Spatial Maps of Hepatocellular Carcinoma Transcriptomes Highlight an Unexplored Landscape of Heterogeneity and a Novel Gene Signature for Survival

Nan Zhao  
Tianjin Medical University

Yanhui Zhang  
Tianjin Medical University

Runfen Cheng  
Tianjin Medical University

Danfang Zhang  
Tianjin Medical University

Fan Li  
Tianjin Medical University

Yuhong Guo  
Tianjin Medical University

Zhiqiang Qiu  
Tianjin Medical University

Xueyi Dong  
Tianjin Medical University

Xinchao Ban  
Tianjin Medical University

Baocun Sun  
Tianjin Medical University

xiulan Zhao  (zhaoxiulan@tmu.edu.cn)  
Tianjin Medical University  https://orcid.org/0000-0002-6333-8717

Research Article

Keywords: HCC, heterogeneity, spatial transcriptomics (ST), satellite nodules, gene signature

Posted Date: August 25th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-812356/v1
Abstract

Background: Hepatocellular carcinoma (HCC) are often present with satellite nodules, rendering current curative treatments ineffective in many patients. The heterogeneity of HCC is a major challenge in personalized medicine. The emergence of spatial transcriptomics (ST) provides a powerful strategy for delineating the complex molecular landscapes of tumors.

Methods: In this study, we investigated tissue-wide gene expression heterogeneity in tumor and adjacent nonneoplastic tissues using ST technology. We analyzed the transcriptomes of nearly 10820 tissue regions and identified main gene expression clusters and their specific marker genes (differentially expressed genes, DEGs) in patients. The DEGs were analyzed from two perspectives. First of all, we identified two distinct gene profiles associated with satellite nodules and conducted a more comprehensive analysis for both gene profiles. Their clinical relevance for human HCC was validated with KM Plotter. Secondly, we screened DEGs with TCGA database to divide the HCC cohort into high- and low-risk groups according to Cox analysis. HCC patients from the ICGC cohort were used for validation. Kaplan Meier analysis was used to compare the overall survival (OS) between high- and low-risk groups. Univariate and multivariate Cox analyses were applied to determine the independent predictors for OS.

Results: Novel markers for the prediction of satellite nodules and a tumor clusters specific marker genes signature model(6 genes) for HCC prognosis was constructed, respectively.

Conclusion: The establishment of marker gene profiles may be an important step towards an unbiased view of HCC and the 6-genes signature can be used for prognostic prediction in HCC. This analysis will help us to clarify one of the possible sources of HCC heterogeneity, uncover pathogenic mechanisms and novel anti-tumor drug targets.

Introduction

Globally, the mortality rate of primary liver cancer ranked fourth in cancer deaths. In many countries, its 5-year survival rate is less than 20% without significant changes in survival rate over time[1–3]. Hepatocellular carcinoma (HCC) is the most common histological type accounting for the highest proportion (75–85%) of primary liver cancer[1]. Recently, progression is considered to be the most important reason of the poor prognosis of patients with HCC, identifying the possible factors affecting the progression of HCC and exploring potential interventional therapies will bring better prognosis for HCC patients[4].

HCC is an extraordinarily heterogeneous malignant disease considering the tumors that have so far been identified. During personalized treatment of tumor patients, intra- and inter-tumor heterogeneity has become a great challenge since it may directly change the predicted biological markers of diagnosis, prognosis and therapy. Even for the identical histological of tumor, variation in the expression of biomarkers among different patients and between different tumor areas of the same individual sample such as tumor and peritumor should be considered seriously.
In HCC, the existence of microsatellite nodules is a well-known risk factor. However, such nodules usually cannot be detected on imaging modalities\[^{[5]}\]. These undetectable satellite nodules by imaging may be risk factors for local recurrence since they are usually ignored by local ablation therapy. Therefore, analyze the risk factors associated with satellite nodules is very important in order to improve the treatment strategies of patients at high risk. Characterization of the tumoral heterogeneity of HCC with satellite nodules using transcriptomic analysis may be important to reduce the incidence of local recurrence.

In this study, we analyzed clusters defined by ST related specific marker genes (DEGs) from two perspectives. We demonstrated the feasibility and merit of using spatial transcriptomics (ST) to dissect inter- and intratumoral heterogeneity across HCC patient specimens with or without satellite nodules. In addition, we constructed a prognostic signature with DEGs in the TCGA cohort and validated the stability and reliability of the model through the ICGC cohort. Assessing the pathogenesis of tumors by ST might provide a better view of the landscape of HCC heterogeneity to facilitate personalized management and ST clusters related genes signature can provide novel anti-tumor drug targets for HCC.

**Materials And Methods**

**Patient Sample**

Tissue samples were obtained from 3 patients who underwent hepatectomy because of HCC in August 2020 from the Tumor Tissue Bank of Tianjin Cancer Hospital. Pathologists have confirmed all sample’s diagnoses. Detailed pathologic and clinical data were listed in Supplementary Table 1. The use of these tissue samples in this study was approved by the institutional research committee.

**Slide preparation**

There are six capture areas (3 tumor area and 3 peritumor area) (6.5×6.5 mm) of Spatial Transcriptomics slides, each with 4999 capture spots of barcoded primers (10×Genomics). The diameter of the spots is 100\(\mu\)m. These spots are arranged in a rectangular shape. Each spot contains millions of oligonucleotides with the following features: a 30 nucleotide poly(dT) sequence for the capture of polyadenylated mRNA molecules, a 12 nucleotide unique molecular identifier (UMI) for the identification of duplicate molecules that arise during the library preparation and sequencing process, a 16 nucleotide Spatial Barcode, which is shared by all oligonucleotides within each individual gene expression spot, and a partial TruSeq Read 1 sequence for use during the library preparation and sequencing portions of the workflow.

**Tissue preparation, fixation and staining**

The 10×Visium protocol was optimized for frozen tissue. Briefly, tumors were frozen in dry ice immediately. Tumors were embedded with OCT and were cryosectioned at 10 \(\mu\)m thickness. We placed the sections on the capture areas and incubated them at 37°C for 1 min, and then fixed them in methanol for 10 min at -20°C. In order to stain, we incubated sections in isopropanol (Millipore Sigma) for 6 min, Mayer’s hematoxylin (Dako, Agilent, Santa Clara, CA) for 7 min, bluing buffer (Dako) for 1 min, and Eosin
(Sigma-Aldrich) diluted 1:5 in Tris-base (0.45M Tris, 0.5M acetic acid, pH 6.0) for 1 min. We washed the slide with deionized water after each of the staining steps. After air-drying, we mounted the slides with 85% glycerol and then coverslip them. H&E images were taken at 40× magnification using Digital slice scanner (Hamamatsu). We removed the coverslip after imaging by immersing slides in RNase and DNase free water.

**Tissue permeabilization, reverse transcription and spatial library preparation**

In order to pre-permeabilize, we incubated sections at 37°C for 24 minutes with permeabilization Enzyme. Wells were washed with SSC (0.1×) (Sigma-Aldrich). SSC was removed and reverse transcription Master Mix was added to each well. Reverse transcription was conducted as ST recommended protocol. After RT, sections were incubated in KOH (0.08M) for 5 min at room temperature, and then were incubated in Second Strand Mix for 15 min at 65°C. After removement of Second Strand Mix, 100 μl Buffer EB were added, and sections were incubated in KOH for 10 min at room temperature. The samples were transfered from every single well to a corresponding tube containing Tris-HCl (1 M, pH 7.0). Next, 1 μl sample were added to the qPCR plate well containing the KAPA SYBR FAST qPCR Master Mix (KAPA Biosystems). A qPCR system was performed and determine the optimal number of cycles. After that, 65 μl cDNA Amplification Mix were added into remaining sample. They were incubated with the recommended protocol.

**Library preparation and RNA sequencing**

After the cDNA amplification products were qualified, the sequencing library was constructed with Library Construction KIT (10×Genomics). First, the cDNA was chemically knocked out. The cDNA fragment was cut into 200 ~ 300 bp fragments, and the cDNA fragment was segmented, terminal repair and an addition. The cDNA fragment was screened. The P7 adapter was connected and introduced into the sample index by PCR amplification. Finally, the sequence library was obtained. Sequencing was performed on Illumina Hiseq 3000/4000 with a 150bp pair-end run by Quick Biology (Pasadena, Ca). A data quality check was done on Illumina SAV. Demultiplexing was performed with Illumina fastq2 v 2.17 program.

**RNA sequencing analysis**

In this study, 10×genomics official software Space Ranger was used for data preprocessing, gene expression quantitative and point identification. Sequencing data preprocessing includes filtering the sequenced sequences, evaluating the quality of sequencing data, and calculating the sequence length distribution. Web-based ST Spot Detector Software Space ranger was used to identify the spatial barcode markers in Reads1 and UMI markers of different transcripts. Read2 were aligned to the genome, using the transcriptome specific alignment software STAR, and selected the sequence with unique alignment position for subsequent analysis. The gene spot matrix was generated by using visium spatial barcodes, and then point clustering and gene expression analysis were performed. Software Seurat was used to
analyze and cluster the samples. The data with low quality were filtered. PCA, including t-SNE and UMAP algorithm, were used to reduce the dimension of data and visualize data.

**Data quality control and normalize with Space Ranger**

Sequencing data Read1 contained barcode and UMI markers that distinguished transcripts. The 10× Genomics official analysis software Space Ranger was used to statistically analyze the UMI-related quality control information of each sample. In order to understand the biological significance of expression differences in different locations, it is necessary to classify the points according to the expression level. The points with similar expression levels may come from the same type. Space Ranger software can be used to preliminarily classify different locations. Firstly, the expression of all points is normalized to compare the expression, then the t-SNE dimension reduction analysis based on PCA is carried out and the t-SNE results are clustered.

**Identification of tumor cluster specific marker genes (DEGs)**

Gene sets with FDR adjusted p-values below 0.05 were considered significantly enriched in the related clusters, these genes were identified the differentially expressed genes (DEGs). KEGG and GO were used to analyze the involving signaling pathway information. Marker genes were identified based on the comprehensive analysis of database and gene rank of Log FC value in differentially expressed genes of the clusters. Pearson correlation was used to reveal the relationship between cluster-specific genes and marker genes.

**Clinical significance of tumor clusters-specific marker genes in HCC**

The prognostic value of tumor clusters-specific marker genes were evaluated in database of KM Plotter. The genes were submitted into the website. Approximately, 155 patient samples with HCC are split into two groups according to median value of marker gene expression from gene chip. These two groups of patient cohorts are compared by a Kaplan-Meier survival plot.

**Data Collection (TCGA-LIHC cohort and ICGC (LIRI-JP) cohort )**

RNA sequencing data and corresponding clinical information of 376 patients with liver cancer were downloaded from the TCGA website up to July 13, 2021 (https://portal.gdc.cancer.gov/repository). RNA-sequencing data and clinical information of another 260 tumor samples were obtained from ICGC website (https://dcc.icgc.org/releases/current/Projects/LIRI-JP). The data from TCGA and ICGC are both public. Thus, the present study was exempted from the approval of local ethics committees. The current research follows the TCGA and ICGC data access policies and publication guidelines.

**Construction and validation of a prognostic spatial clusters-related gene signature**

Univariate Cox analysis of overall survival (OS) was utilized to screen tumor clusters specific marker genes with prognostic values. P values were adjusted by Benjamini & Hochberg (BH) correction. LASSO-
penalized Cox regression analysis was performed to construct a prognostic model in order to minimize the risk of overfitting\textsuperscript{[6]}. The LASSO algorithm was used for variable selection and shrinkage with the "glmnet" R package. The normalized expression matrix of candidate prognostic DEGs was the independent variable in regression, and the dependent variable was the overall survival and status of patients in the TCGA cohort. The risk scores of the patients were calculated according to the normalized expression level of each gene and its corresponding regression coefficients. The formula was established as follows: \( \text{score} = \sum (\text{each gene's expression} \times \text{corresponding coefficient}) \). The patients were stratified into high-risk and low-risk groups based on the median value of the risk score. Based on the expression of genes in the signature, PCA analysis and t-SNE analysis were performed with "Rtsne" and “ggplot2” R packages to explore the distribution of different groups. The survival analysis was implemented to analyze the OS of high- and low-risk groups using the “survminer” R package. The "survival ROC" R package was used to conduct time-dependent ROC curve analyses to evaluate the predictive power of the gene signature. The infiltration scores of 16 immune cells and the activities of 13 immune-related pathways between the high- and low-risk groups were calculated by single-sample gene set enrichment analysis (ssGSEA) with the “GSVA” R package.

**Statistical Analysis**

The Chi-squared test was used to compare the different proportions. The ssGSEA scores of immune cells or immune pathways between high- and low-risk groups were compared by Mann-Whitney test, and the P value was adjusted by BH method. Kaplan-Meier analysis was employed to compare the differences of OS among different groups. Univariate and multivariate Cox analyses were performed to screen the independent predictors for OS. The correlation of prognostic model risk score or prognostic gene expression level with stromal score, immune score and drug sensitivity was tested by Spearman or Pearson correlation analysis. R software (Version 4.0.5) with packages venn, igraph, ggplot2, pheatmap, ggpubr, corrplot and survminer was used to create plots. In all statistical results, a two-tailed P value less than 0.05 indicated statistical significance.

**Results**

We analyzed clusters specific marker genes (DEGs) from two perspectives, therefore results were presented in two parts.

1. The establishment of marker gene profiles for the prediction of satellite nodules

The flow chart of first part is shown in Fig. 1. To spatially analyze the gene expression of each HCC, we analyzed 3 individual patient samples (one with satellite nodules and two without satellite nodules) using the ST methodology. For each specimen, we separated the bulk tumor from the adjacent peritumoral tissue. We performed ST to detect the spatial gene expression between 6 tissue sections. Figure 1 shows the gene number distribution, expression distribution, mitochondrial genes, and hemoglobin gene expression ratio of all the spots. This figure also indicates the spatial gene number
distribution and expression distribution in 6 sections. Overall, we analyze 10820 tissue regions within the 6 samples.

1.1 The spatial transcriptome heterogeneity in HCC

We initially analyzed one pair of tissue sections containing a tumor and a peritumoral tissue with satellite nodules. In order to analyze the different components within the cell in HCC, principal component analysis were used to analyze differentially expressed gene among all cells. According to the results of t-SNE and UMAP, all spots of 2 sections were grouped into 4 main clusters (Fig 2 A-C). These clusters show spatial morphological characteristics that closely reflects those of histologically identifiable structures, including tumors, peritumor tissue and stroma. Each cluster had unique differentially expressed genes, indicating intratumoral and intertumoral heterogeneity in HCC (Figure 2D-F). Analysis of these 4 clusters shows that the gene expression between peritumoral tissues are similar, but they are significantly different from the tumors.

Weaker correlation between tumor samples than between peritumoral tissues were observed. It suggested that there's more pronounced intertumoral heterogeneity (Figure 2E). Thereafter, we performed differential gene expression analysis to identify clusters-specific marker genes and then defined the identity of each cluster (Fig 4F). The gene expression profiles represents the HCC expression phenotype taking tissue origin or the functional respective tissue components into account.

1.2 Spatial Transcriptomics Analysis uncovers changes that could not be detected by single-cell Transcriptomics Analysis

Interestingly, there were two tumor clusters in case 1, which had satellite nodules, while there was only one tumor cluster in the other two cases, which did not have satellite nodules. To understand the possible molecular mechanism of intrahepatic metastases, we sought to investigate functional differences in gene expression between these two tumor clusters in more detail. In tumor cluster A, PPIB, UGT2B15 and IFI27 were among the most highly expressed genes (Fig 3 A and C), and their functions were related to cell survival and apoptosis. In tumor cluster B (Fig 3 B and D), NDRG1, BHLHE40 and VEGFA, all of which have functions related to metastasis and invasion, were among the most highly expressed genes. Besides our gene expression analyze, pathways in two tumor clusters were investigated (Fig 3E). The pathways activated in tumor cluster A are mainly linked to altered cellular metabolism (metabolic pathways, drug metabolism, ascorbate and aldarate metabolism and steroid biosynthesis). Metabolic alteration is a hallmark of cancer[7] and clearly characterizes HCC[8]. A large number of clinical parameters are used to evaluate liver function reflecting changes in both enzyme activity and metabolites. On the other hand, we found that Gene Ontology (GO) categories linked to transcriptional misregulation in cancer and the MAPK signaling pathway to be enriched in tumor cluster B of case 1, possibly reflecting increased cell proliferation and malignancy.

1.3 Spatial expression patterns common to HCC samples
Next, gene expression analyses for another two cases were performed. The resulting gene expression profiles of tumor and nonneoplastic tissues were similar (Supplementary Fig. 1). The gene expression within each region (tumor, peritumor and stroma) obtained from the previous results was used to determine region-specific markers. The expression of some specific genes were observed in the tumor than in nonneoplastic tissues in each case. Specifically, in the tumor region, we observed enrichment of PPIB and UGT2B15 and IFI27, NDRG1, BHLHE40 and VEGFA (Supplementary Fig. 2).

We then investigated more functional differences in pathways between the tumor and the peritumor region(Supplementary Fig.3). Consistent with the pathway analysis in case 1, enrichment analysis indicates that the pathways activated in the tumor (for example, genes involved in chemical carcinogenesis and metabolic pathways) are differentially expressed in comparison to the peritumor.

In a word, the gene profiles obtained from the ST analyses may reflect the different status of the HCC and can manifest intertumoral heterogeneity between patients at the gene expression level.

1.4 The effect of marker gene expression in tumor clusters on HCC patient prognosis

To evaluate the clinical significance of marker genes of tumor clusters, the KM plotter website was used to analyze these gene markers. According to 155 HCC samples in the KM Plotter database, high expression levels of PPIB, UGT2B15, IFI27, NDRG1, BHLHE40 and VEGFA were associated with poor prognosis in HCC (Figure 4).

2. The construct of prognostic Gene Signature for survival

The flow chart of second part is shown in Supplementary Fig. 1 A total of 361 HCC patients from the TCGA-LIHC cohort and 260 HCC patients from the ICGC (LIRI-JP) cohort were finally enrolled. The detailed clinical characteristics of these patients are summarized in Table 1.

2.1 Identifiation of prognostic spatial clusters specific marker genes (DEGs) and construction of a prognostic model in the TCGA cohort

Spatial clusters specific marker genes (DEGs) were differentially expressed in tumor tissues and adjacent non-tumorous tissues, and 7 of them were correlated with OS in the univariate Cox regression analysis (Fig. 5A). The expression profiles of the above 7 genes were analyzed by LASSO-Cox regression analysis, and the prognostic model was established. RPS7 was excluded from this analysis because its lasso coefficient was un available. A 6-gene signatureADH1A,ADH1B,CYP3A4,FCGBP,PABPC1, NDRG1 was identified based on the optimal value of $\lambda$ (Supplementary Fig. 5). The risk score was calculated as follows: score = -0.0354*expression level of ADH1A -0.0035*expression level of ADH1B-0.0100*expression level of CYP3A4 + 0.113*expression level of FCGBP+0.119*expression level of PABPC1 +0.126*expression level of NDRG1. Fig 5B showed that the expression of those 6 genes was significant different in HCC tissues and adjacent normal tissues. The risk ratio of NDRG1, one of
important markers from tumor cluster B, is 1.309 (95% CI = 1.147-1.493, \( P < 0.001 \), Fig. 5C). Figure 5D showed the relationship of those 6 genes. The result showed that these 6 spatial clusters specific marker genes can be used as prognostic indicators. Based on the median cut-off value, we divided the patients into a high-risk group (n=182) or a low-risk group (n=183) a (Fig. 6A). We found that the higher risk group was correlated with higher tumor grade (Table 2). We can see that patients with low risk were less likely to die earlier than those with high risk from the scatter chart (Figure 6B). We can see that patients were distributed in the two subgroups according to they are in high or low risk group through PCA analysis and t-SNE analysis (Figures 6E–F). On the other hand, the survival analysis showed that the patients with low risk had a better OS than those with high-risk within 5 years (Figure 1, \( P < 0.05 \)). Time-dependent ROC curves were performed for displaying sensitivity and specificity of survival prognostic model, and the area under the curve (AUC) reached 0.700 at 1 year, 0.647 at 2 years, and 0.606 at 3 years (Figure 6J).

To explore the relationship between each prognostic gene and prognosis, survival analysis was performed. The result indicated that low expression of ADH1A,CYP3A4 and ADH1B was significantly correlated with poor OS (Supplementary Fig. 6, A-C, \( P < 0.01 \)) and high expression of FCGBP, PABPC1 and NDRG1 was significantly correlated with poor OS (Supplementary Fig. 6D-E and Fig 4.D, \( P < 0.05 \)). The expression levels of FCGBP, PABPC1 and NDRG1 were higher and the expression of ADH1A,CYP3A4 and ADH1B were lower in tumor tissues compared with adjacent non-tumorous tissues (Supplementary Fig. 7).

2.2 Validation of the 6 Gene Signature in the ICGC Cohort

To further validate the stability of the model based on TCGA cohort, we performed the same analysis with ICGC cohort. Refer to the median value obtained from the TCGA cohort, patients from ICGC cohort were also divided into two groups with high or low-risk. Consistent with the results from the TCGA cohort, PCA and t-SNE analyses confirmed that patients were separated in two directions (Figures 6G, H). Meanwhile, patients in the low-risk group were less likely to die earlier (Figure 6D) and had a longer survival time compared with the high-risk group (Figure 6K). Moreover, the AUC of the 6-gene signature was 0.819 at 1 year, 0.764 at 2 years, and 0.759 at 3 years (Figure 6L).

2.3 Independent Prognostic Value of the 6 Gene Signature

To validate whether the risk score was an independent prognostic factor for OS, we carried out univariate and multivariate Cox analyses. In both TCGA and ICGC cohorts, the risk scores were significantly correlated with OS according to univariate Cox analysis (TCGA cohort: HR = 2.994, 95% CI = 1.892-4.736, \( P < 0.001 \), ICGC cohort: HR = 5.358, 95% CI =2.818-10.188, \( P < 0.001 \)) (Figures 7A, C). Even after correcting for other factors, the risk score was still an independent predictor for OS based on multivariate Cox analysis (TCGA cohort: HR = 2.546, 95% CI = 1.617-4.008, \( P < 0.001 \), ICGC cohort: HR =3.780, 95% CI = 1.879-7.604, \( P =0.001 \)) (Figures 7B, D).

2.4 Functional analyses in the TCGA and the ICGC cohort
The association of risk score with the clinical characteristics of HCC patients was analyzed. The result showed that the risk score was significantly higher in tumor tumor stage III-IV (P < 0.001) or grade 3-4 (P < 0.001) compared with tumor stage I-II (Figures 7G, H) or tumor grade 1-2. The same analysis in the ICGC dataset confirmed that the risk score was definitely higher in tumor stage III-IV compared with tumor stage I-II (There was no data about the grade of HCC in the ICGC dataset) (Figures 7K).

To detect whether the risk score was associated with immune components, we analyzed the relationship of risk score and immune infiltrates.

We can see that the factors such as DCs, iDCs, APC-co-stimulation, HLA, MHC class1 which were important for antigen presentation process were significantly induced in the high-risk group in the TCGA cohort (all adjusted P < 0.05, Figures 8A, B). Besides, the scores of macrophages or Treg cells were higher in the high risk group, while the score of type II IFN response and mast cells were just the opposite (adjusted P < 0.05, Fig. 8 A, B). Consistent with the analysis from TCGA cohort, comparison between high and low risk groups of ICGC cohort showed the similar result (adjusted P < 0.05, Figures 8C, D).

Six types of immune infiltrates were identified in human tumors, namely C1 (wound healing), C2 (INF-γ dominant), C3 (inflammatory), C4 (lymphocyte depleted), C5 (immunologically quiet) and C6 (TGF-γ dominant) [9]. C5 and C6 immune subtypes were not included in the study because that no patient sample belonged to C5 immune subtype and C6 immune subtype in HCC. We analyzed the relationship between the immune infiltration and the risk score. We can see the high risk score was significantly associated with C2, while low risk score was significantly associated with C3 (Figure 8E). We also detect tumor stemness with RNA stemness score (RNAss) based on mRNA expression and DNA stemness score based on DNA methylation pattern (DNAss) [10]. We estimated tumor immune microenvironment with stromal score and immune score. We also performed the correlation analysis to explore the relationship between tumor stem cells and the immune environment. Based on the results, we can see that the risk score was positively significantly correlated with DNAss and ImmuneScore, while significantly negative associated with RNAss and StromaScore (P < 0.05) (Figure 8F).

The relationship between the expression of prognostic genes levels and drug sensitivity was performed too. The results showed that all prognostic genes were correlative to some chemotherapy drug sensitivity (P < 0.01) (Figure 9). For example, increased expression of NDRG1 with JNJ-42756493, Simvastatin and Cabozantinib.

**Discussion**

HCC treatment is still not promising because of high recurrence rate even after complete surgical excision. The presence of satellite lesions neglected by imaging because of small size may be one of the key factors that lead to the high recurrence. The purpose of the present study was to perform a full analysis of the spatial transcriptome of HCC with satellite nodules. We investigated tissue-wide gene
expression heterogeneity using ST technology, which quantifies transcriptome arrays of the whole tissue sections.

Recently, a study including multiple histological types of cancer suggested that adjacent tumor tissue may be an intermediate state between normal and tumor tissue\textsuperscript{[11]}. However, no evaluation of tumor and peritumor tissue with spatial resolution has been conducted until now. Therefore, we determined to identify differences between the tumor and the peritumor specimens. The results showed that the gene expression and pathway enrichment were different in tumor and peritumor clusters. For example, known HCC-related genes (NDRG\textsubscript{1}\textsuperscript{[12,13]} and VEGFA\textsuperscript{[14]}) are among the most highly expressed genes in the “tumor” clusters. These highly expressed gene profiles may become gene markers of HCC, suggesting poor prognosis. The present study also offers a new perspective into gene expression differences between tumors with and without satellite nodules, prompting key questions with important implications for the metastasis of HCC. The results revealed two tumor clusters in the case with satellite nodules. We identified marker genes corresponding to the different clusters of tumor cells. The gene expression profiles obtained from the ST analysis may be used to forecast further regions of satellite nodules. PPIB, UGT2B15 and IFI27, which were found in tumor cluster A, were mainly associated with cell survival and apoptosis. PPIB (cyclophilin B, CypB) is a member of the PPIase family. It has been reported to play an important role in protein folding. Recent studies have shown that HCC cell survival can be stimulated by PPIB through a positive feedback loop with hypoxia-inducible factor-1a (HIF-1a)\textsuperscript{[15]}. PPIB is associated with malignant progression and gene regulation has been noted by some researchers\textsuperscript{[16]}. Overexpression of miR-206 promotes apoptosis and inhibits the metastasis of HCC cells by targeting PPIB\textsuperscript{[17]}. UGT2B15 is one functional member belong to the UGT2B subfamily. The expression of UGT2B15 is mainly observed in liver, prostate and breast cancer. It has been found to contribute to glucuronidation of androgenic steroids\textsuperscript{[18–20]}. The role of UGT2B15 in induce of tumor progression and drug resistance has been reported in some study\textsuperscript{[20]}. Bioinformatic analysis suggested that UGT2B15 activates the Hippo-YAP signaling pathway, leading to the pathogenesis of gastric cancer\textsuperscript{[20]}. Interferon alpha-inducible protein 27 (IFI27) consists of 122 amino acids. It belong to a hydrophobic mitochondrial protein family\textsuperscript{[21]}. IFI27 keeps a low level of expression in multiple mammalian cells and involved in a wide range of biological processes, such as apoptosis and innate immunity\textsuperscript{[22,23]}. The development of tumor can be affected by IFI27 downregulation in many cancers. It has been reported that TRAIL-induced apoptosis in animal and cellular models of HCC and gastric cancer can be induced when IFI27 was downregulated. It suggested that IFI27 may play a critical important role in tumor development\textsuperscript{[24]}. IFI27 downregulation resulting in a decrease of the formation of the cyclin A/CDK1 complex, inducing epithelial cell arrest in S phase, and cell proliferation is therefore inhibited\textsuperscript{[25]}. Hong W reported that high level of expression of IFI27 promotes cell proliferation and invasion and reduces apoptosis\textsuperscript{[26]}. NDRG1, BHLHE40 and VEGFA, which were found in tumor cluster B, mainly affect tumor proliferation, metastasis and invasion. N-Myc downstream-regulated gene 1 (NDRG1) is a crucial cytosolic ubiquitously expressed protein. NDRG1 is an important molecule in controlling HCC metastasis and is thus suggested as a novel biomarker for predicting HCC recurrence after liver transplantation\textsuperscript{[13]}. NDRG1 is one of significant markers for
metastasis, recurrence and poor prognosis in HCC. Yan et al. found that NDRG1 expression is generally upregulated in HCC tissues compared with normal samples, particularly in recurrent and metastatic HCC. BHLHE40 (also known as DEC1/BHLHB2/SHARP2/STRA13) belong to the basic helix-loop-helix (bHLH) protein family, which is a large superfamily of transcriptional regulators expressed in many organisms. High expression of BHLHE40 significantly correlated with the activation of a hypoxia-response pathway, elevated metastatic potential, and poor prognosis in many tumors, such as HCC, pancreatic cancer, and invasive breast cancer. BHLHE40 is activated under hypoxic conditions by HIF-1α in HCC, stimulating tumor progression. Vascular endothelial growth factor (VEGF) is an essential angiogenic growth factor in physiological and pathological states. High expression of VEGFA is detected in a large number of solid tumors, including HCC. It is involved in the regulation of metastasis of many solid tumors and their neovasculature. In HCC, VEGF is an extremely important angiogenic factor. VEGFA secreted by tumor cells promotes an epithelial to mesenchymal transition phenotype, consequently inducing tumor invasion. Therefore, the spatial gene expression was measured in HCC tissues for the first time.

Spatially mapping gene expression uncovered a new landscape and another two cases validated the result. The results showed that these marker genes in tumor clusters were also highly expressed in the tumor area. We also detected the clinical significance of marker genes through the database and found that these genes can be used to predict prognosis and survival in patients with HCC. In the future, relevant panels can be detected, and molecules in different regions can help improve the accuracy of clinical predictions of which patients with HCC may be more prone to metastasis. High expression of these markers in both clusters may suggest the possibility of satellite nodules. Aggregation of these two gene clusters may suggest that there are satellite nodules in tumors.

To further explore the role of spacial clusters specific marker genes (DEGs), we screened the DEGs with TCGA and ICGC cohort. The expression of tumor clusters marker genes was significantly different between tumor and adjacent nontumorous tissues, and many of them were associated with OS according to the univariate Cox regression analysis. These results suggested that we could construct a prognostic model using the clusters specific marker gene. Patients were divided into high- and low-risk groups according to the median risk score. The result showed that high-risk group was significantly correlated with higher tumor grade, advanced TNM stage and shorter OS period. Risk score was an independent predictor for OS according to independent prognostic analysis.

We construct a prognostic model with 6 clusters specific marker genes (ADH1A, ADH1B, CYP3A4, FCGBP, PABPC1, NDRG1) in the present study. Some researchers reported that a variation in alcohol dehydrogenase 1A (ADH1A) may contribute to slow alcohol metabolism which induced increased blood acetaldehyde levels in Korean subjects. In HCC patients, a higher expression of ADH1A was associated with good survival and a less aggressive disease state. Alcohol Dehydrogenase 1B (ADH1B) is mainly known for its involvement in the major human ethanol metabolic pathway. As a multifunctional enzyme, the human hepatic cytochrome P-450 3A4 (CYP3A4) has a
wide range of substrates including commonly used drugs[40]. The expression of Fc fragment of IgG-binding protein (FCGBP) is low in some tumors, while high in others[41]. Low expression of FCGBP could be used as a crucial regulator of tumor growth factor 1 (TGF-1)-induced epithelial-mesenchymal transition in gallbladder cancer[42]. On the contrary, high expression of FCGBP significantly decreases the overall survival (OS) in colorectal cancer patients[43]. High expression of PABPC1 was associated with worse overall survival (OS) for HCC and it may contribute to the progression of HCC[44]. Just as discussed above, NDRG1 is a biomarker for metastasis, recurrence and of poor prognosis in HCC.

The results showed higher fractions of macrophages cells in high risk groups in both the TCGA and the ICGC. It has been demonstrated that increased tumor-associated macrophages cells are associated with poor prognosis in HCC patients due to their role in immune invasion in previous studies[45, 46]. Besides, higher risk scores was associated with impaired antitumor immunity, including the activity of the type II IFN response and type I IFN response as well as the fractions of NK cells. Therefore, it is reasonable to assume that anti-tumor immunity of the high-risk group is attenuated, which may be an important reason for its poor prognosis. According to ESTIMATE algorithm, the prognostic gene expression was also significant correlated with stromal score and immune score(P < 0.05) indicated that the tumor tissue in the high-risk group is highly infiltrated by immune cells. The result of the expression of prognostic genes levels and drug sensitivity. The results showed that all prognostic genes were correlative to some chemotherapy drug sensitivity. These data demonstrated that some prognostic genes can be used as therapeutic targets to overcome drug resistance or adjuvant drug sensitivity.

In conclusion, the results in this study have demonstrated that the analysis of tumor gene expression combined with spatial transcriptomics (ST) remarkably increases granularity when compared to a bulk analysis. Tumors with negligible histological differences and various regions of the same tumor showed significant differences in the transcription profiles of tumor cells at each site. High expression of PPIB, UGT2B15, IFI27, NDRG1, BHLHE40 and VEGFA may suggest the formation of satellite nodules that cannot be detected by imaging.

The study also defined a new prognostic signature consisting six clusters specific marker genes. It has been proved that the gene profile is independently associated with OS in TCGC cohort and ICGC validation cohort, and was confirmed to be valuable in functional analysis, tumor microenvironment and drug sensitivity, providing insight for predicting the prognosis of HCC. The specific potential mechanism between clusters specific marker genes and tumor immunity in HCC remains unclear, which is worthy of further study.

In summary, we propose that expression profiles based on spatial analysis can serve as new markers for the prediction of HCC prognosis.

Declarations

Ethics approval and consent to participate
Written informed consent was obtained from all participants. Ethical approval was obtained from the Ethical Committee of Tianjin medical university.

**Funding**

This study was supported by the project of National Nature Science Foundation of China (No.81572872), and the project of Nature Science Foundation of Tianjin (No. 19JCYBJC25800).

**Consent for publication**

Not applicable.

**Availability of data and material**

All data generated or analysed during this study are included in this published article.

**Acknowledgements**

Not applicable.

**Competing interests**

The authors declare that there are no competing interests.

**Author Contributions Section**

Nan Zhao contributed to the conception of the study and contributed significantly to analysis and manuscript preparation, Yanhui Zhang, Runfen Cheng, Danfang Zhang, Fan Li, Yuhong Guo, Zhiqiang Qiu, Xueyi Dong and Xinchao Ban performed the experiment. Baocun Sun contributed to the conception of the study and manuscript preparation, Xiulan Zhao contributed to the conception of the study and performed the experiment.

**References**

1. Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A. & Jemal, A. (2018) Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, *CA Cancer J Clin.* **68**, 394-424.

2. Allemani, C., Matsuda, T., Di Carlo, V., Harewood, R., Matz, M., Niksic, M., Bonaventure, A., Valkov, M., Johnson, C. J., Esteve, J., Ogunbiyi, O. J., Azevedo, E. S. G., Chen, W. Q., Eser, S., Engholm, G., Stiller, C. A., Monnereau, A., Woods, R. R., Visser, O., Lim, G. H., Aitken, J., Weir, H. K., Coleman, M. P. & Group, C. W. (2018) Global surveillance of trends in cancer survival 2000-14 (CONCORD-3): analysis of individual records for 37 513 025 patients diagnosed with one of 18 cancers from 322 population-based registries in 71 countries, *Lancet.* **391**, 1023-1075.
3. Yin, Z., Dong, C., Jiang, K., Xu, Z., Li, R., Guo, K., Shao, S. & Wang, L. (2019) Heterogeneity of cancer-associated fibroblasts and roles in the progression, prognosis, and therapy of hepatocellular carcinoma, *J Hematol Oncol.* **12**, 101.

4. Erstad, D. J. & Tanabe, K. K. (2019) Prognostic and Therapeutic Implications of Microvascular Invasion in Hepatocellular Carcinoma, *Ann Surg Oncol.* **26**, 1474-1493.

5. Ikeda, K., Seki, T., Umehara, H., Inokuchi, R., Tamai, T., Sakaida, N., Uemura, Y., Kamiyama, Y. & Okazaki, K. (2007) Clinicopathologic study of small hepatocellular carcinoma with microscopic satellite nodules to determine the extent of tumor ablation by local therapy, *Int J Oncol.* **31**, 485-91.

6. Simon, N., Friedman, J., Hastie, T. & Tibshirani, R. (2011) Regularization Paths for Cox’s Proportional Hazards Model via Coordinate Descent, *N/A*. 39, 1-13.

7. Pavlova, N. N. & Thompson, C. B. (2016) The Emerging Hallmarks of Cancer Metabolism, *Cell Metab.* **23**, 27-47.

8. De Matteis, S., Ragusa, A., Marisi, G., De Domenico, S., Casadei Gardini, A., Bonafe, M. & Giudetti, A. M. (2018) Aberrant Metabolism in Hepatocellular Carcinoma Provides Diagnostic and Therapeutic Opportunities, *Oxid Med Cell Longev.* **2018**, 7512159.

9. Desgroisellier, J. S. & Cheresh, D. A. (2010) Integrins in cancer: biological implications and therapeutic opportunities, *N/A*. 10, 9-22.

10. Malta, T. M., Sokolov, A., Gentles, A. J., Burzykowski, T., Poisson, L., Weinstein, J. N., Kaminska, B., Huelsken, J., Omberg, L., Gevaert, O., Colaprico, A., Czerwinska, P., Mazurek, S., Mishra, L., Heyn, H., Krasnitz, A., Godwin, A. K., Lazar, A. J., Cancer Genome Atlas Research, N., Stuart, J. M., Hoadley, K. A., Laird, P. W., Noushmehr, H. & Wiznerowicz, M. (2018) Machine Learning Identifies Stemness Features Associated with Oncogenic Dedifferentiation, *Cell.* **173**, 338-354 e15.

11. Aran, D., Camarda, R., Odegaard, J., Paik, H., Oskotsky, B., Krings, G., Goga, A., Sirotta, M. & Butte, A. J. (2017) Comprehensive analysis of normal adjacent to tumor transcriptomes, *Nat Commun.* **8**, 1077.

12. Liu, Y., Wang, D., Li, Y., Yan, S., Dang, H., Yue, H., Ling, J., Chen, F., Zhao, Y., Gou, L., Tang, P., Huang, A. & Tang, H. (2019) Long noncoding RNA CCAT2 promotes hepatocellular carcinoma proliferation and metastasis through up-regulation of NDRG1, *Exp Cell Res.* **379**, 19-29.

13. Cheng, J., Xie, H. Y., Xu, X., Wu, J., Wei, X., Su, R., Zhang, W., Lv, Z., Zheng, S. & Zhou, L. (2011) NDRG1 as a biomarker for metastasis, recurrence and of poor prognosis in hepatocellular carcinoma, *Cancer Lett.* **310**, 35-45.

14. Yang, W., Li, Z., Qin, R., Wang, X., An, H., Wang, Y., Zhu, Y., Liu, Y., Cai, S., Chen, S., Sun, T., Meng, J. & Yang, C. (2019) YY1 Promotes Endothelial Cell-Dependent Tumor Angiogenesis in Hepatocellular Carcinoma by Transcriptionally Activating VEGFA, *Front Oncol.* **9**, 1187.

15. Kim, Y., Jang, M., Lim, S., Won, H., Yoon, K. S., Park, J. H., Kim, H. J., Kim, B. H., Park, W. S., Ha, J. & Kim, S. S. (2011) Role of cyclophilin B in tumorigenesis and cisplatin resistance in hepatocellular carcinoma in humans, *Hepatology.* **54**, 1661-78.

16. Choi, T. G., Nguyen, M. N., Kim, J., Jo, Y. H., Jang, M., Nguyen, N. N. Y., Yun, H. R., Choe, W., Kang, I., Ha, J., Tang, D. G. & Kim, S. S. (2018) Cyclophilin B induces chemoresistance by degrading wild-type p53
via interaction with MDM2 in colorectal cancer, *J Pathol.* **246**, 115-126.

17. Wu, H., Xie, D., Yang, Y., Yang, Q., Shi, X. & Yang, R. (2020) Ultrasound-Targeted Microbubble Destruction-Mediated miR-206 Overexpression Promotes Apoptosis and Inhibits Metastasis of Hepatocellular Carcinoma Cells Via Targeting PPIB, *Technol Cancer Res Treat.* **19**, 153303820959355.

18. Hu, D. G., Selth, L. A., Tarulli, G. A., Meech, R., Wijayakumara, A., Chanawong, A., Russell, R., Caldas, C., Robinson, J. L., Carroll, J. S., Tilley, W. D., Mackenzie, P. I. & Hickey, T. E. (2016) Androgen and Estrogen Receptors in Breast Cancer Coregulate Human UDP-Glucuronosyltransferases 2B15 and 2B17, *Cancer Res.* **76**, 5881-5893.

19. Grosse, L., Paquet, S., Caron, P., Fazli, L., Rennie, P. S., Belanger, A. & Barbier, O. (2013) Androgen glucuronidation: an unexpected target for androgen deprivation therapy, with prognosis and diagnostic implications, *Cancer Res.* **73**, 6963-71.

20. Chen, X., Li, D., Wang, N., Yang, M., Liao, A., Wang, S., Hu, G., Zeng, B., Yao, Y., Liu, D., Liu, H., Zhou, W., Xiao, W., Li, P., Ming, C., Ping, S., Chen, P., Jing, L., Bai, Y. & Yao, J. (2018) Bioinformatic analysis suggests that UGT2B15 activates the HippoYAP signaling pathway leading to the pathogenesis of gastric cancer, *Oncol Rep.* **40**, 1855-1862.

21. Cheriyath, V., Leaman, D. W. & Borden, E. C. (2011) Emerging roles of FAM14 family members (G1P3/ISG 6-16 and ISG12/IFI27) in innate immunity and cancer, *J Interferon Cytokine Res.* **31**, 173-81.

22. Rosebeck, S. & Leaman, D. W. (2008) Mitochondrial localization and pro-apoptotic effects of the interferon-inducible protein ISG12a, *Apoptosis.* **13**, 562-72.

23. Mihalich, A., Viganò, P., Gentilini, D., Borghi, M. O., Vignali, M., Busacca, M. & Di Blasio, A. (2012) Interferon-inducible genes, TNF-related apoptosis-inducing ligand (TRAIL) and interferon inducible protein 27 (IFI27) are negatively regulated in leiomyomas: implications for a role of the interferon pathway in leiomyoma development, *Gynecol Endocrinol.* **28**, 216-9.

24. Liu, N., Zuo, C., Wang, X., Chen, T., Yang, D., Wang, J. & Zhu, H. (2014) miR-942 decreases TRAIL-induced apoptosis through ISG12a downregulation and is regulated by AKT, #N/A. 5, 4959-71.

25. Hsieh, W. L., Huang, Y. H., Wang, T. M., Ming, Y. C., Tsai, C. N. & Pang, J. H. (2015) IFI27, a novel epidermal growth factor-stabilized protein, is functionally involved in proliferation and cell cycling of human epidermal keratinocytes, *Cell Prolif.* **48**, 187-97.

26. Wang, H., Qiu, X., Lin, S., Chen, X., Wang, T. & Liao, T. (2018) Knockdown of IFI27 inhibits cell proliferation and invasion in oral squamous cell carcinoma, *World J Surg Oncol.* **16**, 64.

27. Yan, S., Tang, Z., Chen, K., Liu, Y., Yu, G., Chen, Q., Dang, H., Chen, F., Ling, J., Zhu, L., Huang, A. & Tang, H. (2018) Long noncoding RNA MIR31HG inhibits hepatocellular carcinoma proliferation and metastasis by sponging microRNA-575 to modulate ST7L expression, #N/A. 37, 214.

28. Xiong, J., Yang, H., Luo, W., Shan, E., Liu, J., Zhang, F., Xi, T. & Yang, J. (2016) The anti-metastatic effect of 8-MOP on hepatocellular carcinoma is potentiated by the down-regulation of bHLH transcription factor DEC1, *Pharmacol Res.* **105**, 121-33.
29. Wu, Y., Sato, F., Yamada, T., Bhawal, U. K., Kawamoto, T., Fujimoto, K., Noshiro, M., Seino, H., Morohashi, S., Hakamada, K., Abiko, Y., Kato, Y. & Kijima, H. (2012) The BHLH transcription factor DEC1 plays an important role in the epithelial-mesenchymal transition of pancreatic cancer, *Int J Oncol.* **41**, 1337-46.

30. Chakrabarti, J., Turley, H., Campo, L., Han, C., Harris, A. L., Gatter, K. C. & Fox, S. B. (2004) The transcription factor DEC1 (stra13, SHARP2) is associated with the hypoxic response and high tumour grade in human breast cancers, *Br J Cancer.* **91**, 954-8.

31. Kiss, Z., Mudryj, M. & Ghosh, P. M. (2020) Non-circadian aspects of BHLHE40 cellular function in cancer, *Genes Cancer.* **11**, 1-19.

32. Campagnolo, L., Telesca, C., Massimiani, M., Stuhlmann, H., Angelico, M., Lenci, I., Manzia, T. M., Tariciotti, L., Lehmann, G. & Baiocchi, L. (2016) Different expression of VEGF and EGFL7 in human hepatocellular carcinoma, *Dig Liver Dis.* **48**, 76-80.

33. de Oliveira, A., Castanhole-Nunes, M. M. U., Biselli-Chicote, P. M., Pavarino, E. C., da Silva, R., da Silva, R. F. & Goloni-Bertollo, E. M. (2020) Differential expression of angiogenesis-related miRNAs and VEGFA in cirrhosis and hepatocellular carcinoma, *Arch Med Sci.* **16**, 1150-1157.

34. Veenendaal, L. M., Jin, H., Ran, S., Cheung, L., Navone, N., Marks, J. W., Waltenberger, J., Thorpe, P. & Rosenblum, M. G. (2002) In vitro and in vivo studies of a VEGF121/rGelonin chimeric fusion toxin targeting the neovasculature of solid tumors, *Proc Natl Acad Sci U S A.* **99**, 7866-71.

35. Goel, H. L. & Mercurio, A. M. (2013) VEGF targets the tumour cell, *#N/A.* **13**, 871-882.

36. Sun, H., Zhang, D., Huang, C., Guo, Y., Yang, Z., Yao, N., Dong, X., Cheng, R., Zhao, N., Meng, J., Sun, B. & Hao, J. (2021) Hypoxic microenvironment induced spatial transcriptome changes in pancreatic cancer, *Cancer Biol Med.*

37. Lee, Y. J., Yoo, M. G., Kim, H. K., Jang, H. B., Park, K. J., Lee, H. J., Kim, S. G. & Park, S. I. (2019) The association between alcohol metabolism and genetic variants of ADH1A, SRPRB, and PGM1 in Korea, *Alcohol.* **79**, 137-145.

38. Zahid, K. R., Yao, S., Khan, A. R. R., Raza, U. & Gou, D. (2019) mTOR/HDAC1 Crosstalk Mediated Suppression of ADH1A and ALDH2 Links Alcohol Metabolism to Hepatocellular Carcinoma Onset and Progression in silico, *Front Oncol.* **9**, 1000.

39. Polimanti, R. & Gelernter, J. (2018) ADH1B: From alcoholism, natural selection, and cancer to the human phenome, *Am J Med Genet B Neuropsychiatr Genet.* **177**, 113-125.

40. Yuan, X., Lu, H., Zhao, A., Ding, Y., Min, Q. & Wang, R. (2020) Transcriptional regulation of CYP3A4 by nuclear receptors in human hepatocytes under hypoxia, *Drug Metab Rev.* **52**, 225-234.

41. Wang, K., Guan, C., Shang, X., Ying, X., Mei, S., Zhu, H., Xia, L. & Chai, Z. (2021) A bioinformatic analysis: the overexpression and clinical significance of FCGBP in ovarian cancer, *Aging (Albany NY).* **13**, 7416-7429.

42. Xiong, L., Wen, Y., Miao, X. & Yang, Z. (2014) NT5E and FcGBP as key regulators of TGF-1-induced epithelial-mesenchymal transition (EMT) are associated with tumor progression and survival of patients with gallbladder cancer, *Cell Tissue Res.* **355**, 365-74.
43. Qi, C., Hong, L., Cheng, Z. & Yin, Q. (2016) Identification of metastasis-associated genes in colorectal cancer using metaDE and survival analysis, *Oncol Lett.* **11**, 568-574.

44. YuFeng, Z. & Ming, Q. (2020) Expression and prognostic roles of PABPC1 in hepatocellular carcinoma, *Int J Surg.* **84**, 3-12.

45. Zhou, S. L., Zhou, Z. J., Hu, Z. Q., Huang, X. W., Wang, Z., Chen, E. B., Fan, J., Cao, Y., Dai, Z. & Zhou, J. (2016) Tumor-Associated Neutrophils Recruit Macrophages and T-Regulatory Cells to Promote Progression of Hepatocellular Carcinoma and Resistance to Sorafenib, *Gastroenterology.* **150**, 1646-1658 e17.

46. Zhang, Q., He, Y., Luo, N., Patel, S. J., Han, Y., Gao, R., Modak, M., Carotta, S., Haslinger, C., Kind, D., Peet, G. W., Zhong, G., Lu, S., Zhu, W., Mao, Y., Xiao, M., Bergmann, M., Hu, X., Kerkar, S. P., Vogt, A. B., Pflanz, S., Liu, K., Peng, J., Ren, X. & Zhang, Z. (2019) Landscape and Dynamics of Single Immune Cells in Hepatocellular Carcinoma, *Cell.* **179**, 829-845 e20.

**Tables**

**Table 1 Clinical characteristics of the HCC patients used in this study**
| Characteristics          | TCGA-LIHC cohort | ICGC-LIRP-JI cohort |
|-------------------------|-----------------|--------------------|
| No. of patients         | 371             | 260                |
| Age (median, range)     | 61 (16-90)      | 69 (31-89)         |
| Gender                  |                 |                    |
| FEMALE                  | 120 (32.53%)    | 68 (26.15%)        |
| MALE                    | 251 (67.65%)    | 192 (73.85%)       |
| Grade                   |                 |                    |
| G1                      | 55              | NA                 |
| G2                      | 178             | NA                 |
| G3                      | 120             | NA                 |
| G4                      | 13              | NA                 |
| unknown                 | 5               | NA                 |
| Stage                   |                 |                    |
| Stage I-                | 174             | 40                 |
| Stage II                | 85              | 117                |
| Stage III               | 84              | 80                 |
| Stage IV                | 4               | 23                 |
| unknown                 | 24              | NA                 |
| Survival time           |                 |                    |
| Alive                   | 239             | 214                |
| Deceased                | 132             | 46                 |

Table 2. Characteristics of the patients in different risk groups
| Characteristics | TCGA-LIHC cohort | ICGC-LIRP-JI cohort | Pvalue | TCGA-LIHC cohort | ICGC-LIRP-JI cohort | Pvalue |
|-----------------|-----------------|---------------------|--------|-----------------|---------------------|--------|
|                 | High risk       | Low risk            |        | High risk       | Low risk            |        |
| Age             |                 |                     |        |                 |                     |        |
| <=65            | 118(69.01%)     | 101(60.12%)         | 0.1102 | 67(38.95%)      | 22(37.29%)          | 0.9428 |
| >65             | 53(30.99%)      | 67(39.88%)          |        | 105(61.05%)     | 37(62.71%)          |        |
| Gender          |                 |                     |        |                 |                     |        |
| FEMALE          | 54(31.58%)      | 54(32.14%)          | 1      | 49(28.49%)      | 12(20.34%)          | 0.2918 |
| MALE            | 117(68.42%)     | 114(67.86%)         |        | 123(71.51%)     | 47(79.66%)          |        |
| Grade           |                 |                     |        |                 |                     |        |
| G1+2            | 87(50.88%)      | 125(74.4%)          | <0.001 | NA              | NA                  | NA     |
| G3+4            | 84(49.12%)      | 43(25.6%)           | NA     | NA              | NA                  | NA     |
| Stage           |                 |                     |        |                 |                     |        |
| Stage I- II     | 121(70.76%)     | 134(79.76%)         | 0.0729 | 96(55.81%)      | 45(76.27%)          | 0.0086 |
| Stage III- IV   | 50(29.24%)      | 34(20.24%)          |        | 76(44.19%)      | 14(23.73%)          |        |

Figures
Study design for spatial transcriptomics in HCC (A) Work flow of HCC. HCC specimens of three patients were dissected to separate bulk tumor from peritumor tissue. Three pairs of tumor and peritumoral tissues were analyzed. (B) Distribution of all expression gene numbers, distribution of all expression gene, distribution of mitochondrial gene and distribution of hemoglobin gene expression in three pairs of samples.
Figure 2

Spatial gene expression heterogeneity within case1 tissue sample. (A) Tissue plot with spots colored by UMI count and t-SNE projection of spots colored by UMI counts. (B) Tissue plot with spots colored by Clustering and t-SNE projection of spots colored by Clustering. (C) Sequencing Saturation and Median Genes per Spot. (D) Violin plots displaying the expression of top 100 different expressed genes in 4 main clusters. (E) Spearman correlation between areas used for tumor versus peritumor area comparison. (F) Heatmap showing expression levels of specific markers in 4 clusters.
Figure 3

Spatial gene expression comparison of tumor cluster A and cluster B. (A) Gene expression in the tumor cluster A. (B) Gene expression in the tumor cluster B. (C) and (D) Violin plots displaying the expression of representative marker gene identified in tumor and peritumor clusters. (E) Pathway enriched in tumor and peritumor clusters.
Figure 4

Kaplan-Meier survival plot of top 6 marker genes of tumor clusters (A) Kaplan-Meier survival plot of PPIB. (B) Kaplan-Meier survival plot of UGT2B15. (C) Kaplan-Meier survival plot of IFI27. (D) Kaplan-Meier survival plot of NDRG1. (E) Kaplan-Meier survival plot of BHLHE40. (F) Kaplan-Meier survival plot of VEGFA.
Figure 5

Identification of the candidate spatial tumor cluster specific genes in the TCGA cohort. (A) Venn diagram to identify clusters specific marker genes (DEGs) between HCC tissues and adjacent normal tissues correlated with OS. (B) The expression of 7 spatial clusters specific marker genes between HCC tissues and adjacent normal tissues. (C) Forest plots showing the results of the univariate Cox regression analysis between gene expression and OS. (D) The correlation network of candidate genes.
Figure 6

Prognostic analysis of the 6-gene signature model in the TCGA cohort and ICGC cohort. TCGA cohort (A, B, E, F, I, J), ICGC cohort (C, D, G, H, K, L). (A, C) The median value and distribution of the risk scores. (B, D) The distribution of OS status. (E, G) PCA plot. (F, H) t-SNE analysis. (I, K) Kaplan-Meier curves for OS of patients in the high- and low-risk groups. (J, L) AUC time-dependent ROC curves for OS.
Figure 7

The risk score in different groups divided by clinical characteristics and results of the univariate and multivariate Cox regression analyses. TCGA cohort (A–D,H,J), ICGC cohort (E–F,I,K). (A, E) Age. (B, F) Gender. (C) Tumor stage. (D, G) Tumor grade. (H, I) OS-related factors were screened by Univariate Cox regression analyses. (J, K) OS-related factors were screened by Multivariate Cox regression analysis.
Figure 8

Immune status between different risk groups and the association between risk score and tumor microenvironment. TCGA cohort (A, C), ICGC cohort (B, D). (A, B) The scores of 16 immune cells and (C, D) 13 immune-related functions were showed in boxplots. (E) Comparison of the risk score in different immune infiltration subtypes. (D) The relationship between risk score and RNAss, DNAss, Stromal Score and Immune Score. P values were showed as: ns, not significant,*P < 0.05, **P < 0.01, ***P < 0.001.
Figure 9

Scatter plot of relationship between prognostic gene expression and drug sensitivity.

Supplementary Files

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

- SuppleFigLegends.docx
- SupplementaryFig.1.jpg
- SupplementaryFig.2.jpg
- SupplementaryFig.3.jpg
- SupplementaryFig.4.jpg
- SupplementaryFig.5.jpg
- SupplementaryFig.6.jpg
- SupplementaryFig.7.jpg
- SupplementaryTable1.docx