer. *Annals of hepatology*. 2019;18:810-5.
12. Lee YC, Cohet C, Yang YC, Stayner L, Hashibe M, Straif K. Meta-analysis of epidemiologic studies on cigarette smoking and liver cancer. *International journal of epidemiology*. 2009;38:1497-511.
13. Tahmasebi-Birgani M, Ansari H, Carloni V. Defective mitosis-linked DNA damage response and chromosomal instability in liver cancer. *Biochimica et biophysica acta Reviews on cancer*. 2019;1872:60-5.
14. Wang Y, Wang B, Yan S, Shen F, Cao H, Fan J, et al. Type 2 diabetes and gender differences in liver cancer by considering different confounding factors: a meta-analysis of cohort studies. *Annals of epidemiology*. 2016;26:764-72.
15. Kudo M. Surveillance, diagnosis, treatment, and outcome of liver cancer in Japan. *Liver cancer*. 2015;4:39-50.
16. Yamashita T, Kaneko S. [Liver Cancer]. *Rinsho byori The Japanese journal of clinical pathology*. 2016;64:787-96.
17. 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: a cancer journal for clinicians*. 2018;68:394-424.
18. Nozaki Y, Yamamoto M, Ikai I, Yamamoto Y, Ozaki N, Fujii H, et al. Reconsideration of the lymph node metastasis pattern (N factor) from intrahepatic cholangiocarcinoma using the International Union Against Cancer TNM staging system for primary liver carcinoma. *Cancer*. 1998;83:1923-9.
19. Liu Li WQ, Liang Minlian, Li Le, Xu Yunsheng. From The "Officer Of General" Take A Fancy To Cure Cancer. *Journal of Zhejiang University of Chinese Medicine*. 2013;37:847.
20. Anwanwan D, Singh SK, Singh S, Saikam V, Singh R. Challenges in liver cancer and possible treatment approaches. *Biochimica et biophysica acta Reviews on cancer*. 2020;1873:188314.
21. Kaffe E, Magkrioti C, Aidinis V. Deregulated Lysophosphatidic Acid Metabolism and Signaling in Liver Cancer. *Cancers*. 2019;11:1626.
22. Mato JM, Elortza F, Lu SC, Brun V, Paradela A, Corrales FJ. Liver cancer-associated changes to the proteome: what deserves clinical focus? 2018;15:749-56.
23. Wrighton PJ, Oderberg IM, Goessling W. There Is Something Fishy About Liver Cancer: Zebrafish Models of Hepatocellular Carcinoma. *Cellular and molecular gastroenterology and hepatology*. 2019;8:347-63.
24. Charrière B, Maulat C, Suc B, Muscari F. Contribution of alpha-fetoprotein in liver transplantation for hepatocellular carcinoma. *World journal of hepatology*. 2016;8:881-90.

25. Muscari F, Maulat C. Preoperative alpha-fetoprotein (AFP) in hepatocellular carcinoma (HCC): is this 50-year biomarker still up-to-date? *Transl Gastroenterol Hepatol.* 2020;5:3.
26. Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, et al. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. 2017;45:D362-d8.
27. Subramanian A, Kuehn H, Gould J, Tamayo P, Mesirov JP. GSEA-P: a desktop application for Gene Set Enrichment Analysis. *Bioinformatics (Oxford, England).* 2007;23:3251-3.
28. Obuchowski NA, Bullen JA. Receiver operating characteristic (ROC) curves: review of methods with applications in diagnostic medicine. *Physics in medicine and biology.* 2018;63:07tr1.
29. Dang CV. MYC on the path to cancer. *Cell.* 2012;149:22-35.
30. Hsieh AL, Walton ZE, Altman BJ, Stine ZE, Dang CV. MYC and metabolism on the path to cancer. *Semin Cell Dev Biol.* 2015;43:11-21.
31. Stine ZE, Walton ZE, Altman BJ, Hsieh AL, Dang CV. MYC, Metabolism, and Cancer. *Cancer Discov.* 2015;5:1024-39.
32. Oshi M, Takahashi H. G2M Cell Cycle Pathway Score as a Prognostic Biomarker of Metastasis in Estrogen Receptor (ER)-Positive Breast Cancer. *Int J Mol Sci* 2020;21(8):2921.
33. De Meyer T, Bijsmans IT, Van de Vijver KK, Bekaert S, Oosting J, Van Criekinge W, et al. E2Fs mediate a fundamental cell-cycle deregulation in high-grade serous ovarian carcinomas. *The Journal of pathology.* 2009;217:14-20.
34. Maekawa M. Palmar erythema as a sign of cancer. *Cleveland Clinic journal of medicine.* 2017;84:666-7.
35. Li H, Wang R, Méndez-Sánchez N, Peng Y, Guo X, Qi X. Impact of spider nevus and subcutaneous collateral vessel of chest/abdominal wall on outcomes of liver cirrhosis. *Archives of medical science : AMS.* 2019;15:434-48.
36. Serrao R, Zirwas M, English JC. Palmar erythema. *American journal of clinical dermatology.* 2007;8:347-56.
37. Kuper H, Mantzoros C, Lagiou P, Tzonou A, Tamimi R, Mucci L, et al. Estrogens, testosterone and sex hormone binding globulin in relation to liver cancer in men. *Oncology.* 2001;60:355-60.

Tables

Due to technical limitations, table 1 and 2 is only available as a download in the Supplemental Files section.

Figures

GSE25097

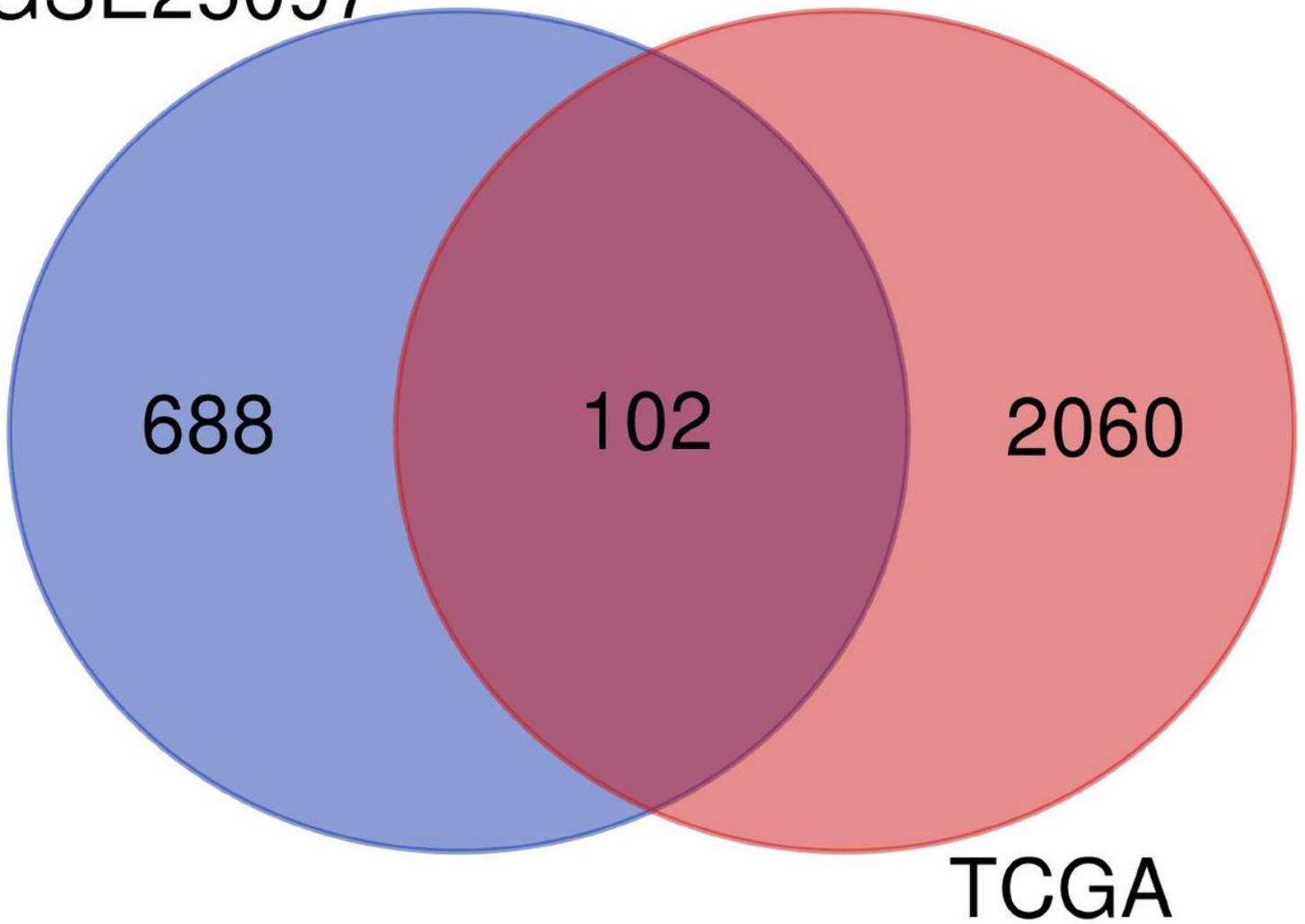


Figure 1

To confirm the reliability of DEGs in liver cancer, we obtained Common DEGs of the two datasets, including 102 genes

GSE25097

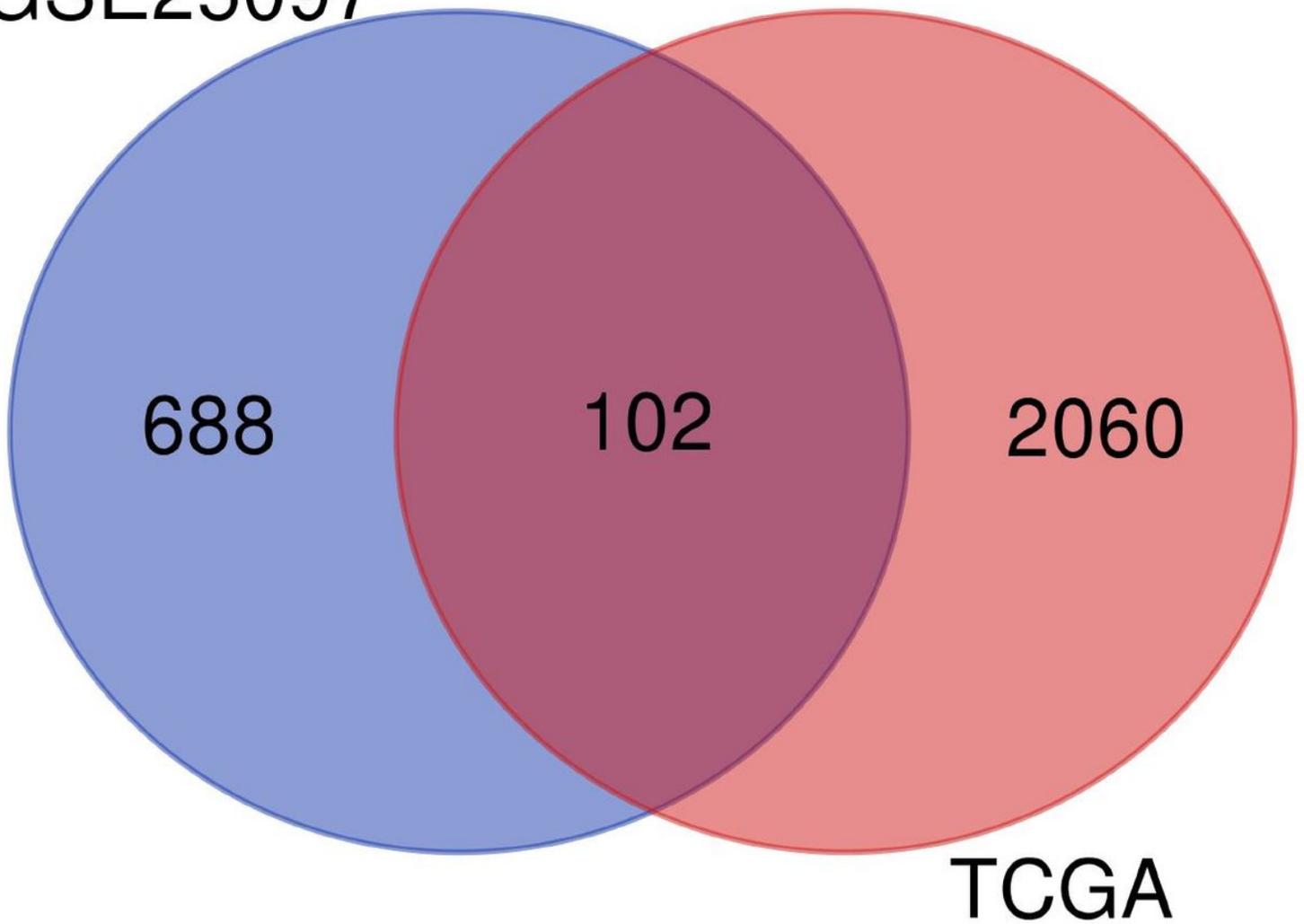


Figure 1

To confirm the reliability of DEGs in liver cancer, we obtained Common DEGs of the two datasets, including 102 genes

GSE25097

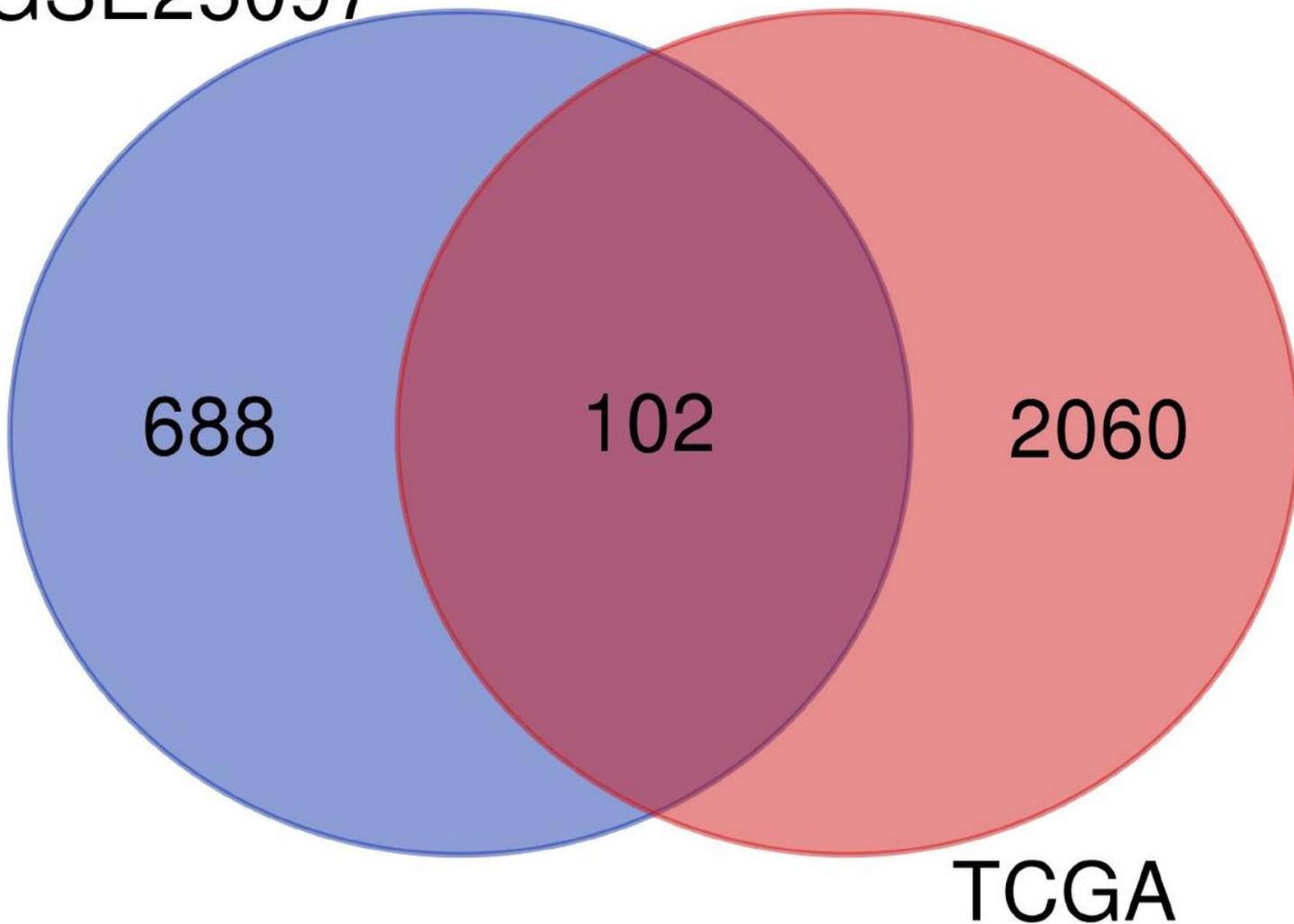


Figure 1

To confirm the reliability of DEGs in liver cancer, we obtained Common DEGs of the two datasets, including 102 genes

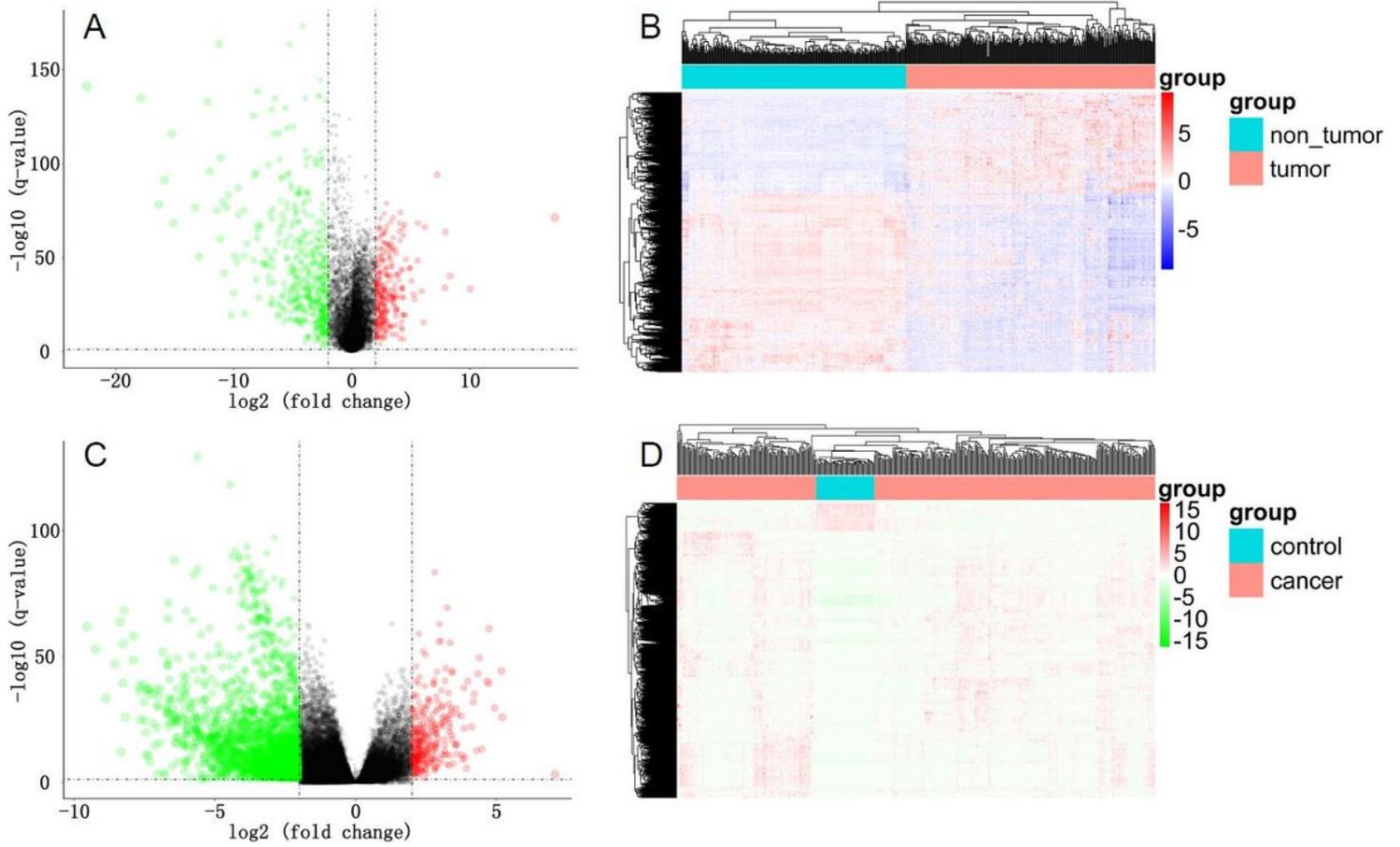


Figure 2

The volcano map (Figure 2A, Figure 2C) and heatmap (Figure 2B, Figure 2D) were drawn based on the differential genes obtained from data set GSE25097 and TCGA LIHC, respectively.

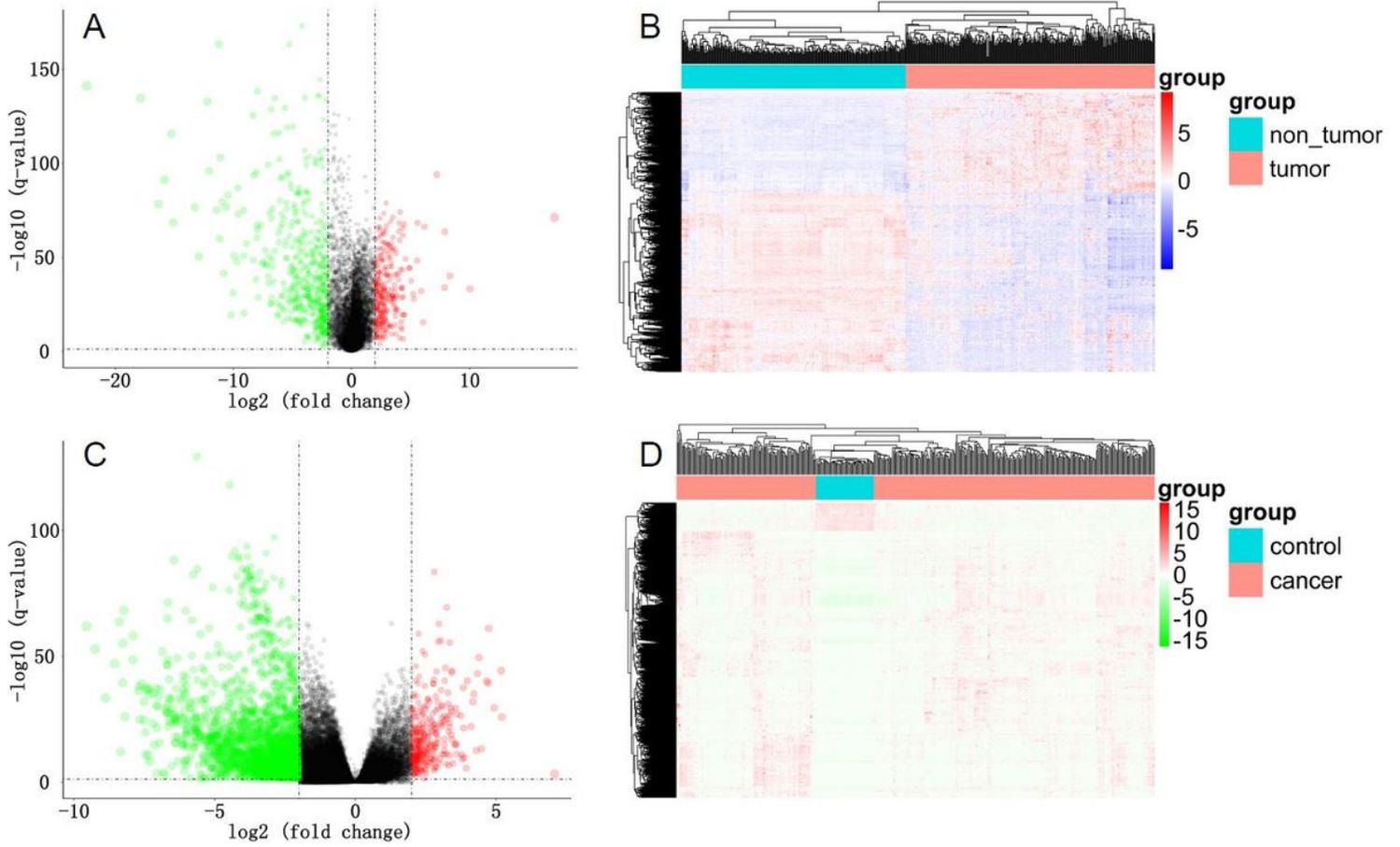


Figure 2

The volcano map (Figure 2A, Figure 2C) and heatmap (Figure 2B, Figure 2D) were drawn based on the differential genes obtained from data set GSE25097 and TCGA LIHC, respectively.

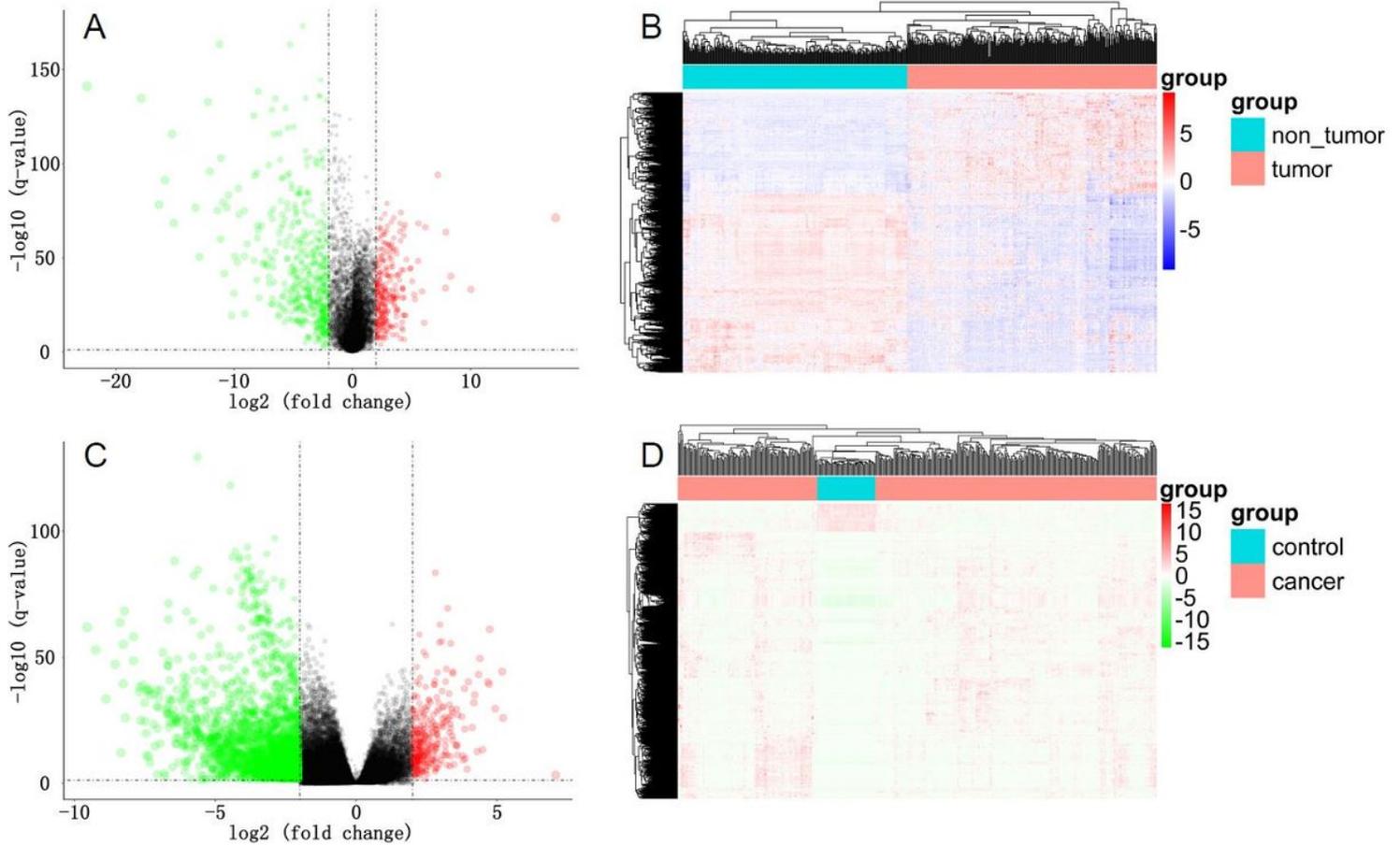


Figure 2

The volcano map (Figure 2A, Figure 2C) and heat map (Figure 2B, Figure 2D) were drawn based on the differential genes obtained from data set GSE25097 and TCGA LIHC, respectively.

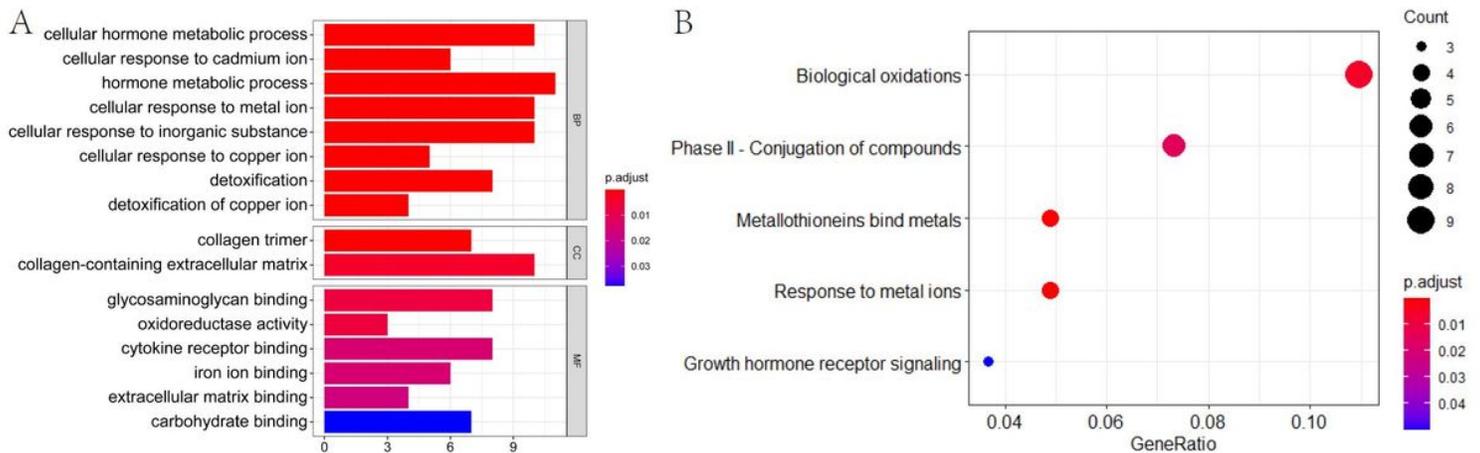


Figure 3

The results showed that liver cancer showed changes in collagen at the cellular level, changes in hormone metabolism and reaction to metal ions at the biological function, and abnormalities in molecular binding and oxidoreductase activity at the molecular level

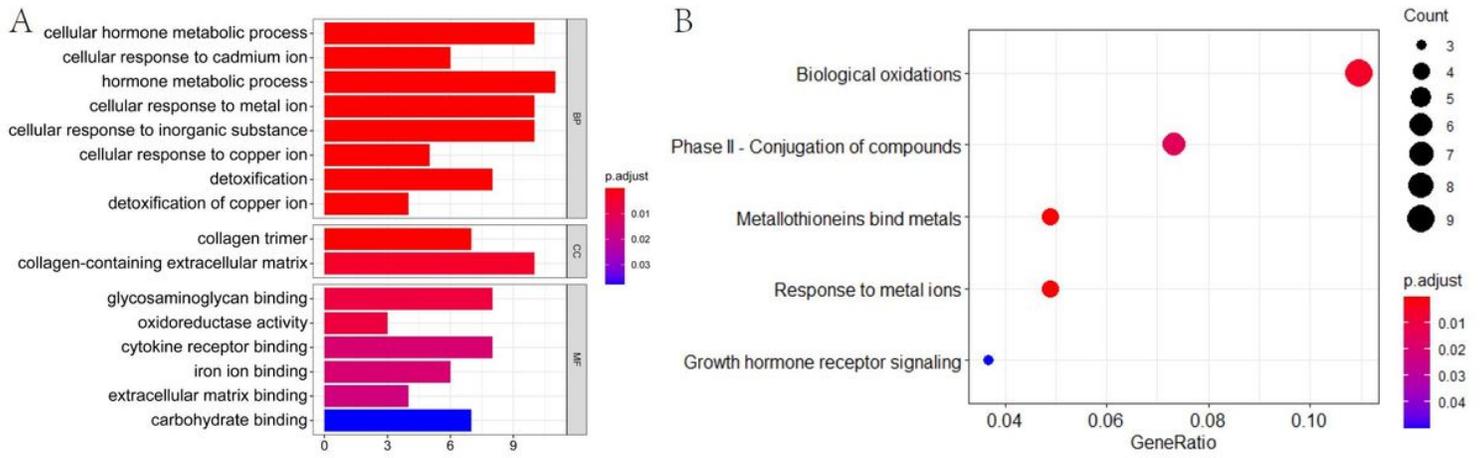


Figure 3

The results showed that liver cancer showed changes in collagen at the cellular level, changes in hormone metabolism and reaction to metal ions at the biological function, and abnormalities in molecular binding and oxidoreductase activity at the molecular level

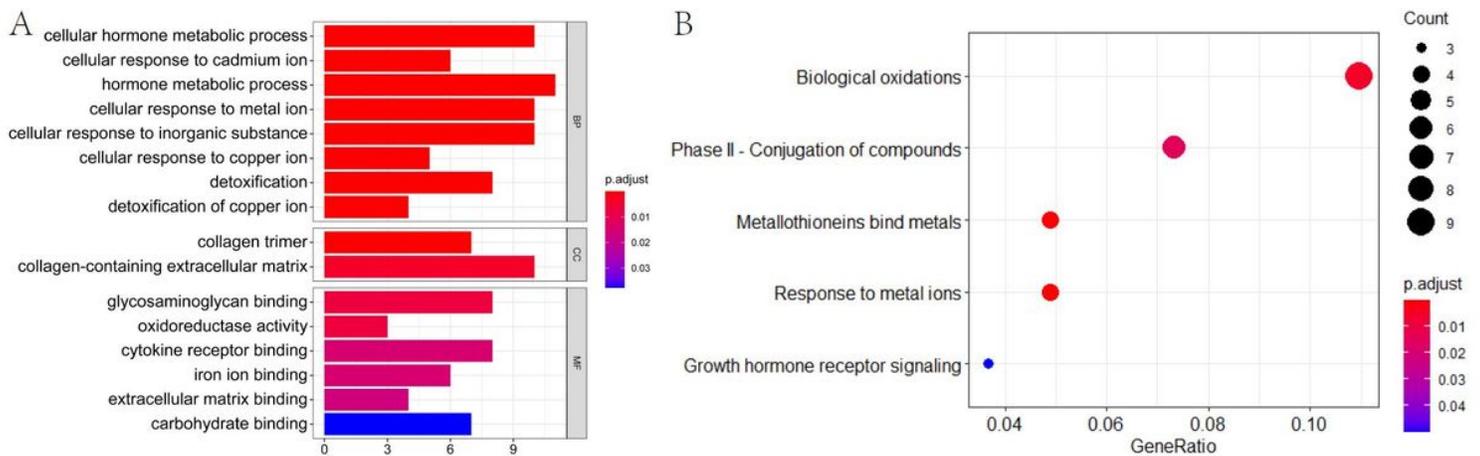


Figure 3

The results showed that liver cancer showed changes in collagen at the cellular level, changes in hormone metabolism and reaction to metal ions at the biological function, and abnormalities in molecular binding and oxidoreductase activity at the molecular level

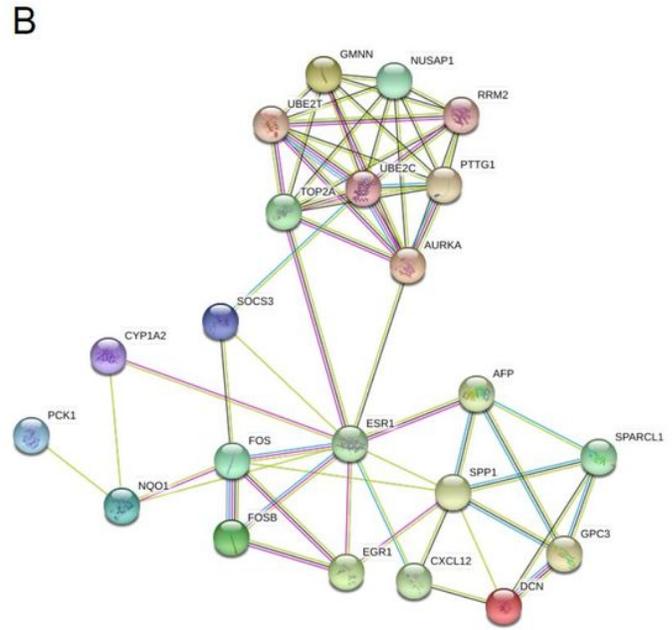
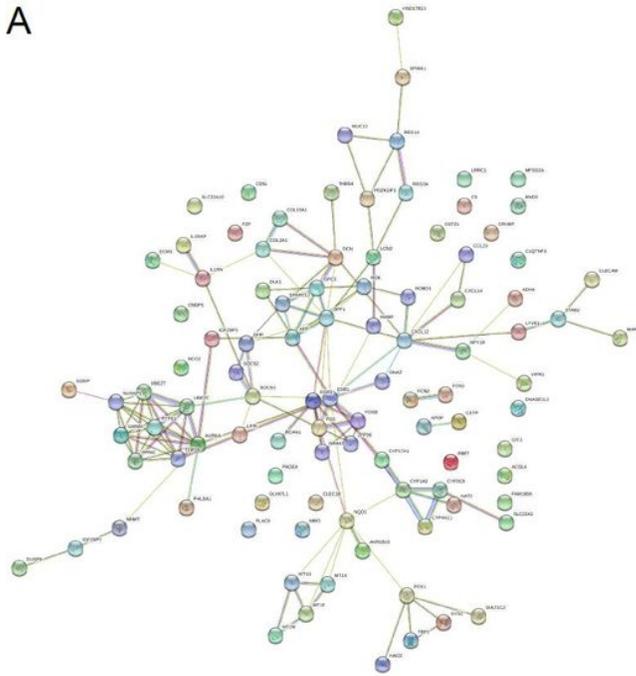


Figure 4

102 the DEGs were input into STRING to build PPI network (Figure 4A). Then the PPI network diagram was imported into Cytoscape(3.2.1). CytoHubba of app plug-in was used to calculate the Degree Value and other parameter values(Supplement table 3). Genes whose Degree Value is greater than or equal to 5 are taken as Hub genes, and a total of 22 Hub genes were obtained(Table 2). See Figure 4B for the relationship between 22 Hub genes.

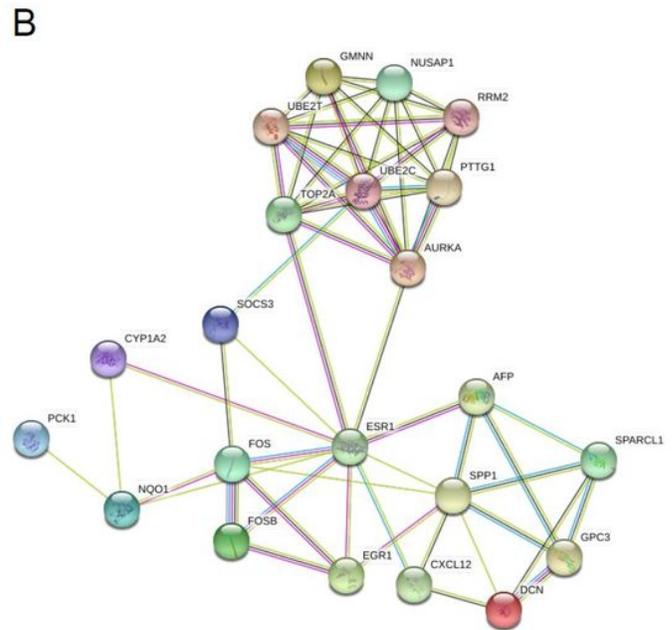
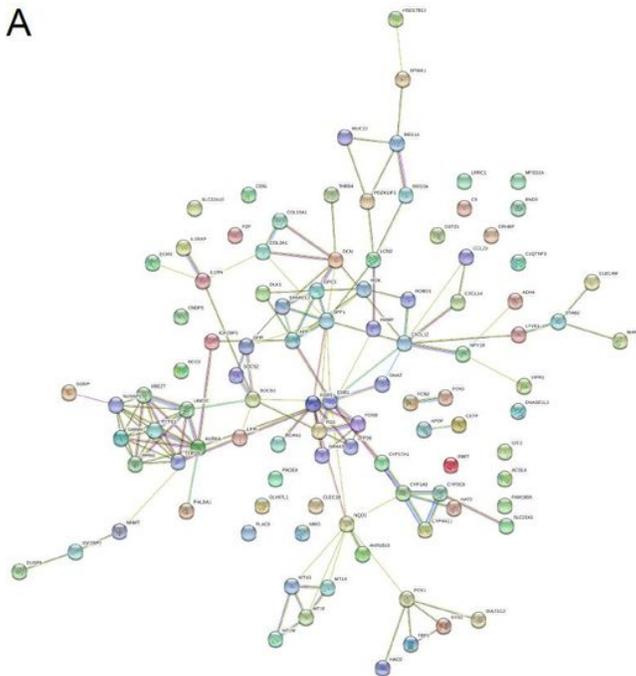


Figure 4

102 the DEGs were input into STRING to build PPI network (Figure 4A). Then the PPI network diagram was imported into Cytoscape(3.2.1). CytoHubba of app plug-in was used to calculate the Degree Value and other parameter values(Supplement table 3). Genes whose Degree Value is greater than or equal to 5 are taken as Hub genes, and a total of 22 Hub genes were obtained(Table 2). See Figure 4B for the relationship between 22 Hub genes.

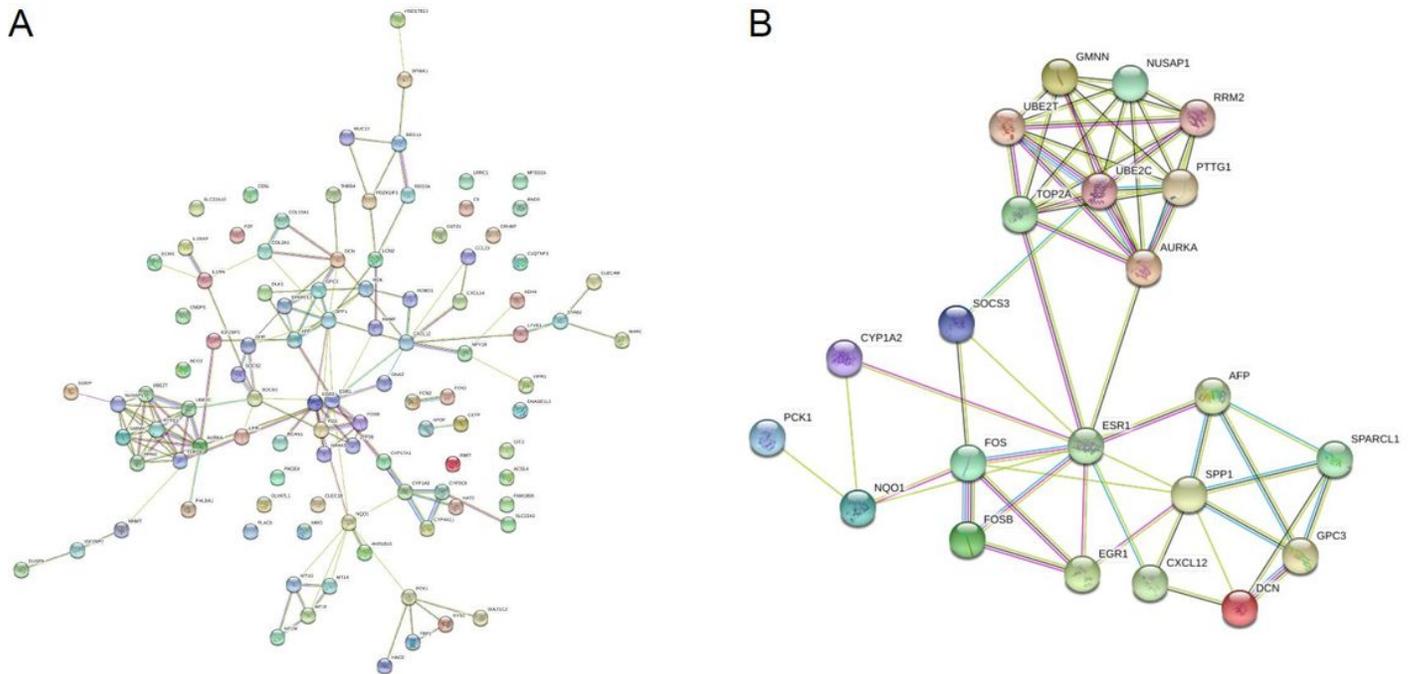


Figure 4

102 the DEGs were input into STRING to build PPI network (Figure 4A). Then the PPI network diagram was imported into Cytoscape(3.2.1). CytoHubba of app plug-in was used to calculate the Degree Value and other parameter values(Supplement table 3). Genes whose Degree Value is greater than or equal to 5 are taken as Hub genes, and a total of 22 Hub genes were obtained(Table 2). See Figure 4B for the relationship between 22 Hub genes.

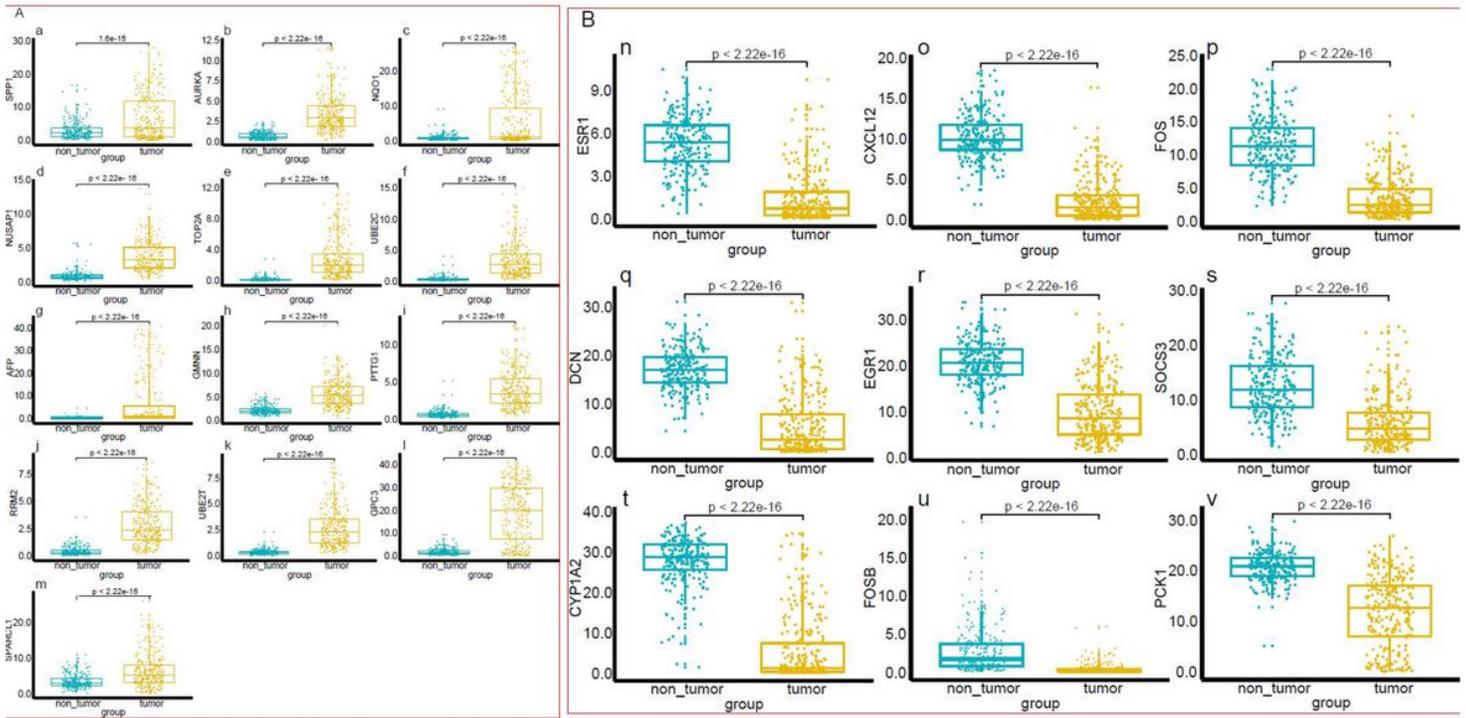


Figure 5

Therefore, changes in the expression levels of ESR1, FOBS and SPP1 genes in liver cancer inhibited all three pathways

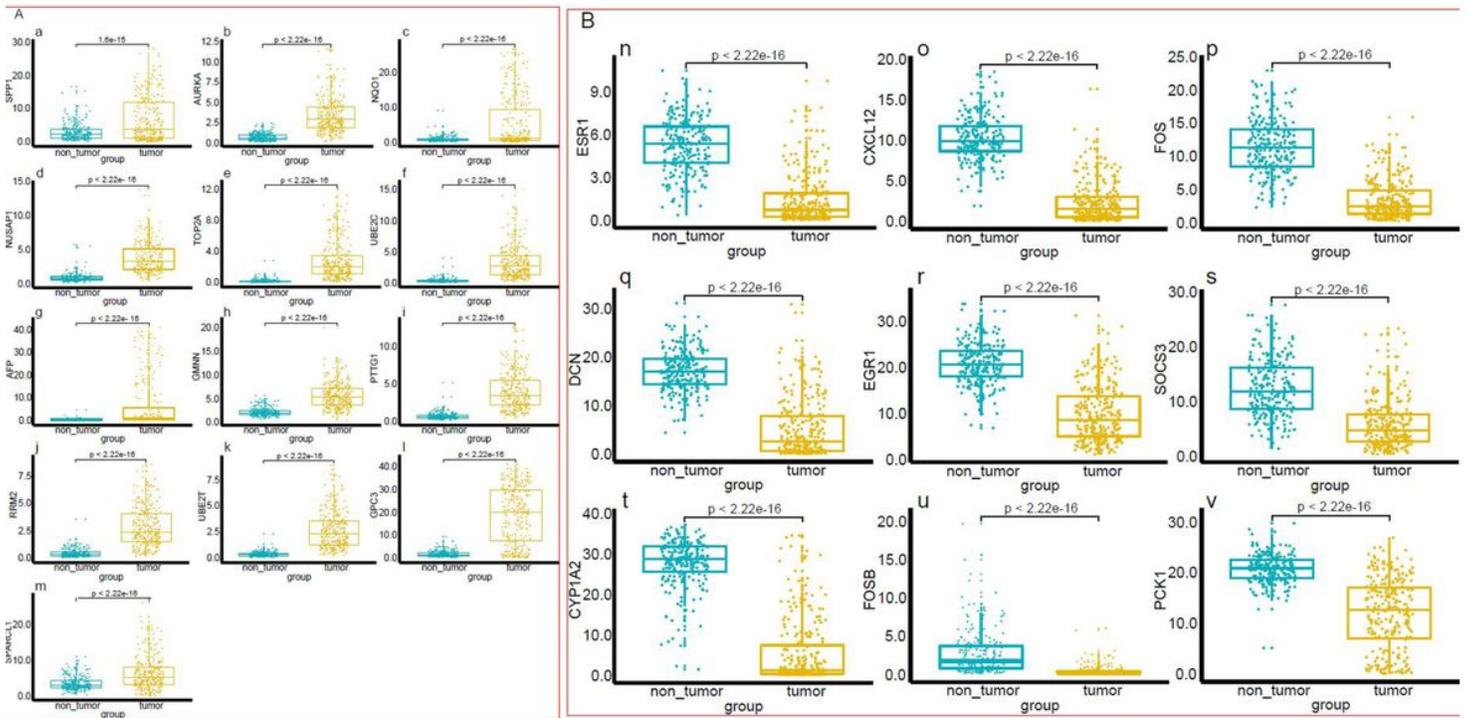


Figure 5

Therefore, changes in the expression levels of ESR1, FOBS and SPP1 genes in liver cancer inhibited all three pathways

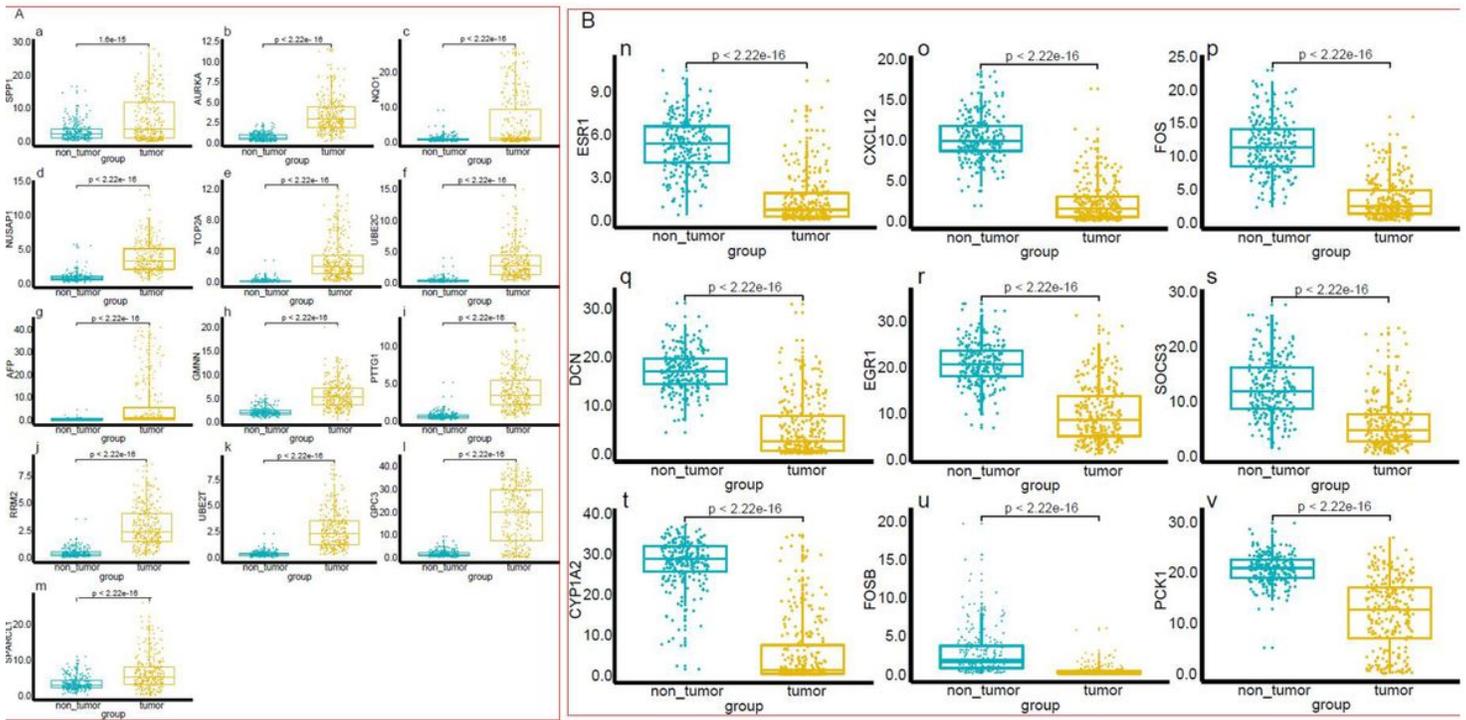


Figure 5

Therefore, changes in the expression levels of ESR1, FOSB and SPP1 genes in liver cancer inhibited all three pathways

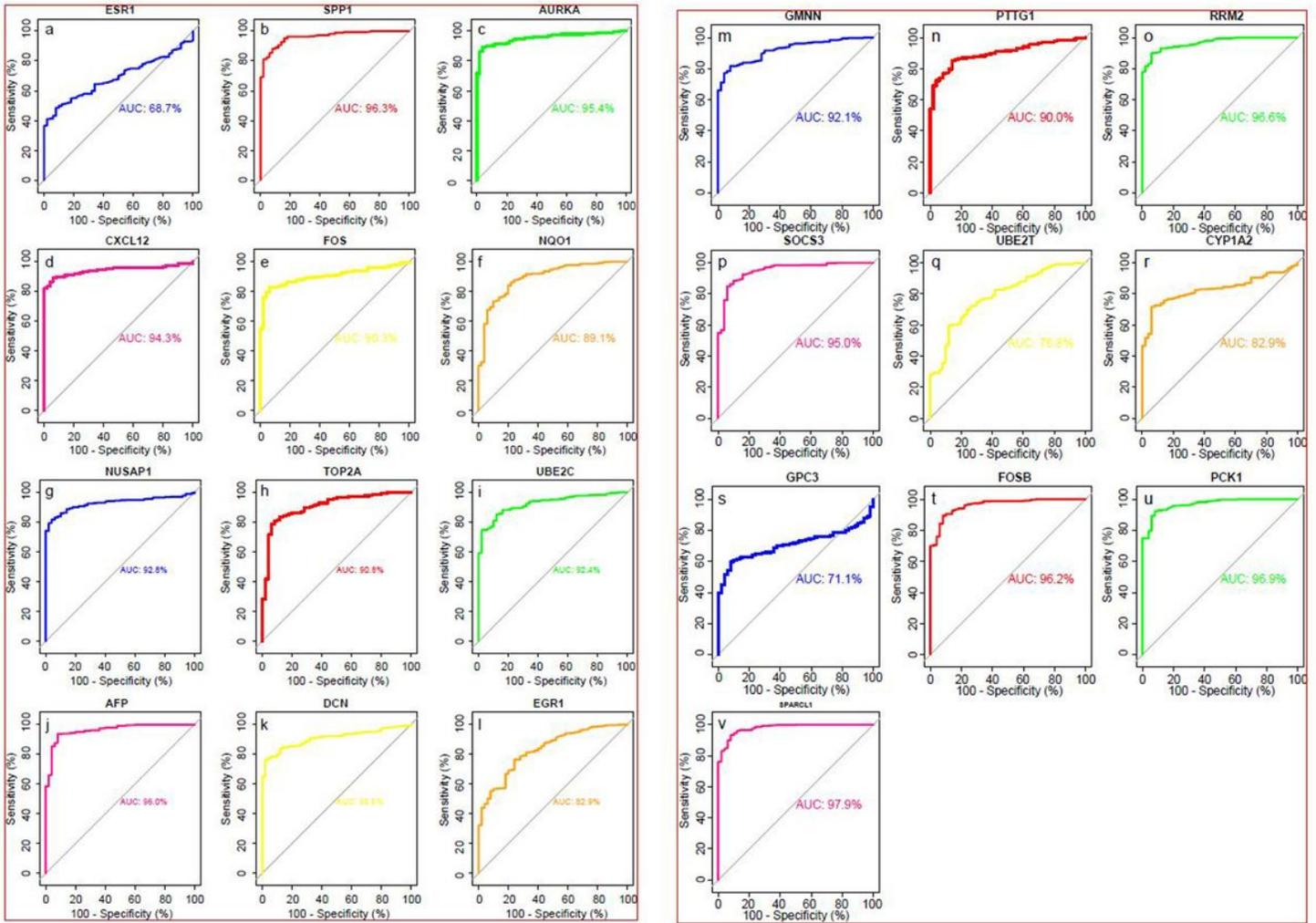


Figure 6

At the same time, it can be used as a biomarker for the diagnosis of liver cancer, which has important significance for the accurate diagnosis of liver cancer

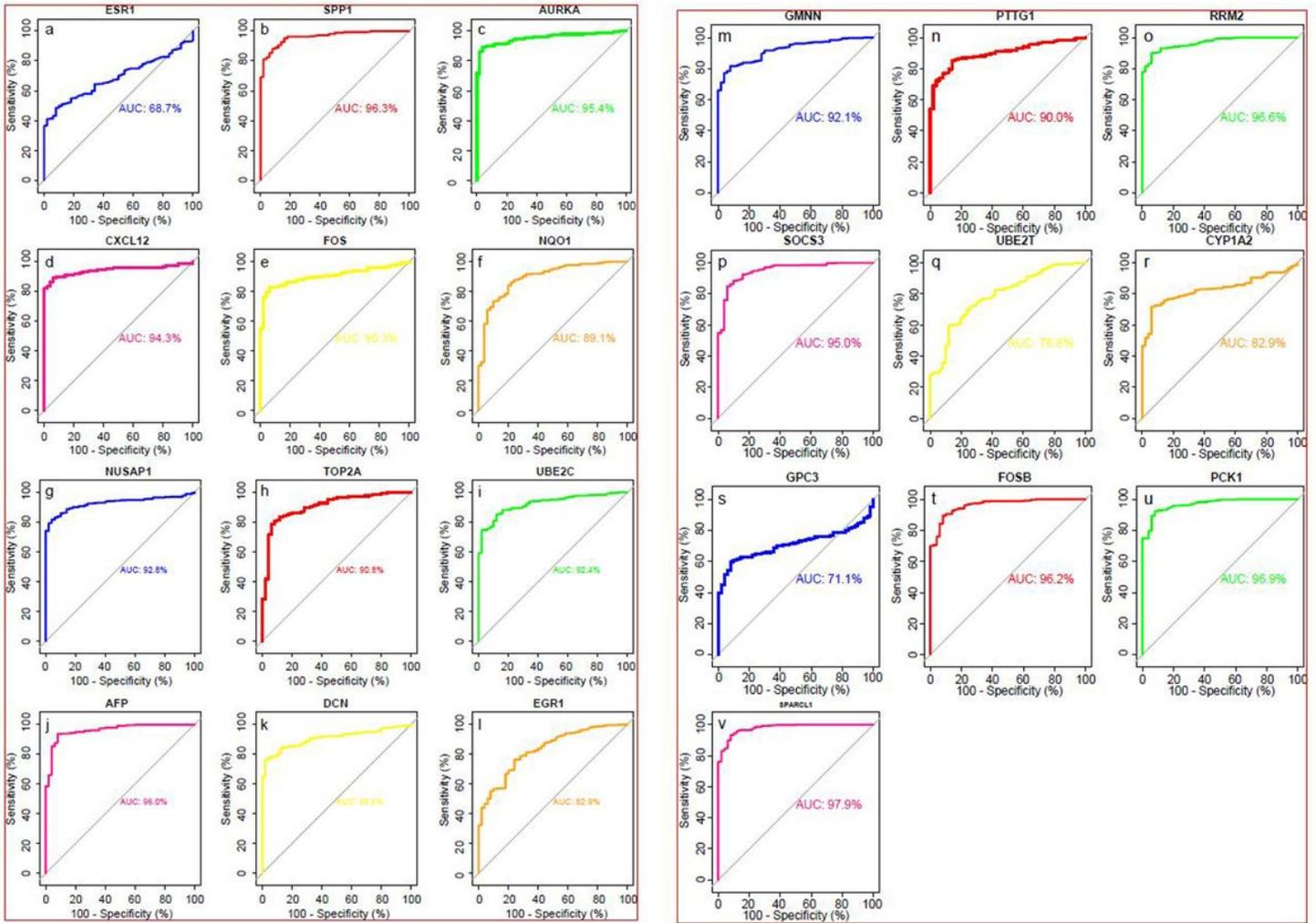


Figure 6

At the same time, it can be used as a biomarker for the diagnosis of liver cancer, which has important significance for the accurate diagnosis of liver cancer

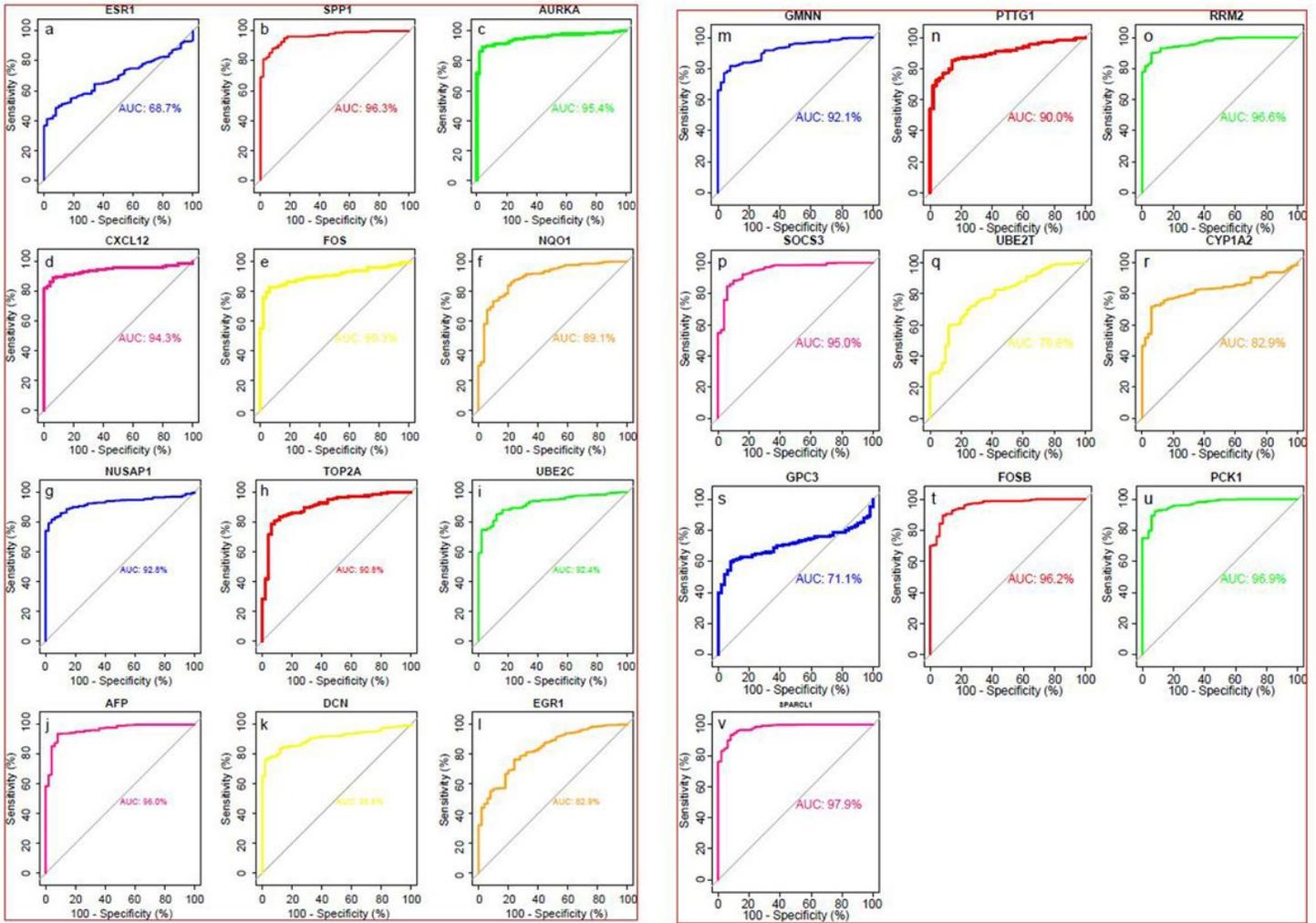


Figure 6

At the same time, it can be used as a biomarker for the diagnosis of liver cancer, which has important significance for the accurate diagnosis of liver cancer

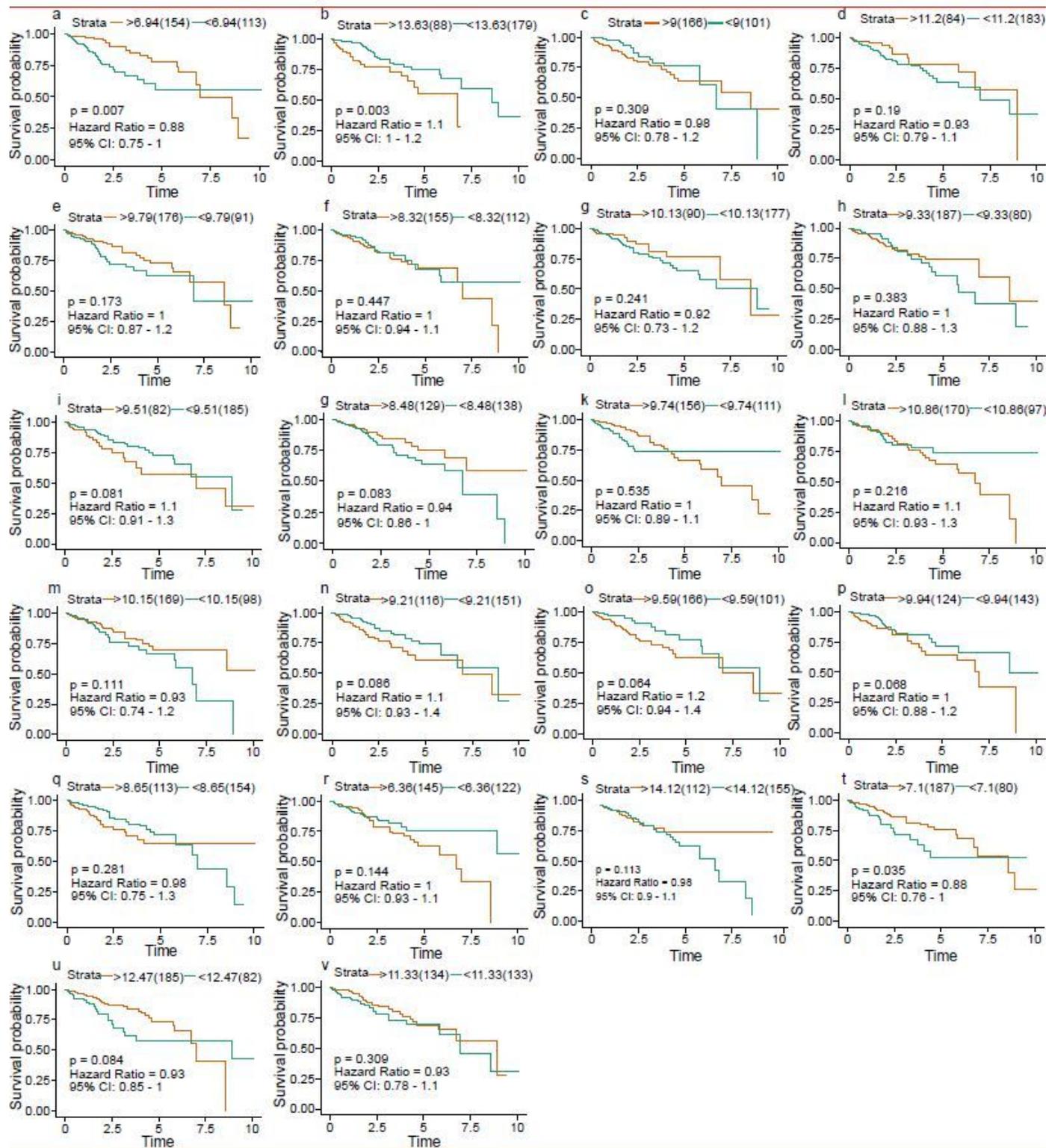


Figure 7

Survival curves were plotted from Kaplan-Meier estimations

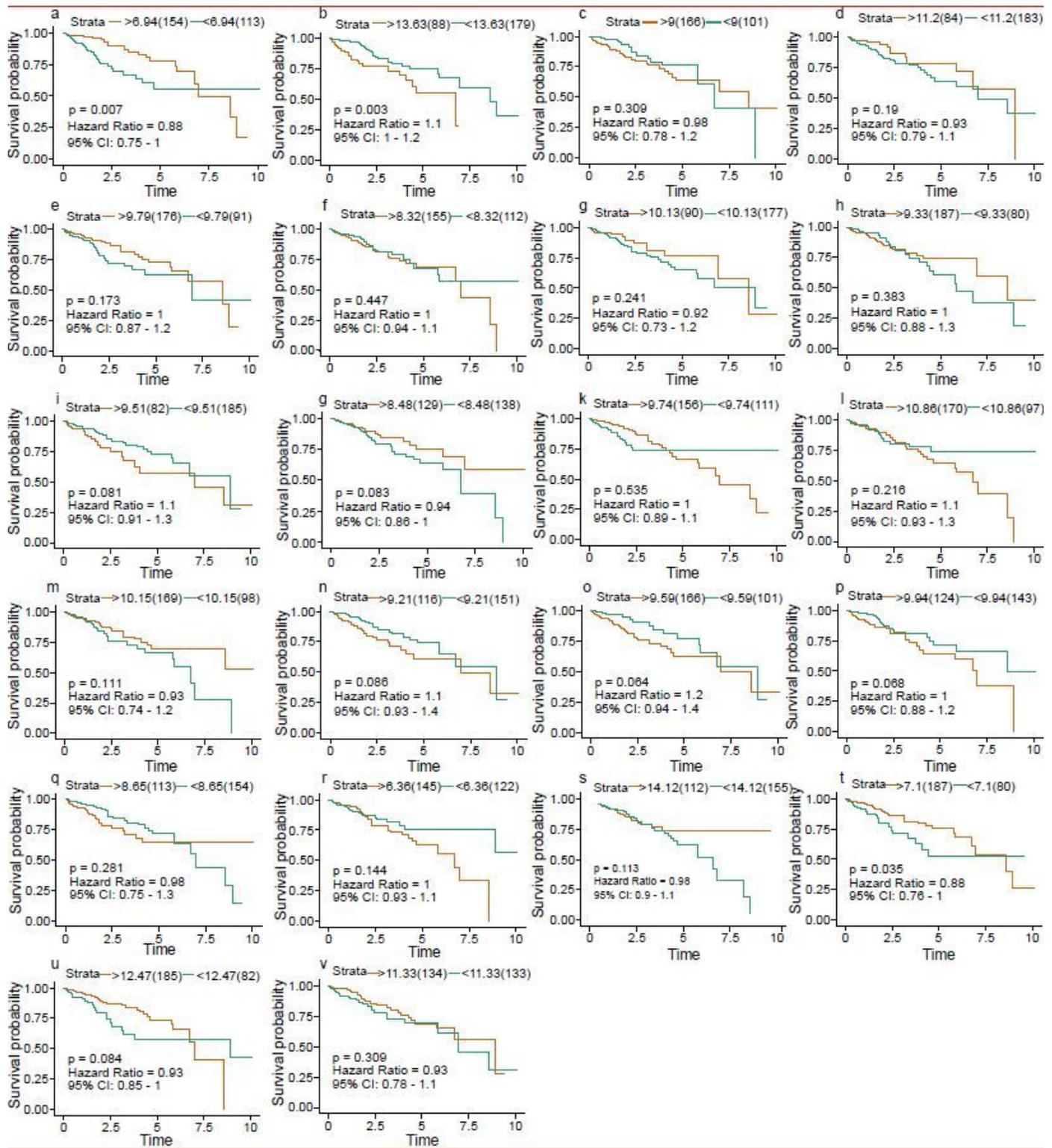


Figure 7

Survival curves were plotted from Kaplan-Meier estimations

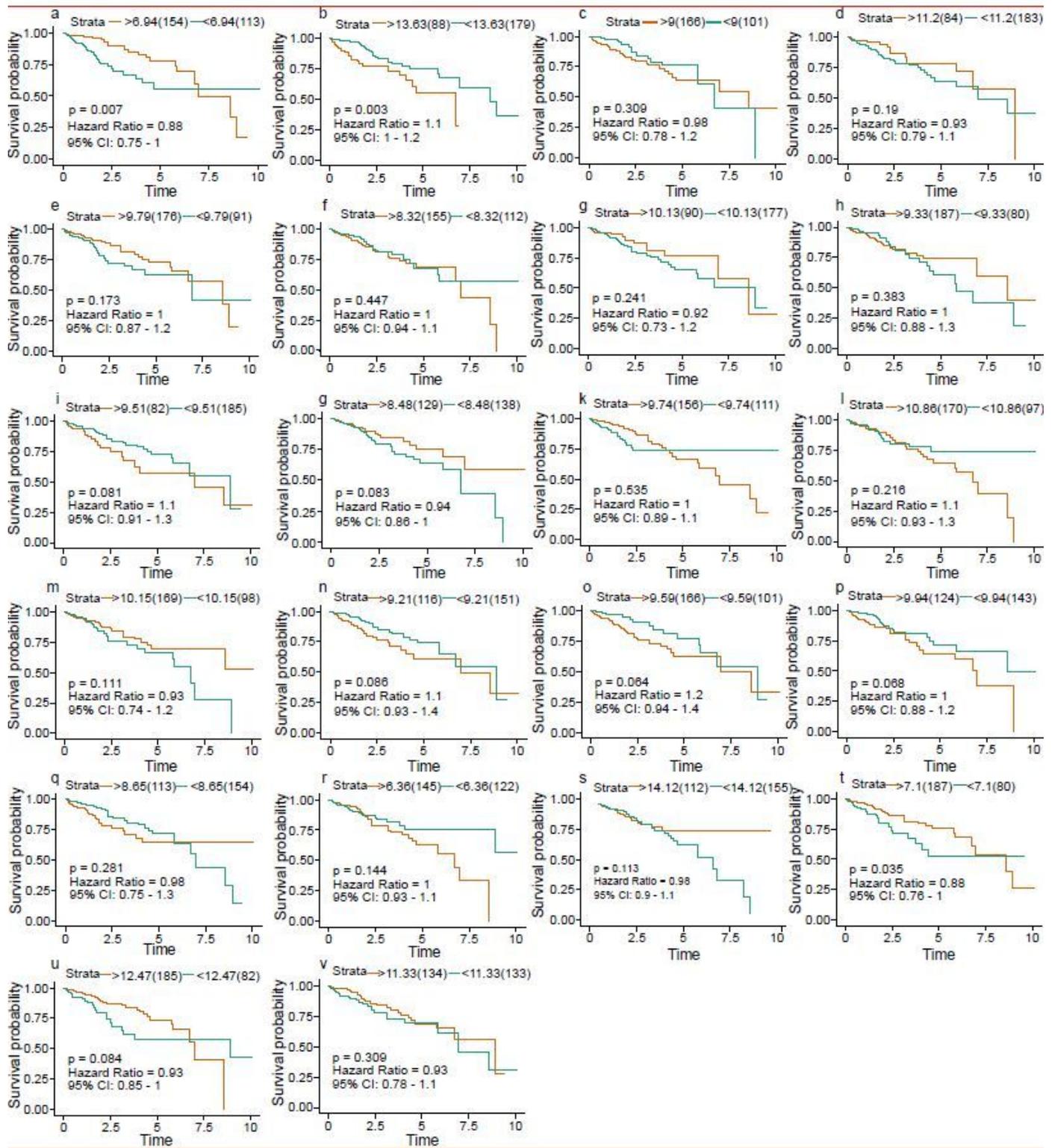


Figure 7

Survival curves were plotted from Kaplan-Meier estimations

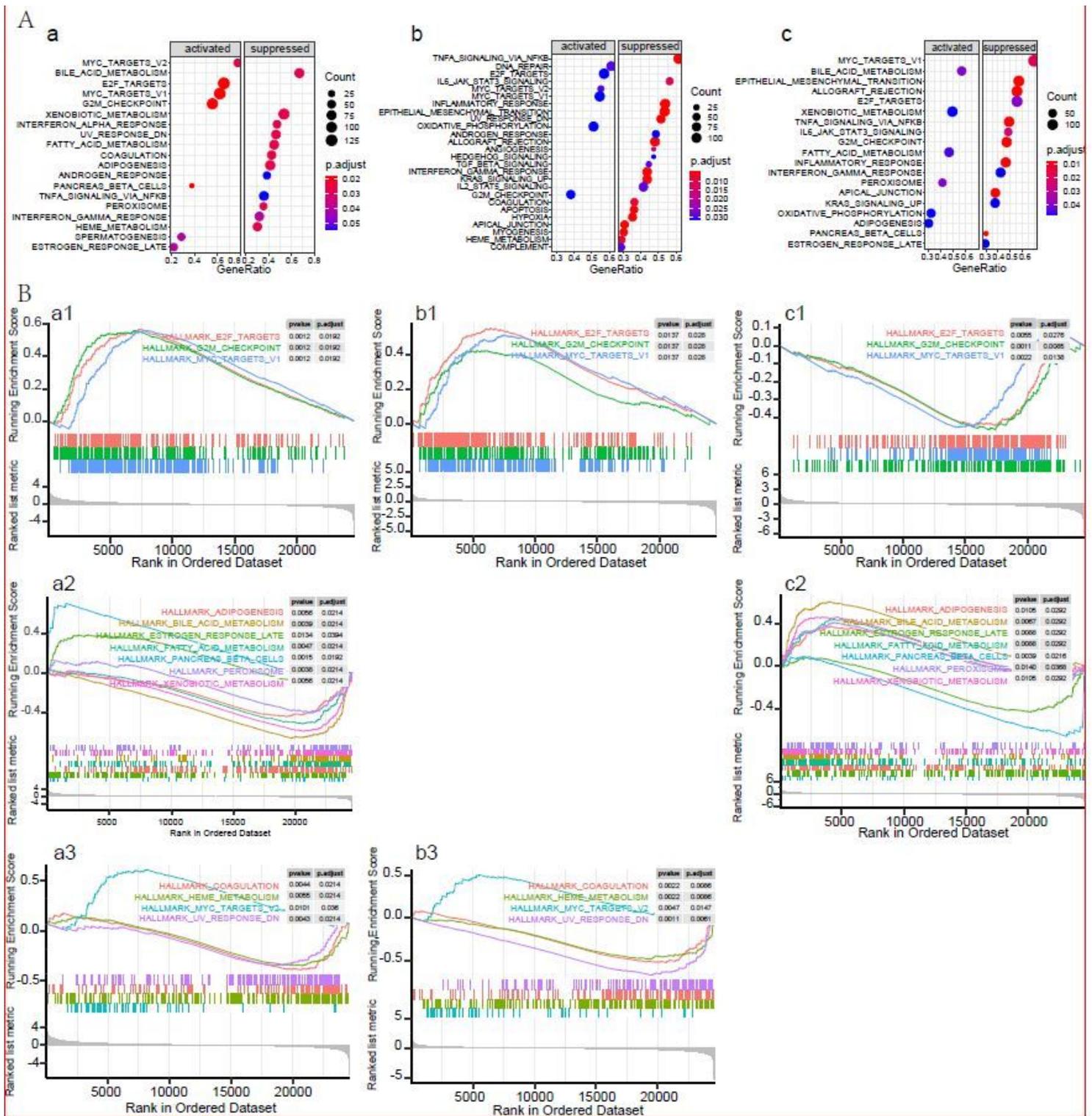


Figure 8

Single-gene GSEA analysis was performed on the three prognostic genes, ESR1, SPP1 and FOSB, that affect the survival time of liver cancer patients

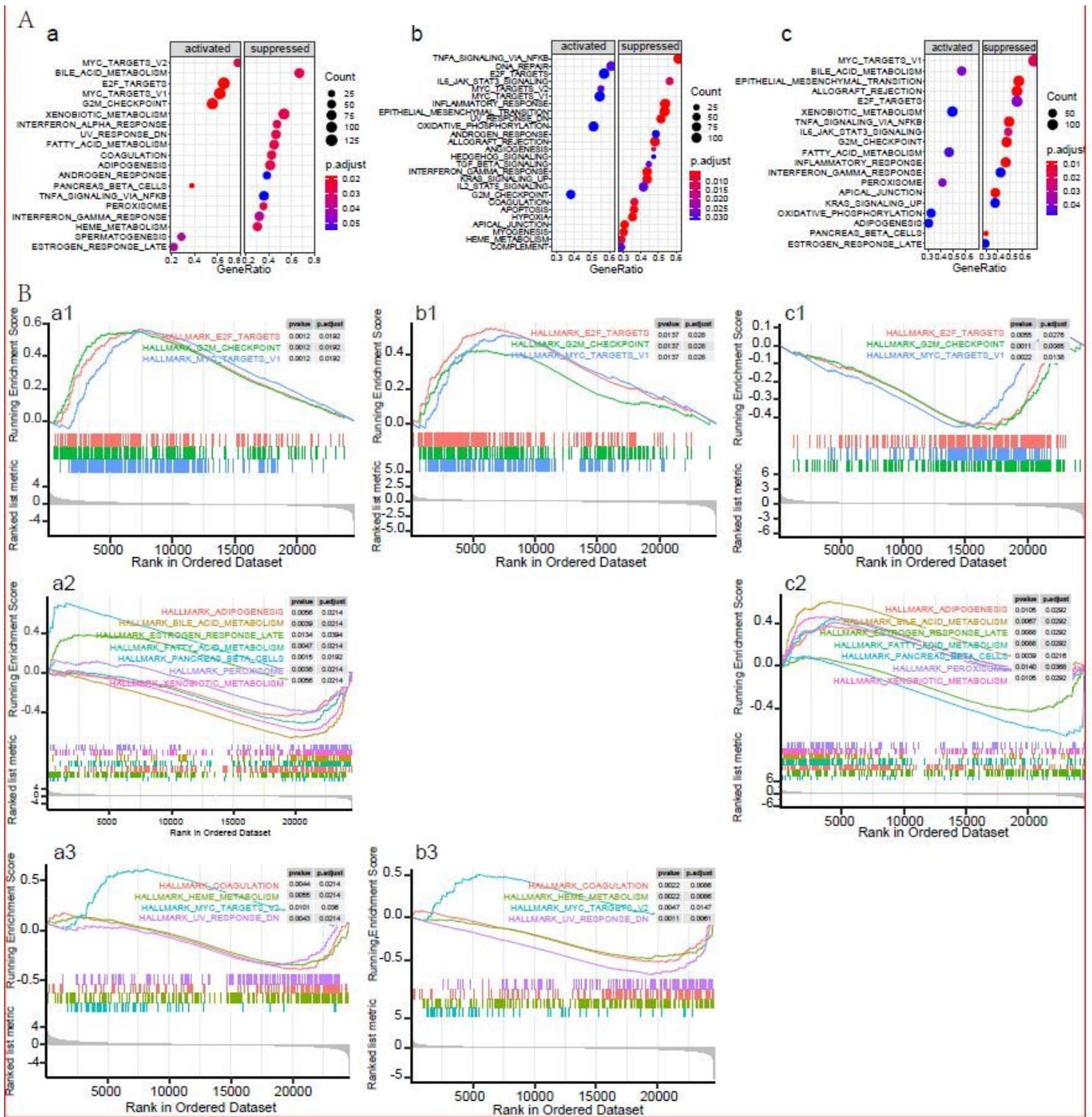


Figure 8

Single-gene GSEA analysis was performed on the three prognostic genes, ESR1, SPP1 and FOSB, that affect the survival time of liver cancer patients

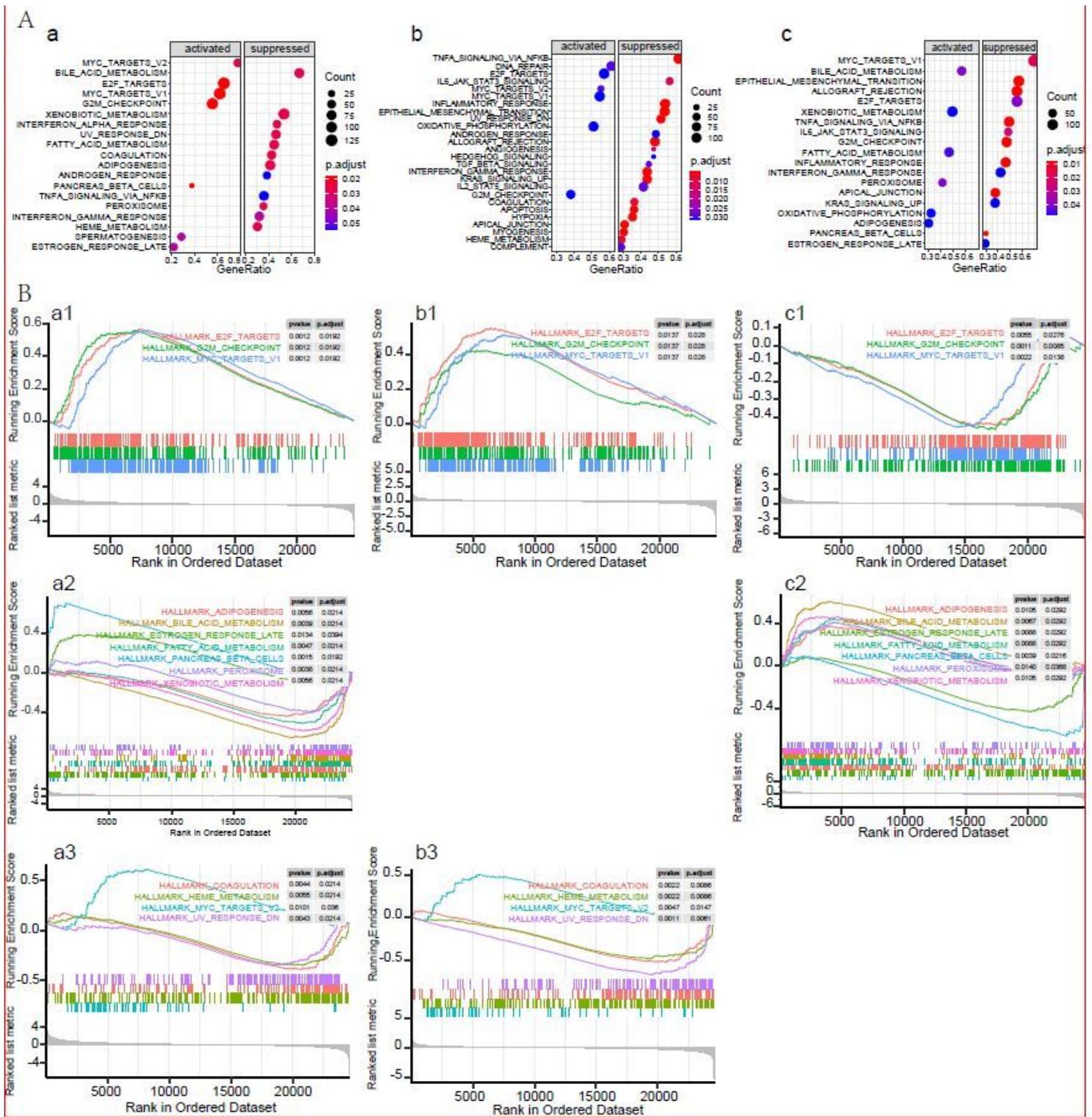


Figure 8

Single-gene GSEA analysis was performed on the three prognostic genes, ESR1, SPP1 and FOSB, that affect the survival time of liver cancer patients

Supplementary Files

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

- [Supplementtable3.xlsx](#)
- [Supplementtable3.xlsx](#)
- [Supplementtable3.xlsx](#)
- [Supplementtable2.xlsx](#)
- [Supplementtable2.xlsx](#)
- [Supplementtable2.xlsx](#)
- [Supplementtable1.xlsx](#)
- [Supplementtable1.xlsx](#)
- [Supplementtable1.xlsx](#)
- [Table1.xlsx](#)
- [Table1.xlsx](#)
- [Table1.xlsx](#)
- [Table2.xlsx](#)
- [Table2.xlsx](#)
- [Table2.xlsx](#)