ARTICLE   OPEN

Prediction of CAF-related genes in immunotherapy and drug sensitivity in hepatocellular carcinoma: a multi-database analysis

Yi Yao[1], KaiQing Yang[1], Qiang Wang[1], Zeming Zhu[2], Sheng Li[1], Bin Li[1], Bin Feng[3 ✉] and Caixi Tang [ID][3 ✉]

© The Author(s) 2024

This study aims to identify the cancer-associated fibroblasts (CAF)-related genes that can affect immunotherapy and drug sensitivity in hepatocellular carcinoma (HCC). Expression data and survival data associated with HCC were obtained in The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) databases. Weighted correlation network analysis (WGCNA) analysis was performed to obtain CAF-related genes. Least Absolute Shrinkage and Selection Operator (LASSO) regression was used for regression analysis and risk models. Subsequently, Gene Set Enrichment Analysis (GSEA) analysis, Gene Set Enrichment Analysis (ssGSEA) analysis, Tumor Immune Dysfunction and Exclusion (TIDE) analysis and drug sensitivity analysis were performed on the risk models. Survival analysis of CAF scores showed that the survival rate was lower in samples with high CAF scores than those with low scores. However, this difference was not significant, suggesting CAF may not directly influence the prognosis of HCC patients. Further screening of CAF-related genes yielded 33 CAF-related genes. Seven risk models constructed based on CDR2L, SPRED1, PFKP, ENG, KLF2, FSCN1 and VCAN, showed significant differences in immunotherapy and partial drug sensitivity in HCC. Seven CAF-related genes may have important roles in immunotherapy, drug sensitivity and prognostic survival in HCC patients.

INTRODUCTION

Currently, hepatocellular carcinoma (HCC) is considered one of the most common and deadly malignancies in the world, with limited treatment options available because of the complications in disease progression and therapeutic response caused by heterogeneity within the liver tumor or between each tumor case [1]. Intrahepatic or systemic metastases are the leading cause of poor prognosis in patients with advanced HCC [2], among which lung metastasis is the most common and one of the leading causes of death in HCC patients [3, 4]. Previous data shows that cirrhosis or liver fibrosis occurs in approximately 80~90% of HCC progression and is a necessary intermediate state to develop HCC. About one-third of patients with cirrhosis carry the risk of HCC through their lifetime [5, 6]. Therefore, further understanding the impact of fibroblasts in HCC and exploring the genes associated with cancer-associated fibroblasts (CAFs) in HCC may provide new ideas for therapeutic strategies for this disease.

CAFs, originating mainly from activated hepatic stellate cells and increasingly accumulating in HCC stroma, are significant sources of extracellular matrix (ECM) [7–11]. The growing evidence supported the role of CAF in tumor infiltration, migration and invasion [12, 13]. In the past decades, many CAF-related genes, such as CCL2 and CCL26, have been suggested to be involved in tumor development [14, 15]. CXCL11 in CAFs was reported for its implication in HCC migration [16]. HCC-related CAFs and their related genes may be essential in regulating HCC cells [7, 17–19]. Studies on CAFs have provided new directions for further insights into tumor development. Despite the evidence supporting the regulatory role of CAFs in HCC, research focusing on CAF-related genes in HCC remains scarce.

Weighted gene co-expression network analysis (WGCNA) is a systematic bioinformatics algorithm that integrates highly coordinated genes into multiple gene modules and analyzes the correlation between modules and target phenotypes [20]. In this study, CAF levels in HCC samples were obtained by scoring CAF to identify gene modules associated with CAF scoring. Through integrating the data in multiple databases, new CAF-related genes associated with HCC were located to analyze their effect on immunotherapy and drug sensitivity to provide a theoretical basis for further insight into the mechanisms of CAF-related genes in HCC, as well as a new direction for exploring new therapeutic tools for HCC.

MATERIALS AND METHODS

Data download and processing

Liver Hepatocellular Carcinoma (LIHC) samples in The Cancer Genome Atlas (TCGA) (https://portal.gdc.cancer.gov/) were selected and downloaded to obtain HCC gene expression data and clinical data, in which 371 HCC samples and corresponding clinical data were extracted. Microarray GSE76427 with the sequencing platform GPL10558 in GEO

[1]Division 1, Department of Hepatobiliary and Pancreatic Surgery, Zhuzhou Central Hospital, Zhuzhou, Hunan, China. [2]Division 2, Department of Hepatobiliary and Pancreatic Surgery, Zhuzhou Central Hospital, Zhuzhou, Hunan, China. [3]Department of Hepatobiliary and Pancreatic Surgery, Zhuzhou Central Hospital, Zhuzhou, Hunan, China. ✉email: Binfengfb72@126.com; Tangcaixi64@126.com

(https://www.ncbi.nlm.nih.gov/geo/) database was selected, which includes 115 HCC samples and corresponding clinical expression files.

CAF scores and survival analysis

The expression data was normalized using the R language "limma" package v3.54.2. TCGA data were processed using the "averesps" function and GSE76427 microarray was processed using the "normalizeBetweenArrays" function [21]. The R (v4.2.3) "MCPcounter" package v1.2.0 was used to calculate CAF scores for data in TCGA data and GSE76427 microarray, using the "MCPcounter.probesets" and "MCPcounter.probesets" [22, 23]. The CAF scores of TCGA and GEO were merged with the clinical survival data and the survival rate was analyzed using the "survival" v3.5-3 and "surviminer" v0.4.9 packages. The surv_cutpoint function in the surviminer package was used to calculate the best cutoff value and to plot survival curves.

WGCNA analysis

The CAF scores of TCGA and GSE76427 were extracted separately for analysis using WGCNA package v1.72-1 [20] to identify CAF-related module genes. This method analyzes the association between genes and classifies genes into different colored modules. Genes not associated with other genes are organized into insignificant gray modules. The recommended soft threshold values automatically calculated were used for the soft threshold selection procedure. The similarity between modules was computed using the R language cor function based on the eigengenes of the modules. The genes with high similarity were merged and the merged modules were analyzed to determine the correlation between module and CAFs.

Functional enrichment analysis

Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) functional enrichment analyses were performed on the intersecting genes of the CAF-related modules from TCGA and GSE76427 analyses using R language "clusterProfiler" package v4.6.2 [24, 25]. For GO enrichment analysis, enrichGO function was performed using $p < 0.05$ as the significant pathway filter condition, while enrichKEGG function was used for KEGG pathway enrichment analysis.

Construction or risk models

The expression data of those intersecting genes in TCGA were extracted and merged with corresponding clinical data in TCGA using for univariate cox analysis using the "survival" and "surviminer" packages, and the cxoph function to obtain the candidate genes that significantly affected the prognosis of HCC patients. The "glmnet" v4.1-7 and "surviviner" package [26, 27], as well as glmnet and cv.glmnet function, were used for LASSO regression analysis based on the candidate genes. The optimal prognostic genes ($p < 0.05$) were located based on multi-factor cox regression analysis. The point with the lowest cross-validation error were selected for feature construction using the coef function to calculate the risk score (risk score $= \Sigma$(coefficient$_i$×expression of signature gene$_i$)). Based on this score, risk values were assigned to each HCC sample to construct a risk model. The median risk value was used as the median value to classify the samples into high- and low-risk samples. Survival analysis was then performed on the high- and low-risk samples using the "survival" package. The calibration curves for survival period of respective 1 year, 3 year and 5 years were drawn using "survival", "regplot" v1.1 and "rms" v6.7-1 packages, with coxph used as the function. The generated candidate genes were subjected to differential analysis using "limma" v3.54.2 and "ggpubr" v0.6.0 packages, and "wilcox.test".

GSEA and ssGSEA analysis

The GSEA pathway enrichment analysis was performed using the "limma" package and the "clusterProfiler" package for the high- and low-risk samples and the configuration files of "c2.cp.kegg.Hs.symbols.gmt" downloaded from GSEA database, during which the GSEA function was used. The "limma", "GSEABase" v1.60.0 and "GSVA" v1.46.0 packages [28, 29] were used for enrichment analysis for the high- and low-risk samples. The exact configuration file "c2.cp.kegg.Hs.symbols.gmt" was used for the ssGSEA enrichment analysis. The ssGSEA scores of the relevant pathways in the samples were obtained using the gsva function and adjusted using the "normalize" function. The correlation between the risk scores of the sample and the ssGSEA scores of the signaling pathways in the risk model was carried out using the "spearman" method.

Cancer cell line encyclopedia (CCLE) analysis

Sample expression files were downloaded from the CCLE database (https://sites.broadinstitute.org/ccle). The database includes expression files for both HCC samples and fibroblast samples. The expression data of the candidate genes in the fibroblast and HCC samples were extracted from the database and the expressions of candidate genes in fibroblast and HCC samples were analyzed using "limma" and "ggpubr" packages. The differential analysis was performed using "limma", "ggpubr" packages and comparisons function, with the default method for analysis and p-value correction.

Tumor Immune Dysfunction and Exclusion (TIDE) prediction

The TIDE database (http://tide.dfci.harvard.edu/) was used to score HCC samples from TCGA to obtain the immunotherapy score for each sample, with the higher score for better immune escape ability and worse immunotherapy efficiency. During the analysis, the tumor type was selected as other due to the absence of a specific HCC category in the database, and the previous immunotherapy type was chosen as no. The file data were uploaded using the HCC expression data downloaded from TCGA database, and the data were log2 processed to meet the data type required by the database. Subsequently, the differences in immunotherapy scores between the high- and low-risk samples were analyzed using the "wilcox.test" method. The analysis of differences in immunotherapy response in the high- and low-risk samples was carried out using "Chi-squared Test ".

Drug sensitivity analysis

The Genomics of Drug Sensitivity in Cancer (GDSC) database (https://www.cancerrxgene.org/) was used to download tumor expression data and drug data, based on which "oncoPredict" v0.2 package was used to analyze drug sensitivity [30]. The GDSC data were used as the train group and the risk model data as the test group, and the batch correction method was "eb ". A fluctuation deletion threshold of 0.2 and a minimum sample size of 10 were used to obtain the drug sensitivity analysis results and the sensitivity scores for each drug for each sample. Subsequently, "limma" and "ggpubr" package were used to analyze the differences in drug sensitivity between the high- and low- risk samples in the risk model using "wilcox. test".

RESULTS

CAF scoring and survival analysis in HCC

CAF scoring was performed on HCC expression data using the MCPcounter algorithm to obtain a fibroblast score for each tumor (Supplementary Tables 1 and 2). Subsequently, survival analysis was performed based on the CAF scoring. The results showed that among HCC data in TCGA, the survival rate of high CAF-scoring samples was lower than that of the low CAF-scoring samples, but this comparison showed no significant difference (Fig. 1A). Meanwhile, survival analysis of the CAF scoring of the tumor samples from GSE76427 downloaded from the GEO database also showed that the survival rate of the high CAF scoring samples was lower than that of the low CAF scoring samples (Fig. 1B), although no statistically significant difference was detected. Those results imply that CAFs may influence the survival of HCC patients indirectly.

WGCNA analysis

To further screen HCC-associated CAFs, WGCNA analysis was performed on TCGA data and GSE76427 data. WGCNA analysis of TCGA data showed that all genes were classified into 8 different colored modules (Fig. 2A). As CAFs were usually expressed in the cell stroma, these modules were analyzed for correlation with CAF scoring and StromalScore (Fig. 2B) and the results revealed that the green module was highly correlated with CAF and StromalScore, implying that genes in this module may be closely associated with HCC associated CAFs. Similarly, WGCNA analysis of GSE76427 revealed that the genes in this chip were divided into 5 different colored modules (Fig. 2C), and correlation analysis showed a significant positive correlation between the blue
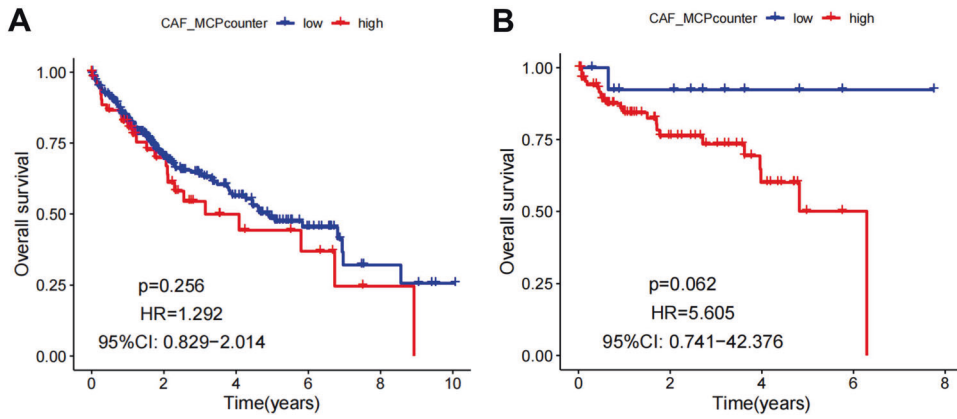
Fig. 1 Overall survival based on CAF scoring in TCGA and GSE databases using MCPcounter. A Survival rate in TCGA database; **B** Survival rate in GEO data, red lines in the graph are high CAF scoring and blue are low CAF scoring.
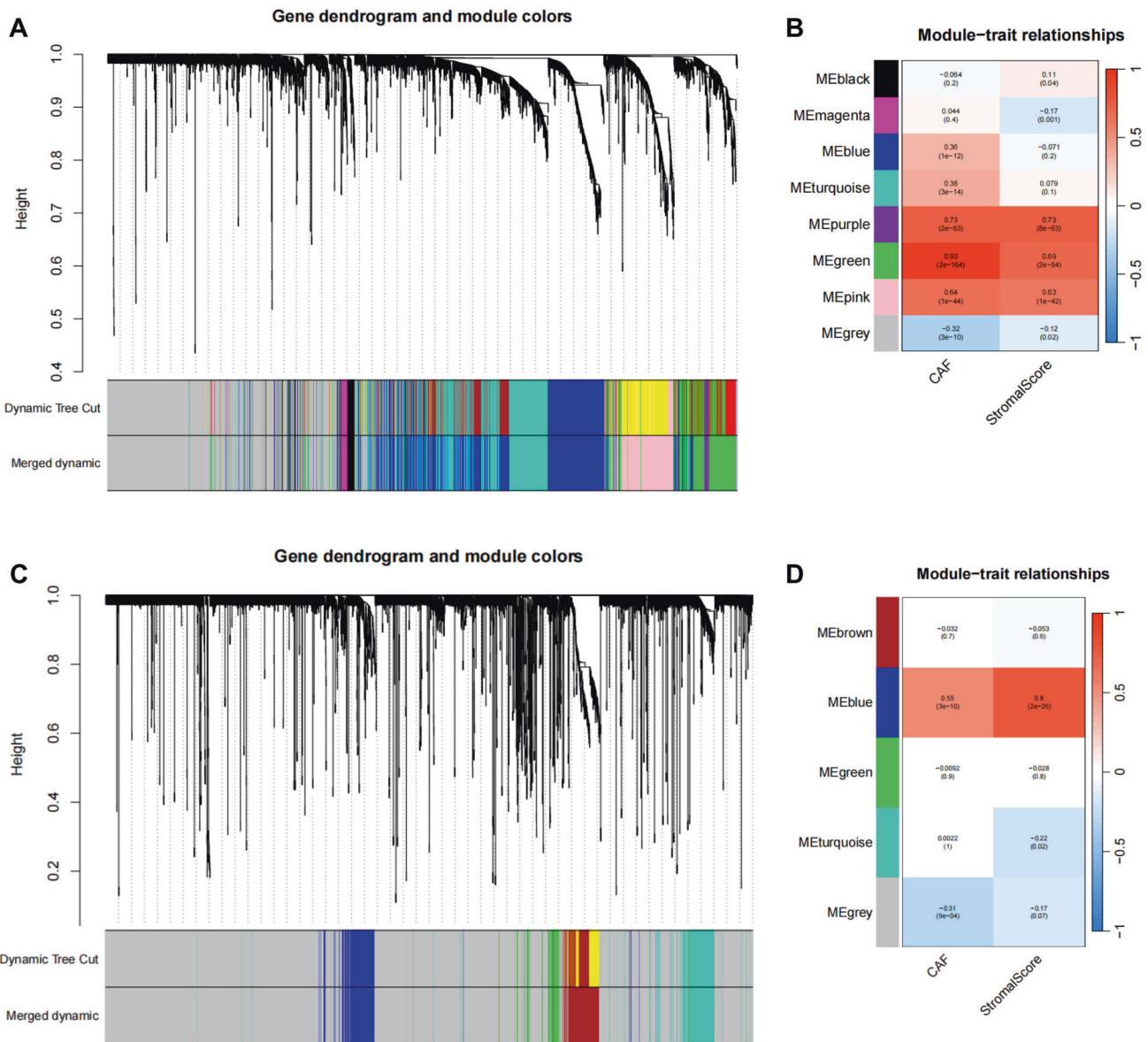

Fig. 2 WGCNA analyses of CAF-related genes. A WGCNA analysis of TCGA data, graph A shows gene module clustering, the upper part shows clustering before module merging, the lower part shows clustering after module clustering; **B** Correlation analysis between module and CAF scoring, the horizontal coordinates indicate CAF scoring and StromalScore respectively, the vertical coordinates indicate gene modules, the right histogram is the color scale; **C, D** WGCNA analysis of the GEO database.
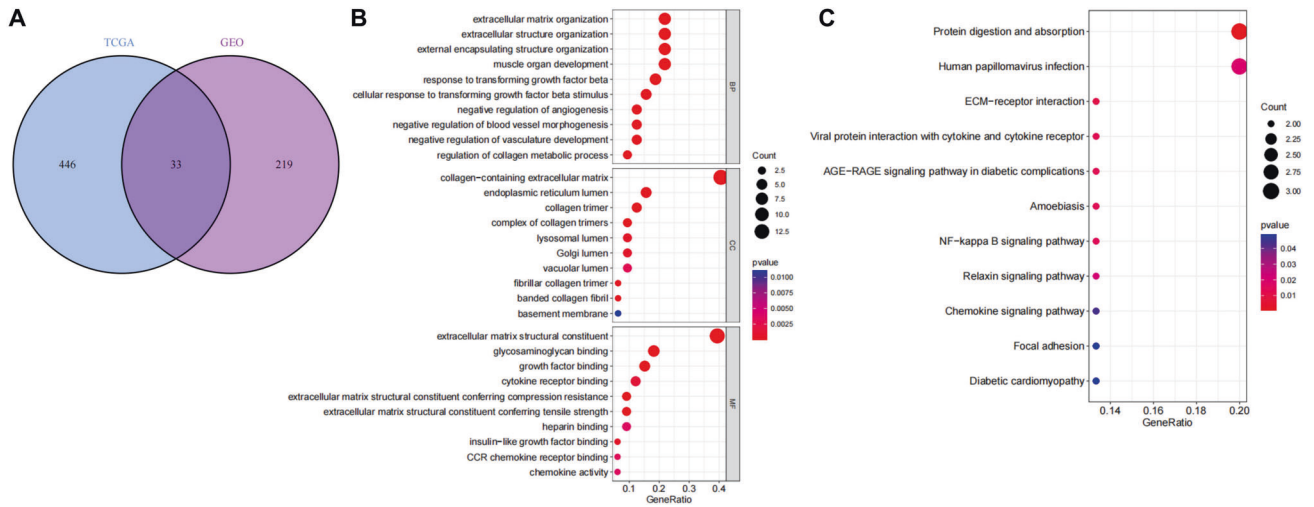
Y. Yao et al.

Fig. 3 Screening and functional enrichment of CAF candidate genes. A Gene intersections in modules significantly associated with CAFs of TCGA and GEO, with intersecting genes in the middle part; **B** GO functional enrichment analysis of intersecting genes, with horizontal coordinates indicating GeneRatio, vertical coordinates indicating functional entries, and color scale in the right histogram; **C** KEGG pathway of intersecting genes enrichment analysis.
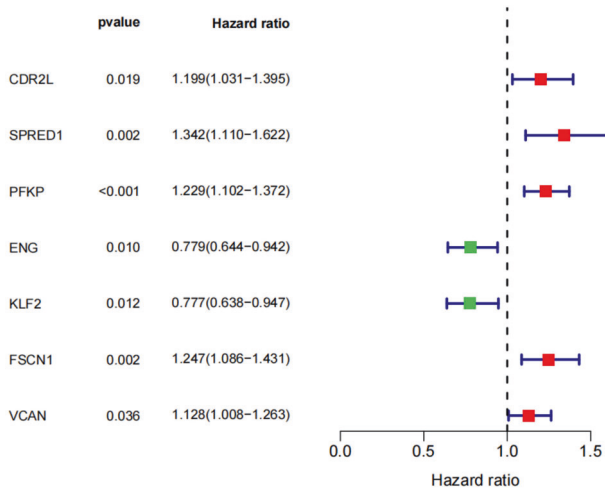

**Fig. 4 One-way cox analysis of candidate genes, which shows genes significantly associated with survival rate of HCC patients after one-way cox analysis.

module and CAF and StromalScore (Fig. 2D), implying that the genes in the blue module may be closely associated with CAFs.

Functional enrichment on candidate genes
The genes in the significantly correlated modules of TCGA and GSE76427 were extracted and the two sets of gene data were intersected (Fig. 3A), and 33 candidate genes were finally obtained. These 33 candidate genes were further analyzed for GO and KEGG functional enrichment, and were found to be enriched mainly under the functional categories of "extracellular matrix organization" and "protein digestion and absorption" (Fig. 3B, C), indicating those functional categories may be associated with the formation or changes of CAFs in HCC. In the signaling pathway analysis, we found that these genes were also significantly enriched in the "AGE-RAGE signaling pathway in diabetic complications" and "NF-kappa B signaling pathway". Those results suggest that alternation in CAF levels may influence these pathways and thus affect the development of HCC.

One-way cox analysis of candidate genes
One-way cox analysis was performed on these 33 candidate genes (Fig. 4, Supplementary Table 3), and the results revealed that 7 genes, including CDR2L, were significantly associated with prognostic survival of HCC patients, suggesting these 7 genes may be related to CAFs in HCC and may play an important role in the prognosis of HCC patients.

LASSO regression risk models
The LASSO regression risk model was constructed based on the 7 genes, CDR2L, SPRED1, PFKP, ENG, KLF2, FSCN1 and VCAN. Further survival analysis of the risk model constructed by LASSO regression (Fig. 5C) revealed that the survival rate in the high-risk samples was significantly lower than that of the low-risk samples, indicating that the 7 genes were not only closely associated with CAFs but also played important roles in the prognosis of HCC patients. Based on the data in the line chart and calibration curves (Fig. 5D, E), there were significant difference in patient risk and stage, which indicates this risk model is of high accuracy to predicting the prognosis of patients. Then the 7 candidate genes were analyzed for differential expression in HCC and normal samples (Fig. 5F), which identified CDR2L, PFKP, FSCN1 and VCAN were highly expressed in HCC samples, whereas the expressions of SPRED1, ENG and KLF2 in HCC samples showed no significant difference with that in normal samples. Those results suggested CDR2L, PFKP, FSCN1 and VCAN may regulate HCC more directly, while SPRED1, ENG and KLF2 may mediate HCC progression indirectly.

Expression analysis of model genes and CAF marker genes
CAF marker genes were obtained from published literature [31, 32]. Subsequently, the expression of genes screened for modeling as well as CAF marker genes in the risk model were further analyzed and the heat map was plotted (Fig. 6A). The results showed that most of the CAF marker genes had higher expression levels in the high-risk samples, while the model genes CDR2L, SPRED1, PFKP, FSCN1 and VCAN also had higher expression levels in the high-risk samples, which was consistent with the above results. The positive correlation between the above 7 model genes and CAF marker genes was found (Fig. 6B), suggesting that the expressions of these 7 genes might be closely related to CAF levels in HCC.
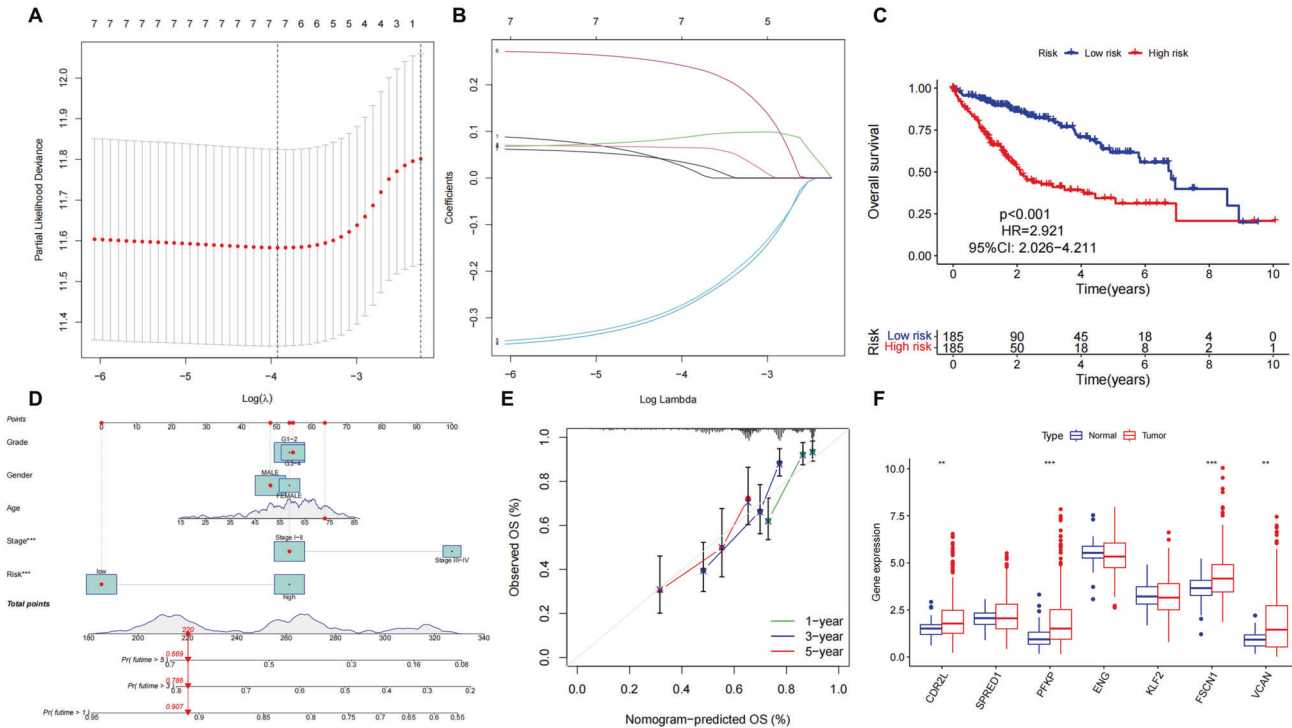
Fig. 5 LASOO regression analysis and risk model constructions of the 7 candidate genes. A, **B** LASOO regression of risk genes obtained from one-way cox analysis; **C** Survival analysis of risk model constructed from LASSO regression, red lines in the graph are the high risk samples and blue lines are the low risk samples. **D** Line chart for risk model; **E** Calibration curves for survival period of respective 1 year, 3 year and 5 years; **F** Differential expression of candidate genes, red box in the graph is HCC samples and blue box is normal sample ($N = 50$, $T = 374$) (*p.val < 0.05, **p.val < 0.01, ***p.val < 0.001).
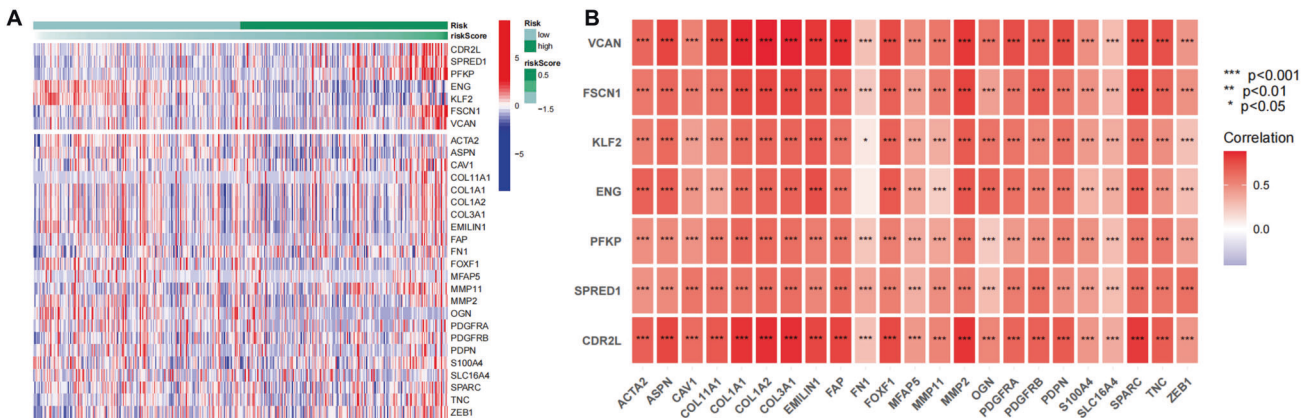

Fig. 6 Expressions of model genes and CAF marker genes. A The expressions of genes used to construct the risk model and the expression of CAF marker genes in the risk model, the vertical coordinates in the graph indicate the gene names, the top on the heat map is the model gene, the bottom of the heat map is the CAF marker genes, and the right histogram is the color scale; **B** Correlation analysis between model genes and CAF marker genes, the horizontal coordinates in the graph indicate the CAF marker gene, the vertical coordinate indicates the model gene, and the right histogram is the color scale.

Risk model GSEA analysis

To further understand the functions of genes in the risk model, GSEA pathway enrichment analysis was performed on high- and low-risk samples (Fig. 7A, B). The results showed that signaling pathways, such as CELL CYCLE, DNA REPLICATION were significantly enriched in the high-risk samples, while COMPLEMENT AND COAGULATION CASCADES, DRUG METABOLISM CYTOCHROME P450 were significantly enriched in the low-risk samples. These pathways may play an important regulatory function in HCC. Meanwhile, ssGSEA analysis was performed on the risk model and the risk values of each sample were calculated for correlation analysis with the pathways obtained from ssGSEA analysis

(Fig. 7C, D, Supplementary Figs. 1–6). The results found a significant correlation between the risk scores of the samples and the signaling pathways such as CELL CYCLE, which implies that CAF most likely regulates HCC progression by indirectly mediating the activation of these pathways.

CCLE expression analysis

To further understand the expression of the 7 candidate genes in HCC samples and CAF samples, the differential expressions of CDR2L, SPRED1, PFKP, ENG, KLF2, FSCN1 and VCAN were analyzed using CCLE database (Fig. 8A, B). The results showed that expression levels of these 7 genes were all
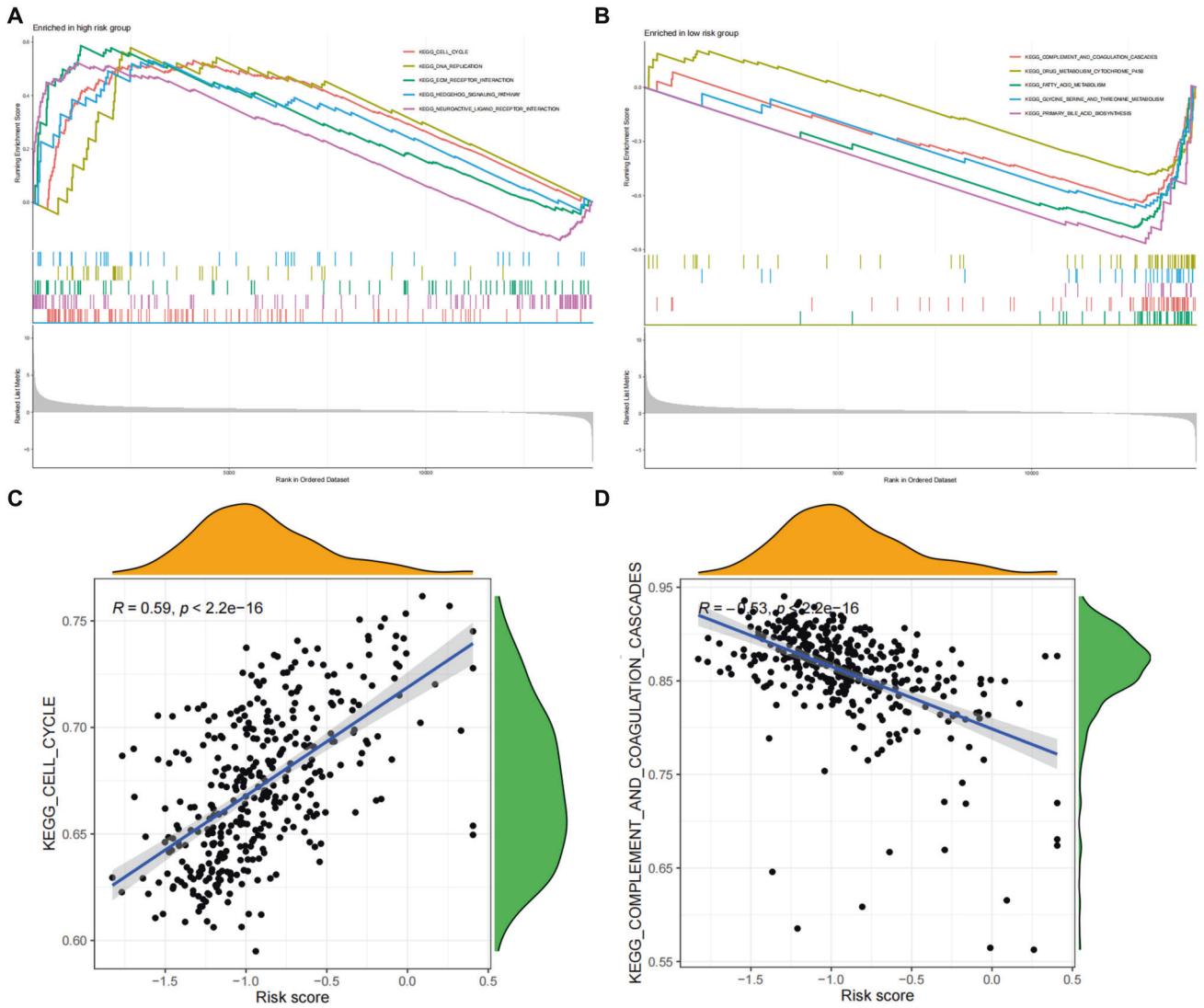
Fig. 7 GESA analysis of the risk model. A, **B** GSEA enrichment for part of significant pathways in the high- and low-risk samples, respectively; **C**, **D** Correlation analysis between the enriched pathway scores obtained from GSEA pathway enrichment and the risk values in the risk model.
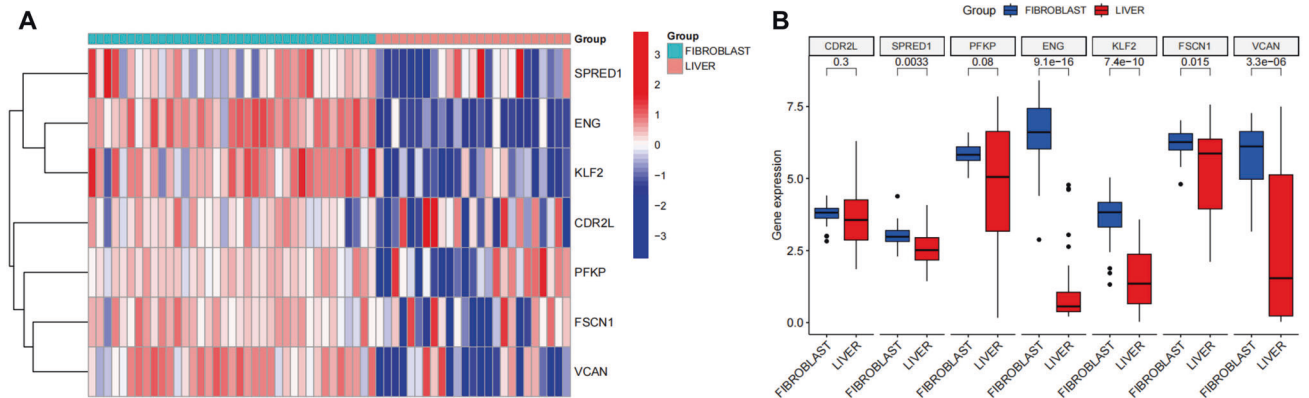

Fig. 8 Candidate gene expression analysis in CCLE. A Heat map of candidate gene expression in CAFs and HCC samples in CCLE, with color scale on the right histogram; **B** Box plot of candidate gene expression in CAFs and HCC in CCLE.

down-regulated in the HCC samples compared to CAF cells, among which 6 genes, SPRED1, PFKP, ENG, KLF2, FSCN1 and VCAN, were significantly down-regulated in HCC. Those results further suggest that these 7 genes are closely related to CAF levels in HCC, and their expression levels may be substantially associated with CAF. Therefore, it is speculated that CAFs regulate HCC progression most likely through regulating the expressions of these 7 genes.
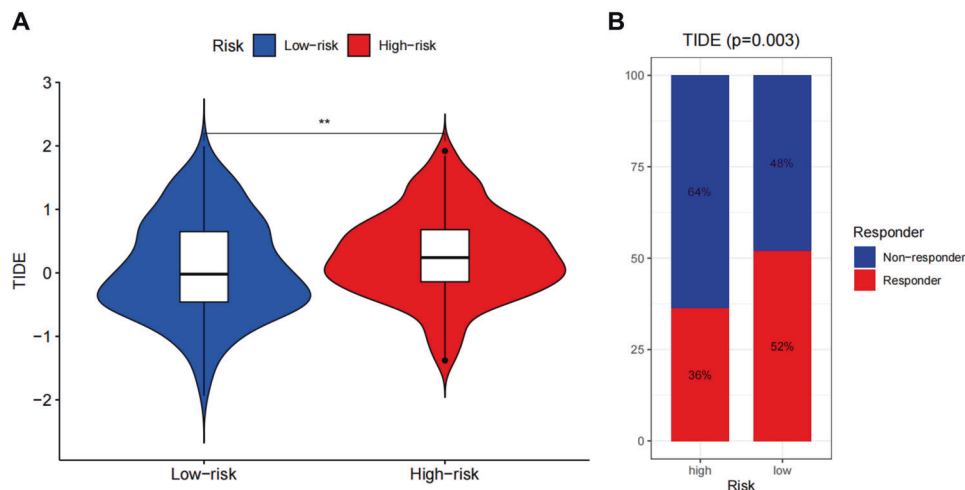
Fig. 9 TIDE analysis on risk models. A Differential analysis in TIDE scores in the risk model, horizontal coordinates indicate TIDE scores, vertical coordinates indicate risk model groupings, the higher the TIDE score, the greater the immune escape capacity of that sample; **B** Proportion of samples responding to immunotherapy in the risk model.

Risk model immunotherapy

To further understand the sensitivity of the risk model to immunotherapy, the samples in TCGA database were scored using TIDE scoring to obtain the immunotherapy scoring (Supplementary Fig. 7 and Supplementary Table 4). Subsequently, a differential analysis of immunotherapy scoring between the high- and low-risk samples in the risk model (Fig. 9A) was performed and showed that the TIDE scores were significantly higher in the high-risk samples than those in the low-risk samples, indicating that the high-risk samples had a higher potential for immune escape and weaker immunotherapy efficiency. The proportion of samples responding to immunotherapy in the high- and low-risk samples was counted (Fig. 9B) and the proportion of non-responders was higher in the high-risk samples than in the responders. In contrast, the opposite expression pattern was found in the low-risk samples. This result suggests that these 7 CAF-related genes, which were used to construct the risk model, may be associated with immunotherapy for HCC, highlighting the potential of CAF on immunotherapy in HCC.

Drug sensitivity prediction on risk models

To further screen the drugs for HCC, drug sensitivities in the risk model were predicted and differences in drug sensitivities between the high- and low-risk samples were analyzed. The results showed that 134 drugs showed significant differences in sensitivity between the high- and low-risk samples (Supplementary Table 5), among which PLX-4720, PF-4708671, JAK1_8709 and AZD2014, had the highest statistical differences (Fig. 10A–D), implying that these drugs had better therapeutic effects in the CAF-related gene constructed HCC risk model. Additionally, drug sensitivity analysis demonstrated that all four drugs show better sensitivity in the high-risk samples, suggesting that these 4 drugs may be more effective in treating CAF-associated HCC.

DISCUSSION

CAFs were reported for their regulation of tumor growth, angiogenesis and metastatic behavior in HCC [16, 33–35]. CAFs can promote metastasis in HCC through chemokines such as CXCL11, and CAF-carried proteins, such as CCL2 and CCL7 have been shown to enhance HCC progression [13, 36–38]. The clinical analysis found CAF abundance is often associated with poor clinical outcomes and is considered an attractive therapeutic target in primary HCC [39]. The bioinformatics method provides clues in predicting CAF-related genes and prognosis in HCC. Dong

et al. found that TOP2A might be closely related to CAF in HCC cancer and affect the prognosis of patients with HCC [40]. Song et al. predicted the prognostic value of CAF in HCC using XCELL calculation and found that the CAF-related risk model constructed based on genes such as LAMB1 is of great potential in predicting the prognosis of patients with HCC [41]. In addition, Yu et al. constructed a CAF-associated risk model in HCC for predicting the prognosis based on single-cell sequencing data and bulk RNA-seq data [42]. Nevertheless, there are few studies identifying the CAF-related genes in HCC, and there is still a paucity of studies exploring the relevant mechanisms of CAF in HCC prognosis. In this study, we obtained CAF scores from HCC data and GSE76427 in TCGA by using the MCPcounter algorithm and found that HCC patients with higher CAF scores had worse survival rates based on the data from TCGA data and the GSE76427. However, CAF scores were not significantly associated with prognostic survival of HCC patients, but in other tumors, such as gastric cancer, CAF abundance was found to be strongly associated with prognosis [43, 44]. This inconsistency may be explained by data limitations, or on the one hand, and suggests that CAF may not directly and dramatically affect the survival rate in HCC patients but in an indirect way.

To understand the regulatory mechanisms associated with CAFs in HCC, CAF-related genes in TCGA and GSE76427 data were predicted by the WGCNA method, which is the first time to search for CAF-related genes in HCC by the WGCNA analysis. A total of 33 shared CAF-associated genes were predicted in the two datasets. Signaling enrichment analysis of these associated genes showed that they were enriched in the "AGE-RAGE signaling pathway in diabetic complications" and the "NF-kappa B signaling pathway". Consistently, several studies showed that these signaling pathways are significantly involved in the development of HCC [45–49], which further confirms that CAFs may indirectly influence the development of HCC through these CAF-related genes.

Risk models for HCC were constructed based on CDR2L, SPRED1, PFKP, ENG, KLF2, FSCN1 and VCAN genes using one-way cox analysis and LASSO regression algorithms. Further analysis of the expression of these 7 CAF-related genes showed that their expression levels in HCC samples were significantly correlated with the expression levels of CAF marker genes in HCC samples, which further proved the association of 7 CAF-related genes with CAF abundance in HCC. The decreased expression of SPRED1 can substantially influence HCC progression [50]. The study by Tan et al. showed miR-126-3p can affect the sensitivity of HCC cells to Sorafenib by regulating SPRED1 [51]. In addition,
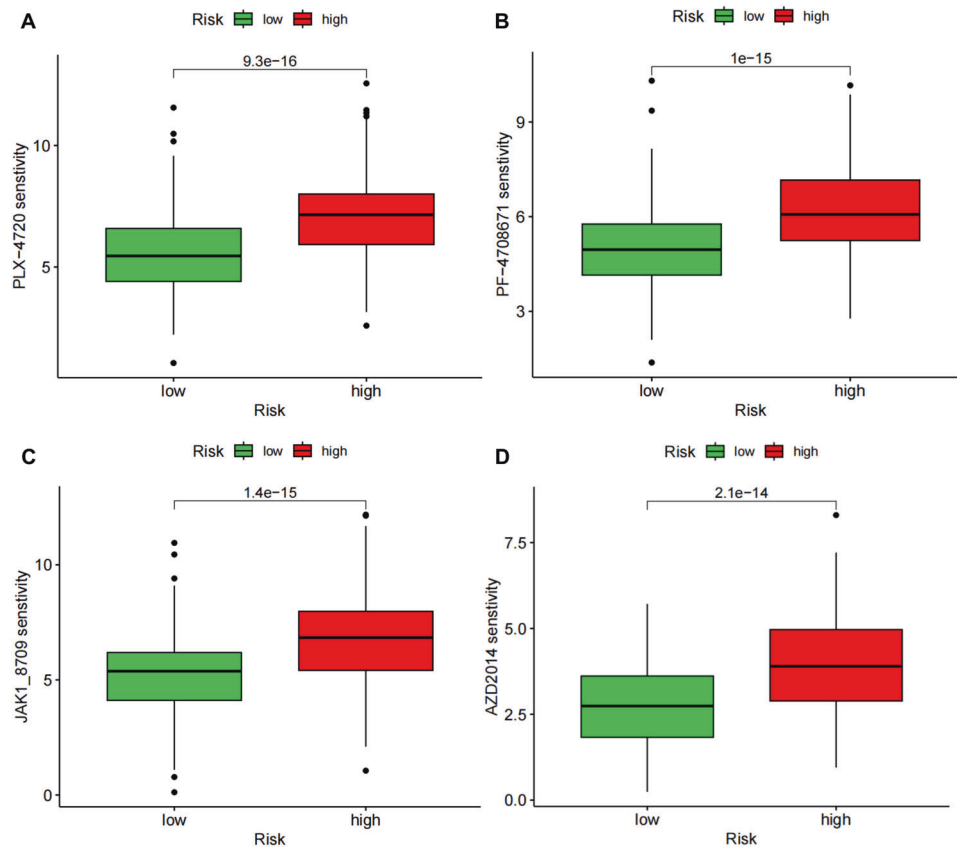
Fig. 10 Drug sensitivity analysis using risk models. A–D: Sensitivity analysis of 4 different drugs in the risk model. Vertical coordinates in the graph indicate drug sensitivity, horizontal coordinates indicate sample subgroups, and *p*-values for differences are shown above.

SPRED1 is thought to be a regulator in several tumors, including gastric cancer and breast cancer [52–54]. SPFKP was found to affect cell stemness of HCC cells [55], whereas in lung cancer, PFKP was found to regulate glucose metabolism [56]. In leukemia, PFKP affected disease progression by influencing CXCR4-dependent T-cell infiltration [57].

However, there are limited studies that have researched the regulatory mechanisms of PFKP in HCC. In the study of Li et al., it was pointed out that KLF2 can inhibit TGF-β-mediated cancer cell motility, thus affecting HCC development [58]. Moreover, KLF2 was reported as a promising marker in HCC and could be used as a prognostic marker associated with fibrosis and immune infiltration for advanced HCC [59]. Accumulating evidence highlighted that KLF2 may be regulated by lncRNAs such as FBXL10-AS1 and GHET1, thus affecting HCC progression [60–64], and that KLF2 can mediate HCC progression through c-myc [65]. It was also noted that miRNAs such as miR-145, miR-133a, and miR-539 could influence proliferation, migration, and invasion ability of HCC cells by targeting FSCN1 [66–69]. LncRNA ADORA2A-AS1 was found to modulate the FSCN1/AKT axis to regulate HCC progression [70]. FSCN1 influenced adriamycin resistance in HCC by modulating epithelial-mesenchymal transition [71]. In the study by Hayashi et al., FSCN1 was also found to modulate the expression of E-cadherin in HCC [72]. The implication of VCAN in HCC was reported in a previous study, which demonstrated that VCAN could affect the proliferation and metastasis of HCC cells through EGFR-PI3K-AKT pathway [73]. Additionally, VCAN protein secreted by CAFs can increase the cellular malignant transformation of HCC. Therefore VCAN was considered to have the potential to serve as a new therapeutic target in HCC [74]. Although less research reported the mechanism of CDR2L and ENG in HCC, CDR2L was found to be a hub gene in paraneoplastic cerebellar

degeneration [75, 76]. CDR2L has also been suggested as a regulatory gene in ovarian cancer [77]. ENG, as an endothelial glycoprotein gene, has been known for its essential role in many tumors with the potential as a target for tumor therapy [78–80]. The expression level of CAF-related ENG was found to influence the invasion and metastasis of rectal cancer cells, as reported by Paauwe et al. [81]. Although the regulatory roles of these genes have been reported in different tumors, however, few studies reported their role in HCC, in HCC related CAFs. In the present study, HCC expression data were analyzed to search for CAF-related genes in HCC. Further expression analysis of these 7 model genes in the CCLE showed that they exhibited significantly elevated expression levels in fibroblasts. TIDE online algorithm in the risk model found that the low-risk samples were closely associated with improved immunotherapy efficiency in HCC patients. Drug sensitivity analysis demonstrated that the high-risk samples were highly sensitive to drugs such as PLX-4720, PF-4708671, JAK1_8709 and AZD2014. Those analyses also provide new research ideas for HCC treatment. In this study, several new CAF-related genes were obtained using different CAF score calculation methods and data from different sources instead of those already reported. The risk model constructed shows high predictive ability for the prognosis of patients with HCC, and therefore, provides a broader idea and direction for further understanding the CAF-related regulatory mechanisms in HCC.

Although this study provides a new direction for further understanding CAF-related genes and their regulatory mechanisms in HCC, we should still be aware of some limitations. First, this is a retrospective bioinformatics analysis based on public expression data, and thus the prognostic and therapeutic value of the model constructed based on these 7 CAF-related genes should be cross-validated in more data. In addition, the specific

biological role of CAF-related genes in HCC and their association with CAFs should be further clarified through molecular and animal experiments. Nevertheless, our results can serve as a theoretical basis for future studies on CAF-related studies in HCC.

CONCLUSION

In conclusion, a comprehensive co-expression network focusing on CAF-related genes in HCC was constructed. Risk models constructed based on CAF-related genes, CDR2L, SPRED1, PFKP, ENG, KLF2, FSCN1 and VCAN were able to predict their association with prognosis, chemotherapy and immunotherapy response of HCC patients. This study provides new insights into CAF-related regulation and therapeutic strategies for HCC. Clinical validation would further added the credibility of this study and future studies with clinical data are encouraged to further validate the association of CDR2L, SPRED1, PFKP, ENG, KLF2, FSCN1 and VCAN with prognosis, chemotherapy and immunotherapy response of HCC patients.

DATA AVAILABILITY

All data are uploaded as supplementary materials.

REFERENCES

1. Liu J, Dang H, Wang XW. The significance of intertumor and intratumor heterogeneity in liver cancer. Exp Mol Med. 2018;50:e416.
2. Liu G, Yang ZF, Zhou PY, Zhou C, Guan RY, Sun BY, et al. ROR-alpha-1 inhibits the proliferation, invasion, and migration of hepatocellular carcinoma MHCC97H via downregulation of chemokine CXCL5. Cytokine. 2020;129:155004.
3. Uka K, Aikata H, Takaki S, Shirakawa H, Jeong SC, Yamashina K, et al. Clinical features and prognosis of patients with extrahepatic metastases from hepatocellular carcinoma. World J Gastroenterol. 2007;13:414–20.
4. Wang H, Chen L. Tumor microenviroment and hepatocellular carcinoma metastasis. J Gastroenterol Hepatol. 2013;28:43–8.
5. El-Serag HB. Hepatocellular carcinoma. N Engl J Med. 2011;365:1118–27.
6. Fattovich G, Stroffolini T, Zagni I, Donato F. Hepatocellular carcinoma in cirrhosis: incidence and risk factors. Gastroenterology. 2004;127:S35–50.
7. Baglieri J, Brenner DA, Kisseleva T. The role of fibrosis and liver-associated fibroblasts in the pathogenesis of hepatocellular carcinoma. Int J Mol Sci. 2019;20:1723.
8. Friedman SL. Molecular regulation of hepatic fibrosis, an integrated cellular response to tissue injury. J Biol Chem. 2000;275:2247–50.
9. Friedman SL. Mechanisms of hepatic fibrogenesis. Gastroenterology. 2008;134:1655–69.
10. Friedman SL. Hepatic stellate cells: protean, multifunctional, and enigmatic cells of the liver. Physiol Rev. 2008;88:125–72.
11. Lee UE, Friedman SL. Mechanisms of hepatic fibrogenesis. Best Pr Res Clin Gastroenterol. 2011;25:195–206.
12. Lin ZY, Chuang WL. Hepatocellular carcinoma cells cause different responses in expressions of cancer-promoting genes in different cancer-associated fibroblasts. Kaohsiung J Med Sci. 2013;29:312–8.
13. Liu J, Chen S, Wang W, Ning BF, Chen F, Shen W, et al. Cancer-associated fibroblasts promote hepatocellular carcinoma metastasis through chemokine-activated hedgehog and TGF-beta pathways. Cancer Lett. 2016;379:49–59.
14. Jiang J, Ye F, Yang X, Zong C, Gao L, Yang Y, et al. Peri-tumor associated fibroblasts promote intrahepatic metastasis of hepatocellular carcinoma by recruiting cancer stem cells. Cancer Lett. 2017;404:19–28.
15. Lin ZY, Chuang YH, Chuang WL. Cancer-associated fibroblasts up-regulate CCL2, CCL26, IL6 and LOXL2 genes related to promotion of cancer progression in hepatocellular carcinoma cells. Biomed Pharmacother. 2012;66:525–9.
16. Liu G, Sun J, Yang ZF, Zhou C, Zhou PY, Guan RY, et al. Cancer-associated fibroblast-derived CXCL11 modulates hepatocellular carcinoma cell migration and tumor metastasis through the circUBAP2/miR-4756/IFIT1/3 axis. Cell Death Dis. 2021;12:260.
17. Fang T, Lv H, Lv G, Li T, Wang C, Han Q, et al. Tumor-derived exosomal miR-1247-3p induces cancer-associated fibroblast activation to foster lung metastasis of liver cancer. Nat Commun. 2018;9:191.
18. Herrero A, Knetemann E, Mannaerts I. Review: challenges of in vitro CAF modelling in liver cancers. Cancers. 2021;13:5914.
19. Jia C, Wang G, Wang T, Fu B, Zhang Y, Huang L, et al. Cancer-associated fibroblasts induce epithelial-mesenchymal transition via the transglutaminase 2-dependent IL-6/IL6R/STAT3 axis in hepatocellular carcinoma. Int J Biol Sci. 2020;16:2542–58.
20. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinform. 2008;9:559.
21. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43:e47.
22. Becht E, Giraldo NA, Lacroix L, Buttard B, Elarouci N, Petitprez F, et al. Erratum to: estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. Genome Biol. 2016;17:249.
23. Becht E, Giraldo NA, Lacroix L, Buttard B, Elarouci N, Petitprez F, et al. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. Genome Biol. 2016;17:218.
24. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, et al. clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. Innovation. 2021;2:100141.
25. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 2012;16:284–7.
26. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Stat Softw. 2010;33:1–22.
27. Tay JK, Narasimhan B, Hastie T. Elastic net regularization paths for all generalized linear models. J Stat Softw. 2023;106:1.
28. Hanzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinform. 2013;14:7.
29. Zhang Z. Reshaping and aggregating data: an introduction to reshape package. Ann Transl Med. 2016;4:78.
30. Maeser D, Gruener RF, Huang RS. oncoPredict: an R package for predicting in vivo or cancer patient drug response and biomarkers from cell line screening data. Brief Bioinform. 2021;22:bbab260.
31. Gascard P, Tlsty TD. Carcinoma-associated fibroblasts: orchestrating the composition of malignancy. Genes Dev. 2016;30:1002–19.
32. Han C, Liu T, Yin R. Biomarkers for cancer-associated fibroblasts. Biomark Res. 2020;8:64.
33. Claperon A, Mergey M, Aoudjehane L, Ho-Bouldoires TH, Wendum D, Prignon A, et al. Hepatic myofibroblasts promote the progression of human cholangiocarcinoma through activation of epidermal growth factor receptor. Hepatology. 2013;58:2001–11.
34. Sun L, Wang Y, Wang L, Yao B, Chen T, Li Q, et al. Resolvin D1 prevents epithelial-mesenchymal transition and reduces the stemness features of hepatocellular carcinoma by inhibiting paracrine of cancer-associated fibroblast-derived COMP. J Exp Clin Cancer Res. 2019;38:170.
35. Xu H, Zhao J, Li J, Zhu Z, Cui Z, Liu R, et al. Cancer associated fibroblast-derived CCL5 promotes hepatocellular carcinoma metastasis through activating HIF1alpha/ZEB1 axis. Cell Death Dis. 2022;13:478.
36. Lau EY, Lo J, Cheng BY, Ma MK, Lee JM, Ng JK, et al. Cancer-associated fibroblasts regulate tumor-initiating cell plasticity in hepatocellular carcinoma through c-Met/FRA1/HEY1 signaling. Cell Rep. 2016;15:1175–89.
37. Loh JJ, Li TW, Zhou L, Wong TL, Liu X, Ma VWS, et al. FSTL1 secreted by activated fibroblasts promotes hepatocellular carcinoma metastasis and stemness. Cancer Res. 2021;81:5692–705.
38. Utispan K, Sonongbua J, Thuwajit P, Chau-In S, Pairojkul C, Wongkham S, et al. Periostin activates integrin alpha5beta1 through a PI3K/AKT-dependent pathway in invasion of cholangiocarcinoma. Int J Oncol. 2012;41:1110–8.
39. Ying F, Chan MSM, Lee TKW. Cancer-associated fibroblasts in hepatocellular carcinoma and cholangiocarcinoma. Cell Mol Gastroenterol Hepatol. 2023;15:985–99.
40. Dong W, Xie Y, Huang H. Prognostic value of cancer-associated fibroblast-related gene signatures in hepatocellular carcinoma. Front Endocrinol. 2022;13:884777.
41. Song L, Li Q, Lu Y, Feng X, Yang R, Wang S. Cancer progression mediated by CAFs relating to HCC and Identification of genetic characteristics influencing prognosis. J Oncol. 2022;2022:2495361.
42. Yu L, Shen N, Shi Y, Shi X, Fu X, Li S, et al. Characterization of cancer-related fibroblasts (CAF) in hepatocellular carcinoma and construction of CAF-based risk signature based on single-cell RNA-seq and bulk RNA-seq data. Front Immunol. 2022;13:1009789.
43. Zhang H, Hong Z, Li P, Jiang H, Wu P, Chen J. Identification and validation of an immune evasion molecular subgroup of patients with colon cancer for implications of immunotherapy. Front Genet. 2022;13:811660.
44. Zheng H, Liu H, Li H, Dou W, Wang X. Weighted gene co-expression network analysis identifies a cancer-associated fibroblast signature for predicting prognosis and therapeutic responses in gastric cancer. Front Mol Biosci. 2021;8:744677.

45. Hollenbach M. The role of glyoxalase-I (Glo-I), advanced glycation endproducts (AGEs), and their receptor (RAGE) in chronic liver disease and hepatocellular carcinoma (HCC). Int J Mol Sci. 2017;18:2466.

46. Liu P, Kimmoun E, Legrand A, Sauvanet A, Degott C, Lardeux B, et al. Activation of NF-kappa B, AP-1 and STAT transcription factors is a frequent and early event in human hepatocellular carcinomas. J Hepatol. 2002;37:63–71.

47. Meng F, Henson R, Patel T. Chemotherapeutic stress selectively activates NF-kappa B-dependent AKT and VEGF expression in liver cancer-derived endothelial cells. Am J Physiol Cell Physiol. 2007;293:C749–60.

48. Moy KA, Jiao L, Freedman ND, Weinstein SJ, Sinha R, Virtamo J, et al. Soluble receptor for advanced glycation end products and risk of liver cancer. Hepatology. 2013;57:2338–45.

49. Zhu D, Fang C, Yang Z, Ren Y, Yang F, Zheng S, et al. Tubulin-binding peptide RR-171 derived from human umbilical cord serum displays antitumor activity against hepatocellular carcinoma via inducing apoptosis and activating the NF-kappa B pathway. Cell Prolif. 2022;55:e13241.

50. Ji JS, Xu M, Song JJ, Zhao ZW, Chen MJ, Chen WQ, et al. Inhibition of microRNA-126 promotes the expression of Spred1 to inhibit angiogenesis in hepatocellular carcinoma after transcatheter arterial chemoembolization: in vivo study. Onco Targets Ther. 2016;9:4357–67.

51. Tan W, Lin Z, Chen X, Li W, Zhu S, Wei Y, et al. miR-126-3p contributes to sorafenib resistance in hepatocellular carcinoma via downregulating SPRED1. Ann Transl Med. 2021;9:38.

52. Ablain J, Liu S, Moriceau G, Lo RS, Zon LI. SPRED1 deletion confers resistance to MAPK inhibition in melanoma. J Exp Med. 2021;218:e20201097.

53. Jiang CF, Shi ZM, Li DM, Qian YC, Ren Y, Bai XM. et al. Estrogen-induced miR-196a elevation promotes tumor growth and metastasis via targeting SPRED1 in breast cancer. Mol Cancer. 2018;17:83

54. Liu W, Fang S, Zuo G. A study on the expression of SPRED1 and PBRM1 (Baf180) and their clinical significances in patients with gastric cancer. Clin Lab. 2020;66:10.

55. Sha X, Wang K, Wang F, Zhang C, Yang L, Zhu X. Silencing PFKP restrains the stemness of hepatocellular carcinoma cells. Exp Cell Res. 2021;407:112789.

56. Shen J, Jin Z, Lv H, Jin K, Jonas K, Zhu C, et al. PFKP is highly expressed in lung cancer and regulates glucose metabolism. Cell Oncol. 2020;43:617–29.

57. Gao X, Qin S, Wu Y, Chu C, Jiang B, Johnson RH, et al. Nuclear PFKP promotes CXCR4-dependent infiltration by T cell acute lymphoblastic leukemia. J Clin Invest. 2021;131:e143119.

58. Li Y, Tu S, Zeng Y, Zhang C, Deng T, Luo W, et al. KLF2 inhibits TGF-beta-mediated cancer cell motility in hepatocellular carcinoma. Acta Biochim Biophys Sin. 2020;52:485–94.

59. Chen XQ, Ma J, Xu D, Xiang ZL. Comprehensive analysis of KLF2 as a prognostic biomarker associated with fibrosis and immune infiltration in advanced hepatocellular carcinoma. BMC Bioinform. 2023;24:270.

60. Chen Y, Yang L. FBXL19-AS1 aggravates the progression of hepatocellular cancer by downregulating KLF2. J BUON. 2021;26:1333–9.

61. Jin L, He Y, Tang S, Huang S. LncRNA GHET1 predicts poor prognosis in hepatocellular carcinoma and promotes cell proliferation by silencing KLF2. J Cell Physiol. 2018;233:4726–34.

62. Huang MD, Chen WM, Qi FZ, Xia R, Sun M, Xu TP, et al. Long non-coding RNA ANRIL is upregulated in hepatocellular carcinoma and regulates cell apoptosis by epigenetic silencing of KLF2. J Hematol Oncol. 2015;8:50.

63. Huang MD, Chen WM, Qi FZ, Xia R, Sun M, Xu TP, et al. Long non-coding RNA ANRIL is upregulated in hepatocellular carcinoma and regulates cell proliferation by epigenetic silencing of KLF2. J Hematol Oncol. 2015;8:57.

64. Huang MD, Chen WM, Qi FZ, Sun M, Xu TP, Ma P, et al. Long non-coding RNA TUG1 is up-regulated in hepatocellular carcinoma and promotes cell growth and apoptosis by epigenetically silencing of KLF2. Mol Cancer. 2015;14:165.

65. Zou K, Lu X, Ye K, Wang C, You T, Chen J. Kruppel-like factor 2 promotes cell proliferation in hepatocellular carcinoma through up-regulation of c-myc. Cancer Biol Ther. 2016;17:20–6.

66. Abdallah HMI, El Awdan SA, Abdel-Rahman RF, Farrag ARH, Allam RM. 1,8 Cineole and Ellagic acid inhibit hepatocarcinogenesis via upregulation of MiR-122 and suppression of TGF-beta1, FSCN1, Vimentin, VEGF, and MMP-9. PLoS ONE. 2022;17:e0258998.

67. Wang G, Zhu S, Gu Y, Chen Q, Liu X, Fu H. MicroRNA-145 and MicroRNA-133a inhibited proliferation, migration, and invasion, while promoted apoptosis in hepatocellular carcinoma cells via targeting FSCN1. Dig Dis Sci. 2015;60:3044–52.

68. Zhao H, Kang X, Xia X, Wo L, Gu X, Hu Y, et al. miR-145 suppresses breast cancer cell migration by targeting FSCN-1 and inhibiting epithelial-mesenchymal transition. Am J Transl Res. 2016;8:3106–14.

69. Liu Y, Hong W, Zhou C, Jiang Z, Wang G, Wei G, et al. miR-539 inhibits FSCN1 expression and suppresses hepatocellular carcinoma migration and invasion. Oncol Rep. 2017;37:2593–602.

70. Pu J, Zhang Y, Wang A, Qin Z, Zhuo C, Li W, et al. ADORA2A-AS1 restricts hepatocellular carcinoma progression via binding HuR and repressing FSCN1/AKT axis. Front Oncol. 2021;11:754835.

71. Zhang Y, Lu Y, Zhang C, Huang D, Wu W, Zhang Y, et al. FSCN-1 increases doxorubicin resistance in hepatocellular carcinoma through promotion of epithelial-mesenchymal transition. Int J Oncol. 2018;52:1455–64.

72. Hayashi Y, Osanai M, Lee GH. Fascin-1 expression correlates with repression of E-cadherin expression in hepatocellular carcinoma cells and augments their invasiveness in combination with matrix metalloproteinases. Cancer Sci. 2011;102:1228–35.

73. Zhangyuan G, Wang F, Zhang H, Jiang R, Tao X, Yu D, et al. VersicanV1 promotes proliferation and metastasis of hepatocellular carcinoma through the activation of EGFR-PI3K-AKT pathway. Oncogene. 2020;39:1213–30.

74. Kato K, Fukai M, Hatanaka KC, Takasawa A, Aoyama T, Hayasaka T, et al. Versican secreted by cancer-associated fibroblasts is a poor prognostic factor in hepatocellular carcinoma. Ann Surg Oncol. 2022;29:7135–46.

75. Herdlevaer I, Krakenes T, Schubert M, Vedeler CA. Localization of CDR2L and CDR2 in paraneoplastic cerebellar degeneration. Ann Clin Transl Neurol. 2020;7:2231–42.

76. Krakenes T, Herdlevaer I, Raspotnig M, Haugen M, Schubert M, Vedeler CA. CDR2L Is the major Yo antibody target in paraneoplastic cerebellar degeneration. Ann Neurol. 2019;86:316–21.

77. Raspotnig M, Haugen M, Thorsteinsdottir M, Stefansson I, Salvesen HB, Storstein A, et al. Cerebellar degeneration-related proteins 2 and 2-like are present in ovarian cancer in patients with and without Yo antibodies. Cancer Immunol Immunother. 2017;66:1463–71.

78. Perez-Gomez E, Del Castillo G, Juan Francisco S, Lopez-Novoa JM, Bernabeu C, Quintanilla M. The role of the TGF-beta coreceptor endoglin in cancer. ScientificWorldJournal. 2010;10:2367–84.

79. Rosen LS, Gordon MS, Robert F, Matei DE. Endoglin for targeted cancer treatment. Curr Oncol Rep. 2014;16:365.

80. Seon BK, Haba A, Matsuno F, Takahashi N, Tsujie M, She X, et al. Endoglin-targeted cancer therapy. Curr Drug Deliv. 2011;8:135–43.

81. Paauwe M, Schoonderwoerd MJA, Helderman R, Harryvan TJ, Groenewoud A, van Pelt GW, et al. Endoglin expression on cancer-associated fibroblasts regulates invasion and stimulates colorectal cancer metastasis. Clin Cancer Res. 2018;24:6331–44.

AUTHOR CONTRIBUTIONS

CT, BF and YY contributed to the conception and design; KQY and QW, ZZ, SL and BL did the acquisition of data and data analysis; ZZ drafted the article; SL and BL did the revision of the article. All authors approved the final version of manuscript.

FUNDING

COMPETING INTERESTS

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not applicable.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41435-024-00252-z.

Correspondence and requests for materials should be addressed to Bin Feng or Caixi Tang.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.