Big Data-Based Identification of Multi-Gene Prognostic Signatures in Liver Cancer

Meiliang Liu¹, Xia Liu², Shun Liu¹, Feifei Xiao³, Erna Guo¹,⁴, Xiaoling Qin¹, Liuyu Wu¹, Qiuli Liang¹, Zerui Liang¹, Kehua Li¹, Di Zhang¹, Yu Yang¹, Xingxi Luo¹, Lei Lei¹, Jennifer Hui Juan Tan⁵, Fuqiang Yin⁶,⁷*, and Xiaoyun Zeng¹,⁷*

¹ School of Public Health, Guangxi Medical University, Nanning, China, ² Key Laboratory of Longevity and Ageing-Related Disease of Chinese Ministry of Education, Centre for Translational Medicine and School of Preclinical Medicine, Guangxi Medical University, Nanning, China, ³ Department of Epidemiology and Biostatistics, University of South Carolina, Columbia, SC, United States, ⁴ School of International Education, Guangxi Medical University, Nanning, China, ⁵ School of Life Sciences and Chemical Technology, Ngee Ann Polytechnic, Singapore, Singapore, ⁶ Life Sciences Institute, Guangxi Medical University, Nanning, China, ⁷ Key Laboratory of High-Incidence-Tumor Prevention and Treatment, Guangxi Medical University, Ministry of Education, Nanning, China

Simultaneous identification of multiple single genes and multi-gene prognostic signatures with higher efficacy in liver cancer has rarely been reported. Here, 1,173 genes potentially related to the liver cancer prognosis were mined with Coremine, and the gene expression and survival data in 370 samples for overall survival (OS) and 319 samples for disease-free survival (DFS) were retrieved from The Cancer Genome Atlas. Numerous survival analyses results revealed that 39 genes and 28 genes significantly associated with DFS and OS in liver cancer, including 18 and 12 novel genes that have not been systematically reported in relation to the liver cancer prognosis, respectively. Next, totally 9,139 three-gene combinations (including 816 constructed by 18 novel genes) for predicting DFS and 3,276 three-gene combinations (including 220 constructed by 12 novel genes) for predicting OS were constructed based on the above genes, and the top 15 of these four parts three-gene combinations were selected and shown. Moreover, a huge difference between high and low expression group of these three-gene combination was detected, with median survival difference of DFS up to 65.01 months, and of OS up to 83.57 months. The high or low expression group of these three-gene combinations can predict the longest prognosis of DFS and OS is 71.91 months and 102.66 months, and the shortest is 6.24 months and 13.96 months. Quantitative real-time polymerase chain reaction and immunohistochemistry reconfirmed that three genes F2, GOT2, and TRPV1 contained in one of the above combinations, are significantly dysregulated in liver cancer tissues, low expression of F2, GOT2, and TRPV1 is associated with poor prognosis in liver cancer. Overall, we discovered a few novel single genes and multi-gene combinations biomarkers that are closely related to the long-term prognosis of liver cancer, and they can be potential therapeutic targets for liver cancer.

Keywords: liver cancer, gene combinations, data mining, disease-free survival (DFS), overall survival (OS)
INTRODUCTION

Liver cancer is the sixth most common cancer and the fourth leading cause of cancer-related deaths (1). Specifically, hepatocellular carcinoma (HCC) accounts for more than 90% of liver cancer cases from a histopathological perspective. According to the GLOBOCAN 2018 database, there are about 841,000 new HCC cases and 782,000 related deaths worldwide each year, with China accounting for nearly half of the total number of global HCC cases and deaths (2, 3). In China, the Guangxi province has higher morbidity and mortality rates than the national average (4). The high mortality and poor prognosis of HCC poses a global challenge. Despite the slight increase in the 5-year survival rate of liver cancer in China from 10.1 to 12.1% over the periods of 2003–2015, it still remains at a low level (5). A survival analysis of 2,887 liver cancer patients in 14 years showed that the 1-year, 3-year, and 5-year survival rates were 49.3, 26.6, and 19.5%, respectively (6).

Although there are many existing therapies for HCC including surgical resection, transplantation, ablation, and transcatheter chemoembolization, etc., the long-term survival of HCC patients remains poor due to their limited indications and different effects on prognosis (7–10). A 20-year prospective cohort analysis reported that the 5-year survival rates of TNM stage I, II, IIIA, and IVA patients after hepatectomy were 81.7, 77.2, 44, and 28.2%, respectively (11). Therefore, it is of crucial importance to explore new prognostic biomarkers and investigate treatment strategies to improve the overall prognosis of HCC patients.

Currently, the research on prognostic molecular markers of HCC is still ongoing, and many single-gene or multi-gene combination molecular markers related to HCC invasion, metastasis and prognosis are being gradually discovered. For example, the expression of HMGA1 in HCC is associated with poor prognosis and is found to promote tumor growth and migration in vitro (12). The overexpression of SYPL1 is associated with epithelial-mesenchymal transition (EMT) of HCC cells and can predict the prognosis of HCC (13). RBM8A and SIRT5 promote the migration and invasion of HCC cells by activating the EMT signaling pathway and targeting E2F1 (14, 15), respectively (16, 17). The EpCAM (18), a liver X receptor (LXR) (19), SPAG5 (20), and KOR (21) have been shown to be strongly correlated with HCC metastasis, invasion, or prognosis. Arginase-1, FTCD, and MOC-31 have a good performance in the diagnosis of HCC (22). TMEM88, CCL14, and CLEC3B can serve as potential prognostic markers of HCC (23). At the same time, some multi-gene combined prognostic studies on HCC have also been reported. For example, three genes (UPB1, SOCS2, RTN3) combination markers (24) and four genes (CENPA, SPP1, MAGEB6, HOXD9) combination models can predict the overall survival in patients with HCC prognosis (25).

However, due to the sample size limitation and the heterogeneity of the samples in different studies, the efficiency of the identified prognostic markers for liver cancer still has ample space to improve. In addition, because of the myriad of gene interaction capabilities and the possibility of synergistic promotion of disease progression, it is of great significance to find some multi-gene combinations that may have better prognostic efficacy than single genes for prognostic targets of liver cancer. Therefore, the leverage of the large sample sizes of the public data platforms, integrating new and effective mining and screening methods, as well as reliable experimental verification is a very promising direction for the discovery of multiple effective single genes and multi-gene combination prognostic markers of liver cancer.

High-throughput profiling technologies and bioinformatics methods are now being applied to all fields of biomedical research. A mass of cancer data, such as the mRNA expression, copy number variation, single nucleotide polymorphism (SNP), and microRNA expression generated by those tools are collected in public archives such as The Cancer Genome Atlas (TCGA) (http://cancergenome.nih.gov/), Coremine (http://www.coremine.com/medical/), Oncomine (https://www.oncomine.org/resource/login.html), Gene Expression Omnibus database (GEO, https://www.ncbi.nlm.nih.gov/geo/), etc. Making full use of the public data from these databases is meaningful for exploring and discovering effective HCC prognostic biomarkers. For instance, Li et al. (24) developed a three-gene prognostic signature composing of three genes UPB1, SOCS2, and RTN3, which was revealed to have prognostic value for HCC patients based on TCGA data. Our previous study used data retrieved from the Coremine, TCGA, and GEO database and discovered that high-expressed E2F transcription factor 3 is associated with poor prognosis of HCC (26).

In this study, we used text mining approach to find the medial related candidate gene list for liver cancer prognosis, and a total of 1,173 genes that might be related to the prognosis of liver cancer were finally obtained. The association of the 1,173 genes with overall survival (OS) and disease-free survival (DFS) was accessed in a large sample of TCGA cohort, in which the subgroups of 319 patients with DFS and 370 with OS were available. The survival analyses are carried out for each of these genes to identify single prognostic markers. Moreover, we performed survival analyses of the gene combinations and performed multiple screening for these HCC prognostic molecular markers, revealing the association between the expression of numerous genes or gene combinations and the survival in HCC patients. We then compared the ability of single genes and multiple gene combinations to predict the prognosis of HCC. Moreover, a huge difference between high and low expression group of these three-gene combinations was detected, with median survival difference of DFS up to 65.01 months, and of OS up to 83.57 months. The high or low expression group of these three-gene combinations can predict the longest prognosis of DFS and OS is 71.91 months and 102.66 months, and of OS up to 83.57 months. The high or low expression group of these three-gene combinations can predict the longest prognosis of DFS and OS is 71.91 months and 102.66 months, and of OS up to 83.57 months. The high or low expression group of these three-gene combinations can predict the longest prognosis of DFS and OS is 71.91 months and 102.66 months, and of OS up to 83.57 months. The high or low expression group of these three-gene combinations can predict the longest prognosis of DFS and OS is 71.91 months and 102.66 months, and of OS up to 83.57 months. The high or low expression group of these three-gene combinations can predict the longest prognosis of DFS and OS is 71.91 months and 102.66 months, and of OS up to 83.57 months. The high or low expression group of these three-gene combinations can predict the longest
(IHC) staining in 90 pairs of HCC and its precancerous tissues. These results validated that the low expression of F2, GOT2, and TRPV1 in liver cancer was associated with the poor prognosis of liver cancer.

**MATERIALS AND METHODS**

**Data Sources**

We combined 3 corresponding concepts of the key word “liver cancer” with 2 concepts of the key word “prognosis” and 10 concepts of the key word “outcome,” respectively, (Supplementary Table S1), and searched for their corresponding genes or proteins in the Coremine database (http://www.coremine.com/medical/). After deleting duplicates, we selected 1,173 gene entries with p-values < 0.05 that might be associated with the prognosis of liver cancer for further analyses (Supplementary Table S2).

The above genes mined in the Coremine database include some genes obtained from other gene-mining reports; however, the number of samples and data standards in each report is different. Therefore, we selected the cohort of The Cancer Genome Atlas (TCGA) (http://cancergenome.nih.gov/), a database with consistent sample size and data standards, to conduct unified batch verification of these genes and conduct three-gene combinations survival analyses.

We studied the relationship between each of the selected 1,173 genes and the prognosis of liver cancer patients in TCGA cohort which downloaded from cBioPortal for Cancer Genomics (https://www.cbioportal.org/) in September 2018 (27, 28), and a subgroup of 319 liver cancer samples with HCC DFS corresponding follow-up data and a subgroup of 370 liver cancer samples with HCC OS corresponding follow-up data were chosen.

**Survival Analysis and Gene Selection**

Kaplan-Meier estimation of survival functions and Log-rank tests were used to evaluate effect of genes on DFS and OS. The Cox proportional hazard model was performed for multivariate analyses of HCC prognosis. Survival analyses were performed using the R survival package in R (version 3.3.1). The Kaplan-Meier survival curves and Cox proportional hazards regression model for DFS and OS were generated by IBM SPSS (version 23.0). The median expression level of a gene was used as a cutoff value for the classification of patients into high and low expression groups (29).

**Human Tissue Samples**

For the validation studies, we used 20 patients who underwent primary and curative hepatectomy from Apr 2016 to Apr 2018 at the First Affiliated Hospital of Guangxi Medical University. Those patients who have distinctive pathologic diagnosis of HCC without preoperative anticancer treatment were eligible for inclusion in this study. The paraffin-embedded pathologic specimens were collected during surgery and stored in a liquid nitrogen tank until the step of mRNA isolation. All patients received an explanation for the purpose of the study and signed informed consent. The Ethics Committee of Guangxi Medical University granted approval for this study. For IHC, a commercial biological tissue microarray containing 90 pairs of HCC and adjacent normal liver tissues was constructed by the Biological sample library of Shanghai Outdo Biotech Company, and the survival information of each case was usable. (Microarray: HLivH180Su14).

**Quantitative Real-Time Polymerase Chain Reaction (qRT-PCR)**

QRT-PCR was performed to evaluate the mRNA expression of selected genes in 20 HCC and their matched precancerous tissues. Total RNA was isolated using Trizol reagent (Life Technologies, Inc., NY, USA) according to the manufacturer’s instructions. The concentration and purity of the total RNA were detected using Microplate reader (Biotek Instruments, Inc., VT, USA). RNA reverse transcription was then performed with the PrimeScript™ RT reagent Kit (Takara Biomedical Technology (Beijing) Co., Ltd.) with gDNA Eraser (Perfect Real Time), and qRT-PCR was performed using the TB Green™ Premix Ex Taq™ II (Tli RNaseH Plus) kit (Takara Biomedical Technology (Beijing) Co., Ltd.) protocol in a StepOnePlus system (Applied Biosystems. Life Technologies Holdings Pte Ltd, Singapore).

The sequences of the primers are as follows: F2: forward primer, 5′-CTGAGGCGTCTGGTACGAACT-3′, reverse primer, 5′-TGGTACGACTCTCCTCATA-3′; GOT2: forward primer, 5′-AAGAGTGCCGTTGTTGTCAC-3′, reverse primer, 5′-AGAAAGACATCTCGGCTGAAT-3′; TRPV1: forward primer, 5′-TGCAGCGGACGAAAC-3′, reverse primer, 5′-GGTTGCAAGGTCTGTCG-3′. The cycle conditions are as follows: after an initial incubation at 95°C for 30 s, the samples were cycled 40 times at 95°C for 5 s and 60°C for 30 s. The relative expression level of each gene in the individual samples was calculated using the 2−ΔΔCt method and normalized using GAPDH as an endogenous control.

**Immunohistochemistry (IHC)**

EnVision™ FLEX+, Mouse, High pH, (Link) (K8002, Dako) was used for the immunohistochemistry. After the tissue chips were baked and placed in LEICAST5010 (LEICA), PT Link (Dako North America, Inc.) was used for antigen retrieval. Primary antibodies were diluted (F2, 1:3000; GOT2, 1:80000; TRPV1, 1:1500) and incubated overnight at 4°C. The secondary antibody reactions were carried out using the Autostainer Link 48 (Dako North America, Inc.), the sections were subjected to color development with the DAB chromogenic kit, and finally counterstained with Hematoxylin (SLBT4555, Sigma Aldrich).

The following antibodies were used: F2, 1: Anti-Thrombin (ab83981; Abcam), GOT2, 1: Anti-FABP-1 (ab171739; Abcam), TRPV1, 1: Anti-VR1 (ab3487; Abcam). All slides were evaluated by two independent pathologists who were blind about the clinicopathologic data.

The expression levels were scored as the staining intensity (0, negative; 1+, weak; 2+, moderate; 3+, strong) multiplied by the proportion of immunopositive staining area (0, < 25%; 1+, 25–50%; 2+, 50–75%; 3+, >75%) intensity of staining. Expression scores <5 were considered as “low expression,” and scores ≥5 were considered as “high expression.”
Selecting Genes Related to Liver Cancer Prognosis and Liver Cancer Samples

We combined 3 corresponding concepts of the key word “liver cancer” [Liver Neoplasms (alias Liver Cancer) (disease) (60,666 connections); Liver carcinoma (alias liver cell cancer) (disease) (55,739 connections); Carcinoma, Hepatocellular (alias Adult Liver Cancer) (mesh) (57,034 connections)] with 2 corresponding concepts of the key word “prognosis” [Prognosis (mesh) (77,312 connections); Prognostic Marker (alias Prognosis Marker) (chemical) (22,056 connections)] and 10 corresponding concepts of the key word “outcome” [Fatal Outcome (mesh) (34,016 connections); Outcome Assessment (Health Care) (alias Outcome Study) (mesh) (48,296 connections); Outcome studies (procedure) (9,545 connections); Treatment Outcome (mesh) (77,246 connections); Outcomes research (procedure) (5,540 connections); Outcome monitoring (procedure) (2,030 connections); Patient-focused outcomes (procedure) (3,830 connections); Treatment outcome in HSR (procedure) (998 connections); Patient Reported Outcome Measures (alias Patient Reported Outcome) (mesh) (48,296 connections); Patient Reported Outcome Assessment (mesh) (9,066 connections)], respectively, (Supplementary Table S1), and searched for their corresponding genes or proteins in the Coremine database (http://www.coremine.com/medical/). With p-values < 0.05 as the criteria, a total of 1,173 genes that might be related to the prognosis of liver cancer were finally obtained after screening and elimination of duplicates. As the samples of liver cancer in the Coremine database were not uniform enough, we selected 319 samples for DFS and 370 samples for OS of liver cancer from the TCGA database and obtained the corresponding survival data as well as the expression information of the above 1,173 genes in these samples. This was necessary to carry out the subsequent survival analyses of these genes for liver cancer.

The Single Genes Prognostic Analyses

To clearly describe our process of screening genes, a flowchart of the analysis procedure was developed (Figure 1). First, we performed the Kaplan-Meier analysis of each of the 1,173 genes. It was found that the mRNA expression of 276 genes and 283 genes was significantly associated with DFS in 319 patients (p < 0.05) and OS in 370 patients (p < 0.05), respectively. Additionally, the mRNA expression of 166 of these genes was significantly associated with both DFS and OS (p < 0.05).

To further investigate the value of the genes in the prognosis of liver cancer, we chose 135 genes and 149 genes with p-values < 0.01 for DFS and OS, respectively. Next, we used the Cox proportional hazards regression model to employ multivariate analyses on the above genes, respectively to determine the DFS and OS prediction potential of these genes.

The DFS-related multivariate analysis results showed that the expression of 39 genes (ALDOB, APOB, AURKB, C5, CCNE, CD4, CNNP1, CETP, COL1A1, CPT2, DAND5, DNAS1, EBPL, F7, FLI1, G6PD, GMNT, ITGB2, KLRK1, KNG1, LMOD1, NEK2, PCLAF, PER1, PKM, POU2F1, PPAT, PPIA, PRF1, PTTPN6, RUNX3, SELP, SLCO1B1, SPPL2A, STAT5A, TCF21, TRPV1, TUSC1, and TYMS) was significantly associated with DFS in HCC patients (p < 0.05, Table 1). The highly significant results of both the DFS-related single-gene survival analyses for each of these 39 genes and the multivariate analysis confirmed that the above 39 genes have a strong association with the DFS of liver cancer, especially the 5-year disease free survival rate of liver cancer.

The OS-related multivariate analysis results showed that the expression of 28 genes (ABCC1, ANXA7, APOB, ATG7, BAK1, CA9, CCNA2, CHD1L, CYP3A4, E2F1, EZH2, F2, G6PC, GMP3, GOT2, HDAC2, HPX, KPN2A, LAPT4M, MAGEB3, MAPT, MPV17, NTF3, PPAT, SLCA21, SLC38A1, SPPL2A, and TRPV1) was significantly associated with OS in HCC patients (p < 0.05, Table 1). The strongly significant results of both the OS-related single-gene survival analyses and multivariate analysis confirmed that these 28 genes are significantly associated with the OS of liver cancer, especially the 5-year survival rate of liver cancer.

Additionally, among the above-mentioned genes selected after single-gene survival analyses and multivariate analyses, 3 genes (APOB, PPAT, and TRPV1) were significantly associated with both DFS and OS in HCC patients.

Heat maps of the expression of the above 39 DFS-related genes and 28 OS-related genes in 1173 TCGA liver cancer samples, respectively, which grouped by prognosis status, were shown in Supplementary Figure S1.

Three-Gene-Combination Prognostic Model

To reflect the association of the expression of the combined genes with the prognosis of HCC, three-gene-combinations of the above 39 and 28 single genes that are significantly associated with DFS and OS, respectively, were formed, resulting in 9,139 and 3,276 three-gene-combinations for DFS and OS, respectively. In each combination, simultaneous high expression of the three genes in the same case was considered as the co-high expression, low expression of the three genes in the same case was considered as the co-low expression, and the low and high expression were deleted combinations which had < 25 cases in the co-high or co-low expression group.

Three-Gene-Combination of Prediction for DFS in Liver Cancer

K-M survival analysis of each of the above 9,139 combinations constituted by 39 DFS-related single genes was first performed. Then, we selected a total of 2,758 combinations with p-values < 0.01 for DFS and OS, respectively, excluding the combinations with no more than 25 cases in the co-high expression or co-low expression groups. Apparently,
these selected 2,758 combinations have significant prognostic implications for DFS in liver cancer.

In addition, 18 of the above 39 single genes have not yet been systematically reported to be associated with HCC prognosis, and these 18 genes can combine into 816 three-gene-combinations. The results of the K-M survival analyses showed that 317 combinations had significant association with DFS of liver cancer ($p < 0.01$).

The top 15 combinations of the above 2,758 and 317 combinations with the smallest $p$-values were chosen. The DFS-related survival analyses diagrams and tables of these combinations and the single genes they contain are as follows (Figures 2, 3; Tables 2, 3).

Three-Gene-Combination of Prediction for OS in Liver Cancer

Similarly, three-gene-combinations of the 28 single genes significantly associated with OS confirmed by the single gene survival analyses and the multivariate analysis were formed, resulting in 3,276 three-gene-combinations. 930 of these 3,276 combinations were screened out on the conditions that the number of cases in both the co-high and co-low expression groups was > 25, and the $p$-values were < 0.01 according to the OS-related K-M analyses results.

Furthermore, 12 of the above 28 single genes that were noted to have an unknown association with liver cancer prognosis formed 220 three-gene-combinations. Out of the 220 combinations, there were 31 combinations in which the number of cases in both the co-high and co-low expression groups was > 25 and the OS-related survival analyses results showed $p < 0.01$.

We found 930 of above 3,276 combinations and 31 of above 220 unreported-gene combinations were significant association with OS related survival of liver cancer patients. Among the 930 combinations and 31 combinations mentioned above, the diagrams and tables of the OS-related survival analyses of the top 15 combinations with the smallest $p$-values and the single genes they contain are as follows (Figures 4, 5; Tables 3, 4). Among the 12 genes that have an unknown association with HCC prognosis, $F2$, $GOT2$, $TRPV1$, and their combination $F2$-$GOT2$-$TRPV1$ were all significantly associated.
TABLE 1 | Multivariate analyses of prognosis of DFS of 319 HCC patients and OS of 370 HCC patients in a TCGA cohort.

| Items         | Genes | B    | SE   | Wald | Sig. | Exp (B) | 95.0% CI     |
|---------------|-------|------|------|------|------|---------|--------------|
|               |       | Lower| Upper|      |      |         |              |
| DFS associated| ALDOB | −0.580| 0.186 | 9.750 | 0.002 | 0.560 | 0.389 0.806 |
|               | APOB  | −0.436| 0.217 | 4.023 | 0.045 | 0.647 | 0.423 0.990 |
|               | AURKB | 0.527 | 0.211 | 6.208 | 0.013 | 1.694 | 1.119 2.564 |
|               | CS*   | −0.420| 0.170 | 6.093 | 0.014 | 0.657 | 0.471 0.917 |
|               | CCNF  | 0.694 | 0.334 | 4.310 | 0.038 | 2.002 | 1.040 3.857 |
|               | CD4*  | −0.774| 0.316 | 6.007 | 0.014 | 0.481 | 0.248 0.856 |
|               | CEP15 | 1.053 | 0.243 | 18.794| 0.000 | 2.867 | 1.781 4.615 |
|               | CETP* | 0.829 | 0.423 | 3.851 | 0.050 | 2.291 | 1.001 5.245 |
|               | COL18A1* | 0.417 | 0.170 | 6.093 | 0.014 | 0.657 | 0.471 0.917 |
|               | CPT2  | 0.558 | 0.247 | 4.514 | 0.033 | 1.672 | 1.056 2.632 |
|               | DAND5* | −0.427| 0.183 | 4.626 | 0.034 | 0.652 | 0.423 0.990 |
|               | DNASE1* | 0.382 | 0.136 | 7.927 | 0.005 | 1.465 | 1.123 1.910 |
|               | EBPL* | −0.766| 0.280 | 7.463 | 0.006 | 0.465 | 0.268 0.805 |
|               | F7*   | −0.496| 0.175 | 8.034 | 0.005 | 0.469 | 0.243 0.858 |
|               | FLT3* | −0.700| 0.240 | 8.512 | 0.004 | 0.497 | 0.261 0.905 |
|               | G6PD  | 0.477 | 0.188 | 6.438 | 0.011 | 1.611 | 1.115 2.328 |
|               | GNMT  | 0.427 | 0.160 | 7.118 | 0.008 | 1.533 | 1.120 2.097 |
|               | ITGB2* | 1.112 | 0.301 | 13.662| 0.000 | 3.042 | 1.686 5.486 |
|               | KLRK1 | 0.932 | 0.384 | 5.883 | 0.015 | 2.539 | 1.196 5.390 |
|               | KNG1* | 0.645 | 0.277 | 5.412 | 0.020 | 1.906 | 1.107 3.282 |
|               | LMOD1* | −0.873| 0.410 | 4.524 | 0.033 | 0.418 | 0.187 0.934 |
|               | NCK2  | −0.546| 0.263 | 4.299 | 0.038 | 0.579 | 0.346 0.971 |
|               | PCLAF | 0.526 | 0.243 | 4.700 | 0.030 | 1.693 | 1.052 2.724 |
|               | PER1  | −0.670| 0.221 | 9.169 | 0.002 | 0.512 | 0.332 0.790 |
|               | PKM   | −0.645| 0.282 | 5.210 | 0.022 | 0.525 | 0.302 0.913 |
|               | POU2F1 | 0.455 | 0.142 | 10.236| 0.001 | 1.577 | 1.193 2.084 |
|               | PPAT1* | 0.968 | 0.310 | 13.662| 0.000 | 3.042 | 1.686 5.486 |
|               | PPAT4 | 0.626 | 0.183 | 11.661| 0.001 | 1.870 | 1.306 2.679 |
|               | PRF1* | −1.676| 0.370 | 20.505| 0.000 | 0.187 | 0.091 0.386 |
|               | PTEN6 | −0.610| 0.227 | 7.203 | 0.007 | 0.543 | 0.348 0.848 |
|               | RUNX3 | 0.967 | 0.375 | 6.659 | 0.010 | 2.629 | 1.262 5.479 |
|               | SELP  | 0.790 | 0.270 | 8.587 | 0.003 | 2.203 | 1.299 3.736 |
|               | SLCO1B1 | −0.524| 0.213 | 6.029 | 0.014 | 0.592 | 0.390 0.900 |
|               | SPP1LA4* | −0.669| 0.217 | 9.528 | 0.002 | 0.512 | 0.335 0.783 |
|               | STAT5A | −1.704| 0.489 | 12.149| 0.000 | 0.182 | 0.077 0.474 |
|               | TCF21* | −0.979| 0.401 | 9.561 | 0.015 | 0.376 | 0.171 0.824 |
|               | TRPV1* | −0.520| 0.189 | 7.604 | 0.006 | 0.595 | 0.411 0.860 |
|               | TUSC1 | 0.423 | 0.188 | 5.044 | 0.025 | 1.526 | 1.055 2.077 |
|               | TWSM  | 0.523 | 0.245 | 4.588 | 0.033 | 1.687 | 1.044 2.727 |
| OS associated  | ABCC1 | 1.097 | 0.369 | 8.841 | 0.003 | 2.994 | 1.453 6.168 |
|               | AKAP7* | −0.554| 0.201 | 7.618 | 0.006 | 0.575 | 0.388 0.852 |
|               | APOB  | −0.791| 0.311 | 6.461 | 0.011 | 0.453 | 0.246 0.834 |
|               | ATG7  | 0.613 | 0.312 | 3.876 | 0.049 | 1.847 | 1.003 3.400 |
|               | BAK1  | −0.490| 0.231 | 4.497 | 0.034 | 0.613 | 0.390 0.964 |
|               | CA9   | 0.761 | 0.363 | 4.399 | 0.036 | 2.140 | 1.051 4.356 |
|               | CCNA2 | 0.502 | 0.203 | 6.094 | 0.014 | 1.652 | 1.109 2.461 |
|               | CDH1L | 0.491 | 0.181 | 7.377 | 0.007 | 1.634 | 1.147 2.330 |
|               | CYP3A4 | 0.999 | 0.364 | 7.539 | 0.006 | 2.717 | 1.331 5.544 |
|               | E2F1  | 0.360 | 0.172 | 4.371 | 0.037 | 1.433 | 1.023 2.008 |

(Continued)
with OS in 370 liver cancer samples from the TCGA data (F2: \( p = 0.005 \); GOT2: \( p < 0.001 \); TRPV1: \( p = 0.002 \); F2-GOT2-TRPV1: \( p < 0.001 \)). The overall survival rate in HCC patients with low expression of F2, GOT2, TRPV1, and the three-gene-combination F2-GOT2-TRPV1 were all significantly lower than that in liver cancer patients with high expression. In addition, the median survival time difference between the high expression group and the low expression group of F2, GOT2, TRPV1, and the three-gene combination F2-GOT2-TRPV1 was 23.62, 32.26, 35.61, and 55.68 months, respectively. The median survival time difference of this combination was greater than that of a single gene, which was one of the main reasons why we selected these three genes for qRT-PCR and immunohistochemically validation.

**Low Expression of F2, GOT2, and TRPV1 Predicts Poor Prognosis**

Based on the above results of the OS-related survival analyses and multivariate analyses on 28 genes, as well as the results of survival analyses on their three-gene-combinations, we selected three genes F2, GOT2, and TRPV1 with strong liver cancer prognostic potential for subsequent validation.

**F2, GOT2, and TRPV1 Were Downregulated in HCC Tissues**

The gene expression in HCC was determined based on three independent microarrays which are all collected in Oncomine database (https://www.oncomine.org/resource/login. html). As shown in Roessler Liver 2 Statistics (225 HCC tissues vs. 220 liver tissues), the expression of F2, GOT2, and TRPV1 in HCC tissues were all significantly down-regulated compared with that in normal liver tissues. (\( p < 0.001 \);

**Figure 6** In addition, based on the Mas Liver Statistics (38 HCC tissue vs. 19 liver tissue), both F2 and TRPV1 were significantly down-regulated in HCC tissues. Based on the Chen Liver Statistics (104 HCC tissues vs. 76 liver tissues), both F2 and GOT2 were significantly down-regulated in HCC tissues.

The qRT-PCR results of F2, GOT2 and TRPV1 showed that 20/20, 19/20, and 16/19 of the HCC tissues exhibited significantly lower expression of F2 (\( p < 0.001 \); Figure 7A), GOT2 (\( p < 0.001 \); Figure 7B), and TRPV1 (\( p = 0.006 \); Figure 7C), respectively, when compared with their corresponding non-tumor tissues.

The protein expression of F2, GOT2, and TRPV1 in HCC tissues was evaluated using IHC. Positive staining of F2, GOT2, and TRPV1 was mainly localized in the cytoplasm of HCC cells. The representative staining of F2, GOT2, and TRPV1 negative and positive protein expression in HCC are shown in Figure 8A.

Among 90 HCC tissues and adjacent non-malignant liver tissues, IHC was employed to measure the protein expression of F2, GOT2, and TRPV1, respectively. Low F2 expression was observed in 62/89 (69.66%) of the HCC tissues, compared to 33/89 (37.08%) in adjacent normal liver tissues (\( p < 0.001 \)); low GOT2 expression was noted in 72/89 (80.90%) of the HCC tissues, compared to 32/89 (35.96%) in adjacent normal liver tissues (\( p < 0.001 \)); low TRPV1 expression was also observed in 59/89 (66.29%) of the HCC tissues, compared to 38/89 (42.70%) in adjacent normal liver tissues (\( p = 0.002 \)).

### Table 1 (Continued)

| Items | Genes | B   | SE  | Wald | Sig.  | Exp (B) | 95.0% CI          |
|-------|-------|-----|-----|------|-------|---------|--------------------|
|       | EZH2  | 0.985 | 0.399 | 6.103 | 0.013 | 2.678   | 1.226 - 5.852      |
|       | F2*   | 0.711 | 0.313 | 5.174 | 0.023 | 2.036   | 1.103 - 3.757      |
|       | G6PC  | -0.677 | 0.341 | 3.937 | 0.047 | 0.508   | 0.260 - 0.992      |
|       | GMP5  | 0.733 | 0.291 | 6.345 | 0.012 | 2.081   | 1.177 - 3.681      |
|       | GOT2* | -1.509 | 0.484 | 9.723 | 0.002 | 0.221   | 0.086 - 0.571      |
|       | HDAC2 | 0.813 | 0.316 | 6.628 | 0.010 | 2.255   | 1.214 - 4.187      |
|       | HPX*  | 0.930 | 0.384 | 5.882 | 0.015 | 2.535   | 1.195 - 5.378      |
|       | KPN2  | 0.835 | 0.284 | 8.664 | 0.003 | 2.305   | 1.322 - 4.018      |
|       | LAPTM4B | -0.492 | 0.168 | 6.161 | 0.003 | 0.611   | 0.440 - 0.849      |
|       | MAGEB3*| 0.393 | 0.179 | 4.824 | 0.028 | 1.482   | 1.043 - 2.105      |
|       | MAPT* | 0.660 | 0.243 | 7.349 | 0.007 | 1.934   | 1.201 - 3.117      |
|       | MPV17*| 1.141 | 0.488 | 5.468 | 0.019 | 3.129   | 1.203 - 8.141      |
|       | NTF3* | 1.089 | 0.357 | 9.318 | 0.002 | 2.973   | 1.477 - 5.983      |
|       | PMP7* | 0.752 | 0.286 | 6.897 | 0.009 | 2.122   | 1.210 - 3.719      |
|       | SLC2A1*| -0.921 | 0.440 | 4.383 | 0.036 | 0.398   | 0.168 - 0.843      |
|       | SLC38A1*| -0.768 | 0.289 | 7.063 | 0.008 | 0.464   | 0.263 - 0.817      |
|       | SPP1  | 0.604 | 0.264 | 5.219 | 0.022 | 1.830   | 1.090 - 3.073      |
|       | TRPV1*| 0.453 | 0.201 | 5.044 | 0.025 | 1.572   | 1.059 - 2.334      |

*The gene has not been systematically reported to be associated with HCC prognosis.

Cox proportional hazard model was used to analyze the impact of 135 genes on DFS and the impact of 149 genes on OS, respectively. \( P < 0.05 \) were considered to be significant. 39 genes and 28 genes were significantly associated with liver cancer DFS and OS, respectively.
FIGURE 2 | Association of DFS and the top 15 three-gene-combinations with smallest p-values, using the data of HCC samples in a TCGA cohort and assessed by Kaplan-Meier analyses. The high expression group (blue line) of the combination consisted of samples with high expression of all three genes, and the low expression group (green line) of the combination consisted of samples with low expression of all three genes. The number of high and low expression groups in each combination was >25. (A) Association of DFS and the top 15 combinations of the overall genes combinations. (B) Association of DFS and the top 15 combinations of the unreported genes combinations.

Expression of F2, GOT2, and TRPV1 and Their Combination F2-GOT2-TRPV1 With OS

Based on the above results of single-genes and three-gene combinations survival analyses of TCGA HCC samples, the low expression of F2, GOT2, TRPV1 and their combination F2-GOT2-TRPV1 was significantly associated with poor OS in HCC. (F2: $p = 0.005$; GOT2: $p < 0.001$; TRPV1: $p = 0.002$; F2-GOT2-TRPV1: $p < 0.001$). In addition, the median survival time difference between the high expression group and the low expression group of
FIGURE 3 | Association of DFS and the individual genes contained in the top 15 combinations with the lowest $P$-values, using the data of HCC samples in a TCGA cohort and assessed by Kaplan-Meier analyses. (A) Association of DFS and the 17 single genes contained in the first 15 total-gene combinations. (B) Association of DFS and the 16 single genes contained in the first 15 unreported-gene combinations.
The associations of three-gene combinations with disease-free survival (DFS) of HCC patients in a TCGA cohort, analyzed by Kaplan-Meier method.

| Combination | Estimate | Std. Error | 95% Confidence Interval | P | Median survival time difference (H-L) |
|-------------|----------|------------|-------------------------|---|-------------------------------------|
| DNASE1-PPIA-TUSC1 | H 9.490 1.597 6.360 12.620 | 0.000 | -62.420 | CD4-F7-TRPV1 | H 71.910 20.619 31.498 112.322 | 0.000 | 65.010 |
| | L 71.910 21.620 24.154 119.666 | Overall 21.620 4.848 12.119 31.121 | | | |
| | CD4-F7-TPRV1 | H 71.910 20.619 31.498 112.322 | 0.000 | 65.010 | | CD4-F7-LMOD1 | H 70.070 6.900 1.657 3.652 10.148 | 0.000 | 63.830 |
| | L 6.900 1.657 3.652 10.148 | Overall 15.740 5.309 5.334 26.146 | | | |
| | CD4-F7-GNMT | H 71.910 22.303 28.196 115.624 | 0.000 | 63.370 | CD4-COL18A1-F7 | H 67.580 21.110 26.205 108.955 | 0.000 | 59.660 |
| | L 8.540 1.241 6.108 10.972 | Overall 21.160 4.039 13.244 29.076 | | | |
| | CD4-F7-LMOD1 | H 70.070 62.420 | 0.000 | 65.010 | CD4-FLT3-SPPL2A | H 70.070 18.005 34.779 105.361 | 0.000 | 62.220 |
| | L 6.240 1.408 3.480 9.000 | Overall 19.650 3.833 10.127 25.153 | | | |
| | CD4-COL18A1-F7 | H 67.580 21.110 26.205 108.955 | 0.000 | 59.660 | CS-CD4-F7 | L 7.920 7.850 1.658 4.937 10.763 | 0.000 | 9.937 |
| | L 6.240 1.408 3.480 9.000 | Overall 17.640 3.833 10.127 25.153 | | | |
| | CD4-CPT2-F7 | H 71.910 22.303 28.196 115.624 | 0.000 | 63.370 | CD4-F7-SEL P | L 7.920 7.920 1.658 4.937 10.763 | 0.000 | 9.937 |
| | L 8.540 1.241 6.108 10.972 | Overall 21.160 4.039 13.244 29.076 | | | |
| | APOB-CD4-SLCO1B1 | H 66.620 13.239 40.672 92.568 | 0.000 | 57.130 | CD4-F7-SEL P | L 7.920 8.710 1.658 4.937 10.763 | 0.000 | 9.937 |
| | L 9.490 0.918 7.691 11.289 | Overall 19.650 3.833 10.127 25.153 | | | |
| | CD4-SELP | H 9.490 1.597 6.360 12.620 | 0.000 | 64.060 | CD4-F7-PRF1 | L 7.920 8.710 1.658 4.937 10.763 | 0.000 | 9.937 |
| | L 7.850 2.024 3.883 11.817 | Overall 19.650 3.833 10.127 25.153 | | | |
| | CD4-F7-PER1 | H 71.910 22.303 28.196 115.624 | 0.000 | 63.370 | CD4-SELP-SPPL2A | L 7.850 2.024 3.883 11.817 | 0.000 | 61.500 |
| | L 8.570 1.225 6.170 10.970 | Overall 19.650 3.833 10.127 25.153 | | | |
| | APOB-CD4-SPPL2A | H 70.070 23.928 23.171 116.969 | 0.000 | 60.940 | CD4-F7-FL3 | L 7.920 8.710 1.658 4.937 10.763 | 0.000 | 9.937 |
| | L 9.130 0.855 7.455 10.805 | Overall 19.650 3.833 10.127 25.153 | | | | (Continued)
| Table 2 | Continued |
|---|---|
| **DFS (Median) of combinations of 39 genes with HCC prognosis** |
| **Estimate** | **Std. Error** | **95% confidence interval** | **P** | **Median survival time difference (H-L)** |
| **Lower boundary** | **Upper boundary** | **Lower boundary** | **Upper boundary** |
| Overall | 19.190 | 4.689 | 10.000 | 28.380 |
| **CD4-FLT3-SPPL2A** |  |
| H | 70.070 | 18.005 | 34.779 | 105.361 | 0.000 | 62.220 |
| L | 7.850 | 1.486 | 4.937 | 10.763 |
| Overall | 19.650 | 7.275 | 5.391 | 33.909 |
| **DNASE1-PPAT-TUSC1** |  |
| H | 7.420 | 1.115 | 5.235 | 9.605 | 0.000 | −39.620 |
| L | 47.040 | 17.350 | 13.035 | 81.045 |
| Overall | 21.160 | 5.101 | 11.162 | 31.158 |
| **APOB-CD4-F7** |  |
| H | 67.580 | 13.500 | 41.120 | 94.040 | 0.000 | 58.840 |
| L | 8.740 | 0.884 | 7.007 | 10.473 |
| Overall | 24.770 | 9.057 | 7.018 | 42.522 |
| **CD4-SLC01B1-SPPL2A** |  |
| H | 71.910 | – | – | – | 0.000 | 62.420 |
| L | 9.490 | 1.171 | 7.194 | 11.786 |
| Overall | 19.650 | 6.519 | 6.873 | 32.427 |
| **C5-CD4-F7** |  |
| H | 67.580 | 15.374 | 37.447 | 97.713 | 0.000 | 59.660 |
| L | 7.920 | 1.414 | 5.149 | 10.691 |
| Overall | 21.160 | 5.704 | 9.981 | 32.337 |
| **CD4-GNMT-LMOD1** |  |
| H | 50.030 | 14.883 | 37.450 | 95.790 | 0.000 | 56.370 |
| L | 8.480 | 1.026 | 5.383 | 11.594 |
| Overall | 18.330 | 1.469 | 15.452 | 21.208 |
| **CD4-FETP-KNG1** |  |
| H | 50.030 | 14.498 | 37.480 | 95.790 | 0.000 | 56.370 |
| L | 8.480 | 0.769 | 6.972 | 9.988 |
| Overall | 18.330 | 1.894 | 14.617 | 22.043 |
| **DAND5-PPAT-PPIA** |  |
| H | 8.540 | 0.797 | 6.978 | 10.102 | 0.000 | −33.480 |
| L | 42.020 | 15.014 | 12.592 | 71.448 |
| Overall | 24.770 | 19.276 | 0.000 | 62.551 |
| **CD4-FETP-KNG1** |  |
| H | 50.030 | 14.498 | 37.480 | 95.790 | 0.000 | 56.370 |
| L | 8.480 | 0.769 | 6.972 | 9.988 |
| Overall | 18.330 | 1.894 | 14.617 | 22.043 |
| **CD4-FETP-SELP** |  |
| H | 67.580 | 14.208 | 40.084 | 95.076 | 0.000 | 59.010 |
| L | 8.570 | 1.316 | 5.991 | 11.149 |
| Overall | 19.650 | 6.519 | 6.873 | 32.427 |
| **CD4-FETP-KNG1** |  |
| H | 67.580 | 14.208 | 40.084 | 95.076 | 0.000 | 59.010 |
| L | 8.570 | 1.316 | 5.991 | 11.149 |
| Overall | 19.650 | 6.519 | 6.873 | 32.427 |
| **CD4-FETP-SELP** |  |
| H | 67.580 | 14.208 | 40.084 | 95.076 | 0.000 | 59.010 |
| L | 8.570 | 1.316 | 5.991 | 11.149 |
| Overall | 19.650 | 6.519 | 6.873 | 32.427 |
| **CD4-FETP-KNG1** |  |
| H | 67.580 | 14.208 | 40.084 | 95.076 | 0.000 | 59.010 |
| L | 8.570 | 1.316 | 5.991 | 11.149 |
| Overall | 19.650 | 6.519 | 6.873 | 32.427 |
| **CD4-FETP-SELP** |  |
| H | 67.580 | 14.208 | 40.084 | 95.076 | 0.000 | 59.010 |
| L | 8.570 | 1.316 | 5.991 | 11.149 |
| Overall | 19.650 | 6.519 | 6.873 | 32.427 |

| **DFS (Median) of combinations of 18 genes have unknown association with HCC prognosis** |
| **Estimate** | **Std. Error** | **95% confidence interval** | **P** | **Median survival time difference (H-L)** |
| **Lower boundary** | **Upper boundary** | **Lower boundary** | **Upper boundary** |
| Overall | 35.580 | 12.142 | 11.781 | 59.379 |
| **CD4-FETP-KNG1** |  |
| H | 35.580 | 12.142 | 11.781 | 59.379 |
| L | 7.920 | 1.864 | 4.266 | 11.574 |
| Overall | 24.770 | 19.276 | 0.000 | 62.551 |
| **CD4-FETP-SELP** |  |
| H | 8.540 | 0.797 | 6.978 | 10.102 |
| L | 7.920 | 1.864 | 4.266 | 11.574 |
| Overall | 24.770 | 19.276 | 0.000 | 62.551 |
TABLE 3 | The associations of single genes contained in the multi-gene combinations with disease-free survival (DFS) and overall survival (OS) of HCC patients in a TCGA cohort, analyzed by Kaplan-Meier method.

|                  | DFS (Median) of single genes of the combinations with HCC prognosis | OS (Median) of single genes of the combinations with HCC prognosis |
|------------------|---------------------------------------------------------------------|------------------------------------------------------------------|
|                  | Estimate | Std. Error | 95% confidence interval | P       | Median survival time difference (H-L) | Estimate | Std. Error | 95% Confidence Interval | P       | Median survival time difference (H-L) |
|                  |          |            | Lower boundary | Upper boundary |          |          |            | Lower boundary | Upper boundary |
| APOB             | 29.300   | 6.376      | 16.802 | 41.798 | 0.008 | 14.450 | ANXA7       | 83.180       | 15.496 | 52.807 | 113.553 | 0.006 | 36.430 |
|                  | 14.850   | 2.049      | 10.834 | 18.866 | Overall | 55.650 | 7.925 | 40.116 | 71.184 |
| Overall          | 20.930   | 2.318      | 16.387 | 25.473 |          |          |          |          |          |
| C5               | 29.960   | 6.762      | 16.706 | 43.214 | 0.001 | 16.330 | ATG7       | 45.070       | 8.031 | 29.330 | 60.810 | 0.009 | -35.610 |
|                  | 13.630   | 2.870      | 8.006  | 19.254 | Overall | 55.650 | 7.925 | 40.116 | 71.184 |
| Overall          | 20.930   | 2.318      | 16.387 | 25.473 |          |          |          |          |          |
| CD4              | 36.700   | 7.693      | 21.622 | 51.778 | 0.000 | 23.070 | CA9        | 37.290       | 8.317 | 20.989 | 53.591 | 0.000 | -32.720 |
|                  | 13.630   | 2.089      | 9.536  | 17.724 | Overall | 55.650 | 7.925 | 40.116 | 71.184 |
| Overall          | 20.930   | 2.318      | 16.387 | 25.473 |          |          |          |          |          |
| CETP             | 35.580   | 5.896      | 24.023 | 47.137 | 0.002 | 21.450 | CCNA2      | 45.070       | 10.298 | 24.885 | 65.255 | 0.001 | -24.940 |
|                  | 14.130   | 1.799      | 10.605 | 17.655 | Overall | 55.650 | 7.925 | 40.116 | 71.184 |
| Overall          | 20.930   | 2.318      | 16.387 | 25.473 |          |          |          |          |          |
| COL18A1          | 27.200   | 4.885      | 17.622 | 36.775 | 0.005 | 11.600 | CHDH1L     | 39.750       | 6.940 | 26.148 | 53.352 | 0.006 | -40.930 |
|                  | 15.600   | 3.114      | 9.497  | 21.703 | Overall | 55.650 | 7.925 | 40.116 | 71.184 |
| Overall          | 20.930   | 2.318      | 16.387 | 25.473 |          |          |          |          |          |
| CPT2             | 29.300   | 4.767      | 19.956 | 38.644 | 0.005 | 14.350 | EZH2       | 37.290       | 10.816 | 17.335 | 57.245 | 0.000 | -43.390 |
|                  | 14.950   | 1.836      | 11.352 | 18.548 | Overall | 55.650 | 7.925 | 40.116 | 71.184 |
| Overall          | 20.930   | 2.318      | 16.387 | 25.473 |          |          |          |          |          |
| DAND5            | 13.630   | 2.561      | 8.610  | 18.650 | 0.001 | -16.330 | F2         | 69.510       | 11.842 | 46.300 | 92.720 | 0.005 | 23.620 |
|                  | 29.960   | 5.455      | 19.269 | 40.651 | Overall | 55.650 | 7.925 | 40.116 | 71.184 |
| Overall          | 20.930   | 2.318      | 16.387 | 25.473 |          |          |          |          |          |
| DNASE1           | 13.140   | 1.997      | 9.226  | 17.054 | 0.001 | -16.160 | GMPS       | 45.070       | 9.687 | 26.123 | 64.017 | 0.003 | -24.440 |
|                  | 29.300   | 4.256      | 20.958 | 37.642 | Overall | 55.650 | 7.925 | 40.116 | 71.184 |
| Overall          | 20.930   | 2.318      | 16.387 | 25.473 |          |          |          |          |          |
| F7               | 33.900   | 8.191      | 17.846 | 49.954 | 0.000 | 18.490 | G0T2       | 70.010       | 12.025 | 46.441 | 93.579 | 0.000 | 32.260 |
|                  | 15.410   | 1.485      | 12.500 | 18.320 | Overall | 55.650 | 7.925 | 40.116 | 71.184 |
| Overall          | 20.930   | 2.318      | 16.387 | 25.473 |          |          |          |          |          |
| FLT3             | 35.580   | 3.640      | 28.446 | 42.714 | 0.000 | 22.440 | HPX        | 69.510       | 10.518 | 48.894 | 90.126 | 0.002 | 23.620 |
|                  | 13.140   | 1.833      | 9.547  | 16.733 | Overall | 55.650 | 7.925 | 40.116 | 71.184 |
| Overall          | 20.930   | 2.318      | 16.387 | 25.473 |          |          |          |          |          |
| GNMT             | 29.300   | 9.167      | 11.334 | 47.266 | 0.002 | 13.370 | HDAC2      | 45.070       | 8.365 | 28.675 | 61.465 | 0.002 | -35.610 |
|                  | 15.930   | 1.821      | 12.360 | 19.500 | Overall | 55.650 | 7.925 | 40.116 | 71.184 |
| Overall          | 20.930   | 2.318      | 16.387 | 25.473 |          |          |          |          |          |
| Gene   | DFS (Median) of single genes of the combinations with HCC prognosis | OS (Median) of single genes of the combinations with HCC prognosis |
|--------|---------------------------------------------------------------|---------------------------------------------------------------|
|        | Estimate | Std. Error | 95% confidence interval | P | Median survival time difference (H-L) | Estimate | Std. Error | 95% Confidence Interval | P | Median survival time difference (H-L) |
|        | Lower boundary | Upper boundary | | | Lower boundary | Upper boundary | |
| **ITGB2** | H | 35.580 | 4.232 | 27.285 | 43.875 | 0.002 | 19.840 | KPNAA2 | H | 35.020 | 8.165 | 17.017 | 49.023 | 0.000 | −47.660 |
|        | L | 15.740 | 2.671 | 10.504 | 20.976 | Overall | 20.930 | 2.318 | 16.387 | 25.473 | Overall | 55.650 | 7.925 | 40.116 | 71.184 |
|        | L | 15.700 | 2.458 | 10.882 | 20.518 | Overall | 20.930 | 2.318 | 16.387 | 25.473 | Overall | 55.650 | 7.925 | 40.116 | 71.184 |
| **KNG1** | H | 25.300 | 6.478 | 12.603 | 37.997 | 0.007 | 9.600 | LAPTMB4 | H | 45.070 | 10.511 | 24.468 | 65.672 | 0.000 | −35.610 |
|        | L | 15.740 | 2.671 | 10.504 | 20.976 | Overall | 20.930 | 2.318 | 16.387 | 25.473 | Overall | 55.650 | 7.925 | 40.116 | 71.184 |
| **LMCO1** | H | 25.300 | 6.478 | 12.603 | 37.997 | 0.007 | 9.600 | LAPTMB4 | H | 45.070 | 10.511 | 24.468 | 65.672 | 0.000 | −35.610 |
|        | L | 15.740 | 2.671 | 10.504 | 20.976 | Overall | 20.930 | 2.318 | 16.387 | 25.473 | Overall | 55.650 | 7.925 | 40.116 | 71.184 |
| **PRF1** | H | 29.660 | 5.120 | 19.625 | 39.695 | 0.004 | 13.960 | MAPT | H | 41.750 | 6.888 | 24.468 | 55.251 | 0.000 | −28.260 |
|        | L | 15.700 | 2.458 | 10.882 | 20.518 | Overall | 20.930 | 2.318 | 16.387 | 25.473 | Overall | 55.650 | 7.925 | 40.116 | 71.184 |
| **SPPL2A** | H | 27.200 | 5.000 | 17.652 | 41.668 | 0.005 | 11.790 | TRPV1 | H | 80.680 | 6.030 | 33.250 | 56.890 | 0.002 | −35.610 |
|        | L | 15.740 | 2.003 | 11.814 | 19.666 | Overall | 20.930 | 2.318 | 16.387 | 25.473 | Overall | 55.650 | 7.925 | 40.116 | 71.184 |
FIGURE 4 | Association of the top 15 three-gene-combinations with smallest p-values with OS, using the data of HCC samples in a TCGA cohort and assessed by Kaplan-Meier analyses. The high expression group (blue line) of the combination consisted of samples with high expression of all three genes, and the low expression group (green line) of the combination consisted of samples with low expression of all three genes. The number of high and low expression groups in each combination was >25. (A) Association of OS and the top 15 combinations with the smallest p-values of the overall genes combinations. (B) Association of OS and the top 15 combinations with the smallest p-values of the unreported genes combinations.

DISCUSSION

Liver cancer is characterized by inconspicuous early symptoms, a high degree of malignancy, recurrence and spread, and unsatisfactory prognosis. With limited treatment options, it is one of the common malignancies that plague the world. Therefore, identification of effective prognostic biomarkers for liver cancer is the key to improving the efficacy of targeted therapy for HCC and reducing the adverse prognostic effects of liver cancer.

In our study, by combining and searching 15 corresponding concepts of the key words “liver cancer,” “prognosis,” and “outcome,” and according to p-values < 0.05, 1,173 genes that

$F2$-$GOT2$-$TRPV1$ was greater than that of any of the three single genes.

The results of IHC for 90 liver cancer cases showed that the low protein expression of $F2$, $GOT2$, and $TRPV1$ was significantly associated with lower 5-year survival in HCC patients ($F2$: $p = 0.033$, $GOT2$: $p = 0.035$, $TRPV1$: $p = 0.046$; K-M survival analyses). However, due to the insufficient number of events in the co-high expression group of the combination $F2$-$GOT2$-$TRPV1$, there was marginally significant difference found in the overall survival rate of HCC patients between the co-high expression group and the co-low expression group of the protein combination $F2$-$GOT2$-$TRPV1$ ($p = 0.051$) (Figure 8B).
FIGURE 5 | Association of OS and the individual genes contained in the top 15 combinations with the lowest $P$-values, using the data of HCC samples in a TCGA cohort and assessed by Kaplan-Meier analyses. (A) Association of OS and the 14 single genes contained in the first 15 total-gene combinations. (B) Association of OS and the 11 single genes contained in the first 15 unreported-gene combinations.
TABLE 4 | The associations of three-gene combinations with overall survival (OS) of HCC patients in a TCGA cohort, analyzed by Kaplan-Meier method.

| OS (Median) of combinations of 28 genes with HCC prognosis | OS (Median) of combinations of 12 genes have unknown association with HCC prognosis |
|-----------------------------------------------------------|-------------------------------------------------------------|
| Estimate Std. Error 95% confidence interval               | Estimate Std. Error 95% confidence interval               |
|                                                    | P     | Median survival time difference (H-L)                  | P     | Median survival time difference (H-L)                  |
|                                                    | Lower | Upper | boundary |                                                    | Lower | Upper | boundary |
| EZH2-KPNA2-MPV17                                      | H     | 21.320| 6.143    | 9.280  | 33.360 | 0.000 | −59.360 |
|                                                    | L     | 80.680| 7.061    | 66.841 | 94.519 |       |         |
|                                                    | Overall| 55.350| 13.443   | 29.001 | 81.699 |       |         |
| EZH2-LAPTM4B-MPV17                                    | H     | 18.230| 5.735    | 6.988  | 29.472 | 0.000 | −62.450 |
|                                                    | L     | 80.680| 7.990    | 65.020 | 96.340 |       |         |
|                                                    | Overall| 48.950| 10.014   | 29.323 | 68.577 |       |         |
| CA9-KPNA2-SPP1                                       | H     | 23.780| 5.368    | 13.259 | 34.301 | 0.000 | −59.400 |
|                                                    | L     | 83.180| 16.292   | 51.248 | 115.112|       |         |
|                                                    | Overall| 51.250| 11.688   | 28.381 | 74.119 |       |         |
| CA9-KPNA2-LAPTM4B                                    | H     | 19.740| 3.699    | 12.490 | 26.990 | 0.000 | −63.440 |
|                                                    | L     | 83.180| 20.669   | 42.669 | 123.691|       |         |
|                                                    | Overall| 46.750| 6.141    | 34.717 | 58.785 |       |         |
| KPNA2-SLC38A1-SPP1                                   | H     | 19.090| 6.876    | 5.614  | 32.566 | 0.000 | −83.570 |
|                                                    | L     | 102.660| 21.958  | 59.622 | 145.698|       |         |
|                                                    | Overall| 69.510| 10.951   | 48.047 | 90.973 |       |         |
| HDAC2-KPNA2-SPP1                                     | H     | 23.780| 5.613    | 12.778 | 34.782 | 0.000 | −78.880 |
|                                                    | L     | 102.660| 14.189  | 74.850 | 190.470|       |         |
|                                                    | Overall| 83.180| 21.672   | 40.704 | 125.665|       |         |
| CHD1L-EZH2-SPP1                                     | H     | 15.410| 4.012    | 7.547  | 23.273 | 0.000 | −67.770 |
|                                                    | L     | 83.180| 9.866    | 63.842 | 102.518|       |         |
|                                                    | Overall| 48.750| 12.305   | 22.633 | 70.867 |       |         |
| EZH2-KPNA2-LAPTM4B                                   | H     | 21.680| 5.445    | 11.008 | 32.352 | 0.000 | −59.000 |
|                                                    | L     | 83.510| 15.702   | 52.734 | 114.286|       |         |
|                                                    | Overall| 45.070| 14.673   | 31.251 | 89.769 |       |         |

(Continued)
| Combination | OS (Median) of combinations of 28 genes with HCC prognosis | OS (Median) of combinations of 12 genes have unknown association with HCC prognosis |
|-------------|---------------|-------------------------------------------------|
|             | Estimate | Std. Error | 95% confidence interval | Median survival time difference (H-L) | Estimate | Std. Error | 95% confidence interval | Median survival time difference (H-L) |
|             |          |            |                          |                                  |          |            |                          |                                  |
| L           | 80.680   | 7.011      | 86.939                  | 94.421                          | L        | 24.870     | 10.561                  | 4.170                            |
| Overall     | 46.750   | 10.389     | 26.388                  | 67.112                          | Overall  | 53.350     | 14.230                  | 25.459                           |
| KPNA2-MPV17-SC38A1 | H 17.580 | 5.820      | 6.172                  | 28.988                          | L        | 80.680     | 7.992                  | 65.015                             |
|             | 53.350   | 11.888     | 30.049                  | 76.651                          | L        | 21.120     | 6.087                  | 9.190                             |
| Overall     | 80.680   | 10.953     | 59.212                  | 102.148                         | Overall  | 45.530     | 11.839                 | 22.325                            |
| GMP5-APAT7-SC2A1 | H 17.970 | 6.680      | 4.876                  | 31.064                          | L        | 83.180     | 16.018                 | 51.784                            |
|             | 53.350   | 13.679     | 23.325                 | 68.735                          | L        | 20.600     | 8.050                 | 5.201                             |
| Overall     | 80.680   | 12.139     | 54.680                 | 102.148                         | Overall  | 53.290     | 13.890                 | 26.066                            |
|                 | 65.210   | 5.820      | 28.988                  | 94.421                          | L        | 83.180     | 14.580                 | 54.602                            |
|                 | 69.214   | 6.650      | 30.136                  | 96.214                          | L        | 27.500     | 8.050                 | 3.201                             |
|                 | 55.530   | 6.190      | 47.057                 | 99.036                          | L        | 37.290     | 6.650                 | 7.016                             |
|                 | 70.010   | 10.953     | 59.122                 | 102.148                         | Overall  | 53.350     | 14.898                 | 22.050                            |
|                 | 48.950   | 14.230     | 30.984                 | 67.112                          | Overall  | 51.250     | 13.888                 | 24.030                            |
|                 | 58.310   | 10.953     | 46.345                 | 102.148                         | Overall  | 48.950     | 13.888                 | 24.030                            |

TABLE 4 | Continued
Liu et al. Multi-Gene Prognostic Signatures Identification

FIGURE 6 | Expression of F2, GOT2, and TRPV1 in HCC and adjacent normal liver tissues confirmed by independent microarrays from the Oncomine database. The expression of (A) F2, (B) GOT2, and (C) TRPV1 were all significantly reduced in HCC tissues by the Roessler Liver 2 Statistics [225 HCC tissues (dark blue) vs. 220 normal liver tissues (light blue)]. ***p < 0.001.

FIGURE 7 | The relative expression levels of F2, GOT2, and TRPV1 were assessed by qRT-PCR in HCC tissues and adjacent liver tissues. (A) The expression levels of F2 were assessed by qRT-PCR in 20 pairs of HCC tissues and peritumoral tissues. Paired t-test, ***p < 0.001. (B) The relative expression levels of GOT2 were assessed by qRT-PCR in 19 pairs of HCC tissues and peritumoral tissues. Paired t-test, ***p < 0.001. (C) The relative expression levels of TRPV1 were assessed by qRT-PCR in 19 pairs of HCC tissues and peritumoral tissues. Paired t-test, log, **p < 0.01.

may be related to the prognosis of liver cancer were mined from the Coremine platform after merging and removing duplicates. However, due to the insufficient sample size and data related to the prognosis of liver cancer in the Coremine platform as well as the large heterogeneity among the samples, we also selected gene expression data and prognosis data of 319 samples for DFS and 370 samples for OS from the TCGA platform. We then separately conducted DFS-related and OS-related K-M survival analysis for each gene, followed by multivariate analyses, respectively. The large-scale genes mining and a large number of homogenous samples gave us a reliable analytical foundation. By far, this is the first large-scale survival analyses for hundreds of genes for subsequent screening.

In addition, the genes selected by K-M survival analyses with a low p-value (p < 0.01) were further screened by multivariate analyses using the Cox proportional hazards regression model. We found that 39 genes and 28 genes were reliably and significantly associated with DFS and OS, respectively, in liver cancer. Many of the above genes have been confirmed to be associated with the prognosis of HCC by previous reports. For example, of the 39 DFS-related genes, ALDOB inhibits metastasis in HCC and can be a valuable novel prognosis predicting marker (30); APOB was found to be a prognostic biomarker for patients with radical resection of HCC (31, 32); CCNF is downregulated in HCC and is a promising prognostic marker (33). In addition, CPT2 (34), G6PD (35), GNMT (36), NEK2 (37), etc. have also been reported to be prognostic markers of HCC by affecting the occurrence or invasion of HCC. The above findings are consistent with what we identified. Other genes, such as C5, CD4, CETP, COL18A1, DAND5, DNASE1, EBPL, F7, FLT3, ITGB2, KNG1, LMOD1, PPAT, PPIA, PRF1, SELP, SPPL2A, and TRPV1 that have not been systematically reported in relation to the prognosis of liver cancer, are our newly discovered prognostic markers for DFS in liver cancer. Similarly, of the 28 OS-related genes, CA9 regulates the epithelial-mesenchymal transition and is a novel prognostic marker in HCC (38), E2F1 expression has an impact on tumor aggressiveness and affects the prognosis of HCC (14, 15), CYP3A4 (39), HDAC2 (40), and KPNA2 (41) have also been identified as prognostic markers of HCC and are reflected in our findings. The other genes, such as ANXA7, F2, GOT2, HPX,
FIGURE 8 | The expression of F2, GOT2, and TRPV1 in 90 pairs of HCC and adjacent normal liver tissues of biological tissue microarray by IHC, and the association with HCC patients prognosis. (A) Negative, weakly positive, intermediately positive, and strongly positive IHC staining of F2, GOT2, and TRPV1. F2, GOT2, and TRPV1 were all low expressed in liver cancer. (B) The lower protein expression levels of F2, GOT2, and TRPV1 were all associated with 5-year OS of 90 HCC patients, examined by Kaplan-Meier analyses and log-rank test. However, there was marginally significant association between the F2-GOT2-TRPV1 combination protein expression levels with the OS of HCC patients. (F2: p = 0.033, GOT2: p = 0.035, TRPV1: p = 0.046, F2-GOT2-TRPV1: p = 0.051).
MAGEB3, MAPT, MPV17, NTF3, PPAT, SLC2A1, SLC38A1, and TRPV1 are all novel prognostic markers associated with liver cancer OS found by our reliable and large-scale screening studies. Three genes (APOB, PPAT, and TRPV1) were associated with both DFS and OS of HCC, suggesting that APOB, PPAT, and TRPV1 may be significant and effective in predicting both the progress and the adverse outcomes of HCC. Moreover, there may be connections among the above selected genes and they can work together to influence the development and prognosis of liver cancer to some extent. Although there are some genes that had been reported as prognostic molecular markers of liver cancer, most reports focused on the impact of a single gene on the prognosis of liver cancer, few studies performed such a large-scale survival analysis. Studies of multiple gene combinations are more effective than the analysis of single genes in predicting the prognosis of liver cancer.

In our study, we performed three-gene combinations of the 39 DFS-related genes and 28 OS-related genes screened from the above survival analyses. In order to further study the predictive effect of the combinations constituted by the selected genes on the prognosis of liver cancer, and to compare the predictive power of single genes and corresponding gene combinations, we carried out thousands of K-M survival analyses on these combinations. To ensure the comparability and credibility, we removed the combinations of which the co-high or co-low expression group cases were fewer than 26, and screened 2,758 DFS-related combinations and 930 OS-related combinations with \( p \)-values < 0.01. Moreover, we also performed three-gene-combination models and K-M survival analyses on the 18 DFS-related genes and 12 OS-related genes we found but have not been systematically reported to be related to the prognosis of HCC. 317 unreported-gene combinations and 31 unreported-gene combinations significantly associated with DFS and OS, respectively, were screened out.

For the above four types of three-gene-combinations (the overall genes combinations associated with DFS, the unreported genes combinations associated with DFS, the overall genes combinations associated with OS, and the unreported genes combinations associated with OS), the top 15 combinations with the lowest \( p \)-values of the survival analyses and the genes they contained were, respectively, selected for comparison (Tables 2, 3, 4).

For example, for the overall gene combinations associated with OS, KPNA2-SLC38A1-SPP1, the median survival time difference between the co-high and the co-low expression group was 83.57 months. In contrast, that of the single genes KPNA2, SLC38A1, and SPP1, was 47.66, 35.61, and 29.64 months, respectively. After combining KPNA2, SLC38A1, and SPP1, the median survival time difference between the high and low expression groups was larger than that of any of the three single genes by at least 36 months. This shows that these three genes KPNA2, SLC38A1, and SPP1, after combination, may be better predictive values for liver cancer prognosis and may be more clinically useful for future treatment target selection.

We also selected genes that have not been previously reported for liver cancer prognosis and compared their prognostic efficacy with the corresponding three-gene combinations (the chart only shows the top 15 groups with the lowest \( p \)-values of the three-gene combinations prognostic models). The expression of one of the combinations F2-GOT2-TRPV1 had a greater effect on the median survival time of OS than any of the three individual genes (The median survival time difference: F2-GOT2-TRPV1: 55.68 months; F2: 23.62 months; GOT2: 32.26 months; TRPV1: 35.61 months).

Coagulation factor II (F2) plays a major role in proteolysis to form thrombin in the first step of the coagulation cascade and eventually generates hemostasis. An enrichment analysis of genetic changes during the development of HCC identified several hub genes, including F2, which interacts in several groups of conditional specific PPI networks (42). Additionally, it was reported that F2 is associated with invasion in neuroendocrine prostate cancer (43). Glutamic-oxaloacetic transaminase 2 (GOT2) plays an important role in amino acid metabolism and the tricarboxylic acid cycle, and it affects the malate-aspartic acid shuttle activity and glycolysis in the liver under the stimulation of liver inflammation. (44, 45) TRPV1 is a regulator of cell homeostasis, previous studies have revealed that the expression of TRPV1 is significantly decreased in renal cell carcinoma, colorectal cancer, and melanoma. In addition, TRPV1 can affect P53 and TRPV1-dependent pathways to inhibit the growth of colorectal cancer and melanoma (46–48), and can cause apoptosis in human osteosarcoma MG63 cells (49).

At present, there are few studies on the above three genes F2, GOT2, TRPV1 and particular their combinations in the prognosis of HCC. In our study, the results of the 20 pairs of HCC and paracancerous tissues for qRT-PCR, as well as 90 pairs HCC biochips for IHC confirmed that all of the F2, GOT2, and TRPV1 genes are significantly and consistently down-expressed in HCC tissues, and this is reconfirmed by three independent microarrays. Moreover, the low expression of F2, GOT2, and TRPV1 were all significantly associated with poor prognosis of HCC. However, due to the number of death events in the F2-GOT2-TRPV1 high expression group of in the HCC biochips being 0, the survival analysis of the F2-GOT2-TRPV1 high and the expression group was marginally significant (\( p = 0.051 \)), but this is still consistent with our above-mentioned big data-based multi-gene combination survival analysis results.

As there may be certain relationships between the genes we screened that are significantly associated with the prognosis of liver cancer, they can work together in the form of multi-gene combinations in the development of liver cancer. However, the predictive potency of different gene combinations varies. Some combinations are better predictors than individual genes, and therefore these combinations may be more valuable than individual genes in determining the target site for liver cancer prognosis. Due to limitations in human and material resources, it still remains unclear how these genes and gene combinations specifically affect the HCC survival. Further investigation and experimentation are needed to elucidate the biological mechanisms of the selected genes, particularly for the significant multi-gene combinations, in the development and progression of HCC.

Our findings cover a large gene level, and we have also explored the predictive efficacy of a number of gene
combinations for the prognosis of liver cancer. We believe that these highly significant prognostic-related genes and gene combinations derived from the above multiple screenings are promising, reliable molecular markers for the prognosis of liver cancer, and our screening methods can be extended to other tumor types.

In conclusion, based on a large sample size of public data platform, novel and effective data mining and multiple screening methods, large-scale survival analyses, as well as supplemental reliable experimental verification, we identified a series of novel genes and multi-gene combinations that are significantly associated with DFS or OS in liver cancer. Moreover, a huge difference between high and low expression group of these three-gene combination was detected. Some of the three-gene combinations can predict much longer or shorter survival time for liver cancer patients than the single genes. QRT-PCR, immunohistochemistry, and three independent microarray results confirmed our findings that three of the selected novel genes \(F2\), \(GOT2\), and \(TRPV1\), as well as the corresponding combination \(F2\)-\(GOT2\)-\(TRPV1\), showed significantly lower expression in HCC and are associated with OS in HCC. Some gene combinations may be more predictors of prognosis than single genes and can be used as potential effective therapeutic targets for liver cancer.

**DATA AVAILABILITY STATEMENT**

The datasets generated for this study are available on request to the corresponding author.

**ETHICS STATEMENT**

The studies involving human participants were reviewed and approved by The Ethics Committee of Guangxi Medical University. The patients/participants provided their written informed consent to participate in this study.

**AUTHOR CONTRIBUTIONS**

ML and XLi performed most analysis. ML led the writing of the manuscript. SL provided the clinical samples and participated in revising the manuscript. FX and JT participated in drafting and reviewing the manuscript. EG conducted a search for genes and preliminary screening work by keyword. XQ obtained and matched the TCGA samples data. ML, LW, and QL performed the single-gene and multi-gene-combination survival analyses. ZL and LL conducted an inquiry about the relevant information of the selected genes. XLu performed validation of the selected genes in three microarrays. KL and DZ performed the mRNA isolation and qRT-qPCR, and collected and analyzed experimental data. YY and XLI were subjected to immunohistochemistry and experimental data processing. FY and XZ participated in designing and reviewing the study. All the authors reviewed the manuscript and all the authors read and approved the final manuscript.

**FUNDING**

This study was supported by the National Natural Science Foundation of China (Grant No. 81760611) and the Key Laboratory of High-Incidence-Tumor Prevention and Treatment (Guangxi Medical University), Ministry of Education (No. GKE 2019–01).

**SUPPLEMENTARY MATERIAL**

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2020.00847/full#supplementary-material

**REFERENCES**

1. IARC. Fact sheets by Population-Globocan-IARC. (2019). Retrieved from: http://gco.iarc.fr/today/fact-sheets-cancers (accessed September 14, 2019).
2. 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. doi:10.3322/caac.21492
3. Qiu WQ, Shi JF, Guo LW, Mao AY, Huang HY, Hu GY, et al. Medical expenditure for liver cancer in urban China: A 10-year multicenter retrospective survey (2002-2011). J Cancer Res Ther. (2018) 14:163–70. doi:10.4103/jcrt.JCRT_709_16
4. Deng W, Long L, Li J, Zheng D, Yu JH, Zheng CY, et al. Mortality of major cancers in Guangxi, China: sex, age and geographical differences from 1971 and 2005. Asian Pac J Cancer Prev. (2014) 15:1567–74. doi:10.7314/APJCP.2014.15.4.1567
5. Zeng H, Chen W, Zheng R, Zhang S, Ji JS, Zou X, et al. Changing cancer survival in China during 2003-15: a pooled analysis of 17 population-based cancer registries. Lancet Glob Health. (2018) 6:e555–67. doi:10.1016/S2214-109X(18)30127-X
6. Wang CY, Li S. Clinical characteristics and prognosis of 2887 patients with hepatocellular carcinoma: A single center 14 years experience from China. Medicine. (2019) 98:e14070. doi:10.1097/MD.00000000000014070
7. Zhu J, Yin T, Xu Y, Lu XJ. Therapeutics for advanced hepatocellular carcinoma: Recent advances, current dilemma, and future directions. J Cell Physiol. (2019) 234:1222–32. doi:10.1002/jcp.28048
8. Finn RS, Zhu AX, Farah W, Almazri J, Zaim F, Prokop LJ, et al. Therapies for advanced stage hepatocellular carcinoma with macrovascular invasion or metastatic disease: a systematic review and meta-analysis. Hepatology. (2018) 67:422–35. doi:10.1002/hep.29486
9. Cai W, Wang Z, Wei C, Wu M, Zheng W, Zhang H, et al. Prognostic evaluation of OCT4 expression for posttransplantation hepatocellular carcinoma recurrence. J Cell Biochem. (2019) 120:8419–29. doi:10.1002/jcb.28128
10. Forner A, Reig M, Bruix J. Hepatocellular carcinoma. Lancet. (2018) 391:1301–14. doi:10.1016/S0140-6736(18)30019-2
11. Fan ST, Mao LO, Poon RT, Yeung C, Leung Liu C, Yuen WK, et al. Continuous improvement of survival outcomes of resection of hepatocellular carcinoma: a 20-year experience. Ann Surg. (2011) 253:745–58. doi:10.1097/SLA.0b013e3182111195
12. Andreozzi M, Quintravalle C, Benz D, Quaglita L, Matter M, Calabrese D, et al. HMGA1 expression in human hepatocellular carcinoma correlates with poor prognosis and promotes tumor growth and migration in vitro models. *Neoplasia*. (2016) 18:724–31. doi: 10.1016/j.neo.2016.10.002

13. Chen DH, Wu QW, Li XD, Wang SJ, Zhang ZM. SYPL1 overexpression predicts poor prognosis of hepatocellular carcinoma and associates with epithelial-mesenchymal transition. *OncoRep*. (2017) 38:1533–42. doi: 10.3892/or.2017.5843

14. Huang YL, Ning G, Chen LB, Lian YF, Gu YR, Wang JL, et al. Promising diagnostic and prognostic value of E2Fs in human hepatocellular carcinoma. *Cancer Manag Res*. (2019) 11:7375–40. doi: 10.2147/CMAR.S182001

15. Han R, Chen X, Li Y, Zhang S, Li R, Lu L. MicroRNA-34a suppresses aggressiveness of hepatocellular carcinoma by modulating E2F1, E2F3, and Caspase-3. *Cancer Manag Res*. (2019) 11:2963–76. doi: 10.2147/CMAR.S20664

16. Liang R, Lin Y, Ye JZ, Yan XX, Liu ZH, Li YQ, et al. High expression of RBM8A predicts poor patient prognosis and promotes tumor progression in hepatocellular carcinoma. *OncoRep*. (2017) 37:2167–76. doi: 10.3892/or.2017.5457

17. Chang L, Xi L, Liu Y, Liu R, Wu Z, Jian Z. SIRT5 promotes cell proliferation and invasion in hepatocellular carcinoma by targeting E2F1. *Med Mol. Rep*. (2018) 17:342–9. doi: 10.3892/mmr.2017.7787

18. Zhou L, Zhu Y. The EpCam overexpression is associated with clinicopathological significance and prognosis in hepatocellular carcinoma patients: A systematic review and meta-analysis. *Int J Surg*. (2018) 56:274–80. doi: 10.1016/j.ijsu.2018.06.025

19. Long H, Guo X, Qiao S, Huang Q. Tumor LXR expression is a prognostic marker for patients with hepatocellular carcinoma. *Pathol Oncol Res*. (2018) 24:339–44. doi: 10.1007/s12253-017-0249-8

20. Zhou H, Wang SC, Ma JM, Yu LQ, Jing JS. SYPL1 overexpression predicts poor prognosis of hepatocellular carcinoma and associates with epithelial-mesenchymal transition. *OncoRep*. (2017) 38:1533–42. doi: 10.3892/or.2017.5843

21. Fu J, Qiu H, Cai M, Pan Y, Cao Y, Liu L, et al. Low cyclin F expression in hepatocellular carcinoma associates with poor differentiation and unfavorable prognosis. *Cancer Sci*. (2013) 104:508–15. doi: 10.1111/cas.12100

22. Fujiwara N, Nakagawa H, Enokou K, Kudo Y, Hayata Y, Nakatsuka T, et al. CPT2 downregulation adapts HCC to lipid-rich environment and promotes carcinogenesis via acylcarnitine accumulation in obesity. *Gut*. (2018) 67:1384–9. doi: 10.1136/gutjnl-2017-315193

23. Chang YY, Yen CJ, Chen CH, Chiu YW, Lee YP, Bao CY, et al. NEK2 Promotes hepatoma metastasis and serves as biomarker for high recurrence risk after hepatic resection. *Ann Hepatol*. (2018) 17:843–56. doi: 10.5604/03.1001.0011.0314

24. Hyuga S, Wada H, Eguchi H, Otsubu T, Iwagami Y, Yamada D, et al. Expression of carbonic anhydrase IX is associated with poor prognosis through regulation of the epithelial-mesenchymal transition in hepatocellular carcinoma. *Int J Oncol*. (2017) 51:1179–90. doi: 10.3892/ijo.2017.4098

25. Ashida R, Okamura Y, Ohshima K, Kakuda Y, Uesaka K, Sugiura T, et al. GYP3A4 Gene is a novel biomarker for predicting a poor prognosis in hepatocellular carcinoma. *Cancer Genomics Proteomics*. (2017) 14:445–53. doi: 10.21873/cgp.20054

26. Quint K, Agaimy A, Di Fazio P, Montalbano R, Steindorf C, Jung R, et al. Clinical significance of histone deacetylases 1, 2, 3, and 7: HDA C2 is an independent predictor of survival in HCC. *Virchows Arch*. (2018) 459:129–39. doi: 10.1007/s00428-011-1103-6

27. Jiang P, Tang Y, He L, Tang H, Liang M, Mai C, et al. Aberrant expression of TRPV1 in renal cell carcinoma: association with tumor shuttle activity and pancreatic tumor growth. *Embo J*. (2015) 34:1110–25. doi: 10.15252/embj.201591041

28. Wu YY, Liu XY, Zhuo DX, Huang HB, Zhang FB, Liao SF. Decreased expression of TRPV1 in renal cell carcinoma: association with tumor Fuhrman grades and histopathological subtypes. *Cancer Manag Res*. (2018) 10:1647–55. doi: 10.2147/CMAR.S166390
47. Hou N, He X, Yang Y, Fu J, Zhang W, Guo Z, et al. TRPV1 Induced apoptosis of colorectal cancer cells by activating calcineurin-NFAT2-p53 signaling pathway. *Biomed Res Int.* (2019) 2019:6712536. doi: 10.1155/2019/6712536

48. Yang Y, Guo W, Ma J, Xu P, Zhang W, Guo S, et al. Downregulated TRPV1 expression contributes to melanoma growth via the calcineurin-ATF3-p53 pathway. *J Invest Dermatol.* (2018) 138:2205–15. doi: 10.1016/j.jid.2018.03.1510

49. Bao Z, Dai X, Wang P, Tao Y, Chai D. Capsaicin induces cytotoxicity in human osteosarcoma MG63 cells through TRPV1-dependent and -independent pathways. *Cell Cycle.* (2019) 18:1379–92. doi: 10.1080/15384101.2019.1618119

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

*Copyright © 2020 Liu, Liu, Xiao, Guo, Wu, Liang, Li, Zhang, Yang, Luo, Lei, Tan, Yin and Zeng. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*