

ORIGINAL RESEARCH ARTICLE

A spatial transcriptome-based perspective on highly variable genes associated with the tumor microenvironment in hepatocellular carcinoma

Hang Yang* and Qi Nian Jiang

Department of Anesthesiology, Clinical College, Guizhou Medical University, Guiyang, China

Abstract

Hepatocellular carcinoma is a common pathologic and histologic subtype of primary liver cancer and a common cause of cancer-related death. The tumor microenvironment (TME) is involved in the composition and deterioration of the cancer pathological environment and thus understanding the components of the TME of hepatocellular carcinoma and the mechanisms involved will aid in the development of strategies for inhibiting hepatocellular carcinoma progression and metastasis. The spatial transcriptomic data were used to uncover specific cell subclusters based on Seurat's process, followed using Monocle to perform pseudotemporal analysis. Spatial gene pathways were analyzed using the gene set variation analysis (GSVA) package process, and single-cell sequencing data were processed using Seurat. Gene sets were scored using the AddModuleScore approach. Quantitative polymerase chain reaction and immunohistochemistry as well as molecular docking assays were used to determine the biological effects of S100A6. The tumor tissue region revealed a specific subpopulation called cluster 4, which was rich in cancer-associated fibroblasts, particularly myofibroblasts. GSVA analysis suggested that this subpopulation was involved in the epithelial-mesenchymal transition process, and Kyoto Encyclopedia of Genes and Genomes analysis uncovered its involvement in the composition of extracellular matrix components. Pathological sections of the spatial transcriptome revealed that this subpopulation was specifically expressed at the tumor's broad envelope boundary and in stroma-rich locations, as confirmed with subsequent sections. The most prognostically relevant gene, *S100A6*, was detected using TCGA sequencing, and expression patterns and tissue locations were consistent with the similar trends of this subpopulation. In conclusion, we identified subgroups with tissue-specific expression patterns and marker genes that are most substantially associated with cancer prognosis, allowing us to gain a more comprehensive understanding of the intricate compositional alterations in the TME.

*Corresponding author:

Hang Yang
(hangyang1999@163.com)

Citation: Yang H, Jiang QN. A spatial transcriptome-based perspective on highly variable genes associated with the tumor microenvironment in hepatocellular carcinoma. *Cancer Plus*. 2023;5(4):1917.
doi: 10.36922/cp.1917

Received: August 26, 2023**Accepted:** November 2, 2023**Published Online:** November 18, 2023

Copyright: © 2024 Author(s). This is an Open-Access article distributed under the terms of the Creative Commons Attribution License, permitting distribution, and reproduction in any medium, provided the original work is properly cited.

Publisher's Note: AccScience Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Keywords: Spatial transcriptome; Hepatocellular carcinoma; Multidimensional integrated analysis; S100A6

1. Introduction

The majority of primary liver malignancies can be histologically classified as hepatocellular carcinoma, which contributes to most of the cancer-related mortalities globally. Hepatocellular carcinoma has a rather bad prognosis since it has a great

potential to recur and spread, which contributes to its poor prognosis. The American Cancer Society projects that hepatocellular carcinoma will cause 29,380 deaths in the United States by 2023.¹

Cancer growth is defined as a process of ongoing evolution of tumor and persistent alterations in response to external factors.² This involves the dynamic interaction between cancer cells and tumor microenvironment (TME), which consists of all non-cancerous cells of the tumor with a high concentration of fibroblasts, endothelial cells, immune cells, cytokines, and other components.^{3,4} Cancer-associated fibroblasts (CAFs) can modify the extracellular matrix (ECM) in the TME and support tumor progression through the secretion of different growth factors, chemokines, and cytokines.⁵ Given the specific expression and heterogeneity of marker genes owned by CAFs across various tumor types, their biological significance warrants further investigations.⁶⁻⁸

The recent advancements in single-cell and spatial transcriptome technologies have enabled us to explore TME in diverse dimensions of tissue regions and cell subclusters.^{9,10} A previous study has shown that spatial transcriptomes are useful for colorectal cancer research.¹¹ The present study seeks to identify unique CAF subclusters in the liver cancer TME by means of spatial transcriptomics and to validate findings using multidimensional multiomics data to provide new insights for targeted therapy in liver cancer.

2. Materials and methods

2.1. Tissue sectioning and data collection

To investigate the changes in spatial transcriptomes of genes in pathological tissue, we first collected data from published literature^{12,13} of pathological tissue section samples from one patient (HCC1), including normal tissue (HCC1-N), tumor tissue with an intact envelope (HCC1-L), and tumor tissue sections with an incomplete envelope (HCC1-T). We also gathered additional transcriptomic data¹⁴ to confirm our findings. The liver cancer cohort we utilized was collected from TCGA-LIHC and is available for download through the UCSC XENA website (<https://xena.ucsc.edu/>). RNA sequences of single cells can be obtained from the TISCH website (tisch.comp-genomics.org/).¹⁵

2.2. Spatial transcriptomic data quality control and microenvironment identification in tissue regions

According to the Seurat package's procedure,¹⁶ we first performed a quality-control check on the spatial transcriptome to eliminate spots that were not in the

tissue section region and then on the regions with high mitochondrial expression before conducting a principal component analysis and visualizing them using UMAP plots. Using the SPARK package,¹⁷ genes with spatially high expression patterns were employed as a means for distinguishing between malignant and normal tissues. We measured the cellular composition of distinct locations based on the collection of marker genes (see "Signature genes" in Table A1) to better understand the TME.

2.3. Tumor regional subgroup clustering analysis and microenvironmental mapping identification

We re-clustered the tumor locations, collected data for scale, performed an analysis with the function RUNPCA, and then used UMAP and T-SNE to display our subgroup distribution. The tool FindAllMarkers was used to determine the top five highly variable genes of the subgroup, and violin plots were used to display the marker genes of certain cells, such as inflammatory CAFs (iCAFs) and myofibroblastic CAFs (myoCAFs).

2.4. Pseudotime series analysis of tumor tissue development

The Monocle package¹⁸ was used to conduct pseudotime analysis of sub-tumor tissue areas to explore the differentiation of distinct subgroups through time, with a focus on the expression patterns of highly variable genes and marker genes linked with the epithelial-mesenchymal transition (EMT) process.

2.5. Analysis of spatial gene enrichment pathways and biological pathways

Using the gene set variation analysis (GSVA) package's ssGSEA method,¹⁹ we investigated the highly active signaling pathways in the tumor location. Furthermore, using the Sanger Assistant website (<http://vip.sangerbox.com/essay.html>), we performed Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis of the gene set function and mapped the gene coordinates identified from tissue sections to a two-dimensional coordinate system using the function SpatialPlot ("Sequencing coordinate" Supplementary File).

2.6. Scoring and multidimensional joint validation of TME-enriched gene sets

We identified cluster 4 as a subgroup enriched in active TME and obtained a set of 56 genes by searching for high-variance genes in cluster 4 and adjusting the absolute value of log2FC to be >1 and the *P*-value to be <0.01, which we termed the "spatial tumor microenvironment gene set" (STME). Using the AddModuleScore function of the Seurat package, we scored the gene set in additional slices

to confirm our findings. We mapped gene sets to single-cell datasets by TISCH, one with the immune-treated GSE125449 dataset and the other without GSE16635.

2.7. Identification of TME-related genes with differential prognosis and spatial expression patterns

Survival curves and receiver operating characteristic (ROC) curves were generated using the survival package and the survminer package (<https://www.bioconductor.org/>). Univariate Cox analysis was used to identify prognosis-related genes. Intersection with differential genes was used to calculate the correlation between genes, and finally, the most significant gene, *S100A6*, was identified and displayed using the SpatialFeaturePlot function.

2.8. Quantitative reverse-transcription polymerase chain reaction (qRT-PCR)

Human hepatocellular carcinoma cell lines (Huh-7, HepG2, MHCC-97L) and human normal hepatocyte lines (LO-2) were purchased from Platzer Biologics. Trizol (500 μ L) was added to the cell or tissue sample. A total of 100 μ L chloroform were added to a 1.5 mL Eppendorf tube spiked with Trizol, shaken for mixing, and then let to stand for 5 min before centrifugation at 4°C for 10 min at 12,000 rpm. The upper colorless and clear aqueous layer were aspirated from the spun tube and transferred into another clean 1.5 mL Eppendorf tube. An equal volume of isopropanol was added, followed by mixing through gentle inversion of the tube. The tube was left to stand for 10 min before centrifugation at 4°C for 10 min at 12,000 rpm. After centrifugation, a gelatinous precipitate was seen on the wall or bottom of the tube; the supernatant was removed carefully and the pellet was kept for further experiments. On washing with 1 mL of 75% ethanol (in DEPC-treated water preparation), the RNA pellet was then centrifuged at 4°C and 7000 rpm for 5 min; after the centrifugation process, the supernatant was removed as much as possible. Twenty-five microliter of DEPC-treated water was added to all tubes to fully dissolve the RNA; the RNA solution samples were then stored at -80°C. The reverse transcription reaction solution was prepared separately, and the cDNA was immediately experimentally synthesized and then stored at 4°C. The *S100A6* primer sequences are as follows: forward, 5'-CTGACTTCAACAGCGACACC-3' and reverse, 5'-GTGGTCCAGGGTCTTACTC-3'. *GAPDH* was used as a standardized control gene. The *GAPDH* primer sequences are as follows: forward, 5'-AGCAGACATTGACCTCACCA-3' and reverse, 5'-CCTCTATGGGCCCGAATCTT-3'. The melting amplification curves of the experiments and the raw data of the experiments are presented in the Supplementary File.

2.9. Immunohistochemistry and molecular docking experiments

Immunohistochemistry images were acquired from the Human Protein Atlas database (<https://www.proteinatlas.org/>). The antibody numbers used are Antibody CAB040549 and Antibody CAB002601. To assess the binding energies and interaction patterns between drug candidates/small molecules and their targets, information about their molecular structures was obtained from PubChem (<https://pubchem.ncbi.nlm.nih.gov/>). 3D coordinates of the *S100A6* protein and the drug sorafenib were downloaded from the protein database (<http://www.rcsb.org/>). We first prepared the protein and ligand files by converting all protein and molecule files to PDBQT format, removing all water molecules, and adding polar hydrogen atoms *in silico*. The grid boxes were centered to cover the structural domains of each protein and to accommodate free molecular motion. The docking pocket was set as a 30 Å × 30 Å × 30 Å square pocket with a grid point distance of 0.05 nm. Molecular docking studies were performed by Autodock Vina 1.2.2 (<http://autodock.scripps.edu/>) for model visualization.

2.10. Statistical analysis

All statistical analyses were performed using R language version 4.2.0, and $P < 0.05$ were deemed statistically significant for all comparisons.

3. Results

3.1. Landscape of the microenvironment in the tumor tissue region

We performed a quality control check on sections of HCC1-L patients with intact tumor envelope areas, which were observed to have high characteristics in the tumor region, and a preliminary descending grouping into four subgroups (Figure 1A). The SPARK package found differentially expressed genes with significant spatial patterns of regional expression. *APOE*, *CRP*, and other cancer-related genes were also identified (Figure 1B). Sections of tumor tissue and normal tissue were distinguished and separately visualized in terms of area annotation (Figure 1C). When tumor and normal tissue were scored for cell content using specific gene sets, we discovered that tumor tissue contained much larger amounts of immune cells than normal tissue, particularly CD4+ and CD8+ T cells, among others. Furthermore, we found that the tumor tissue had more M1-type macrophages than the normal tissue (Figure 1D). The main function of M1 macrophages is to phagocytose tumor cells, whereas anti-inflammatory M2 macrophages, such as tumor-associated macrophages, promote tumor growth and invasion.²⁰

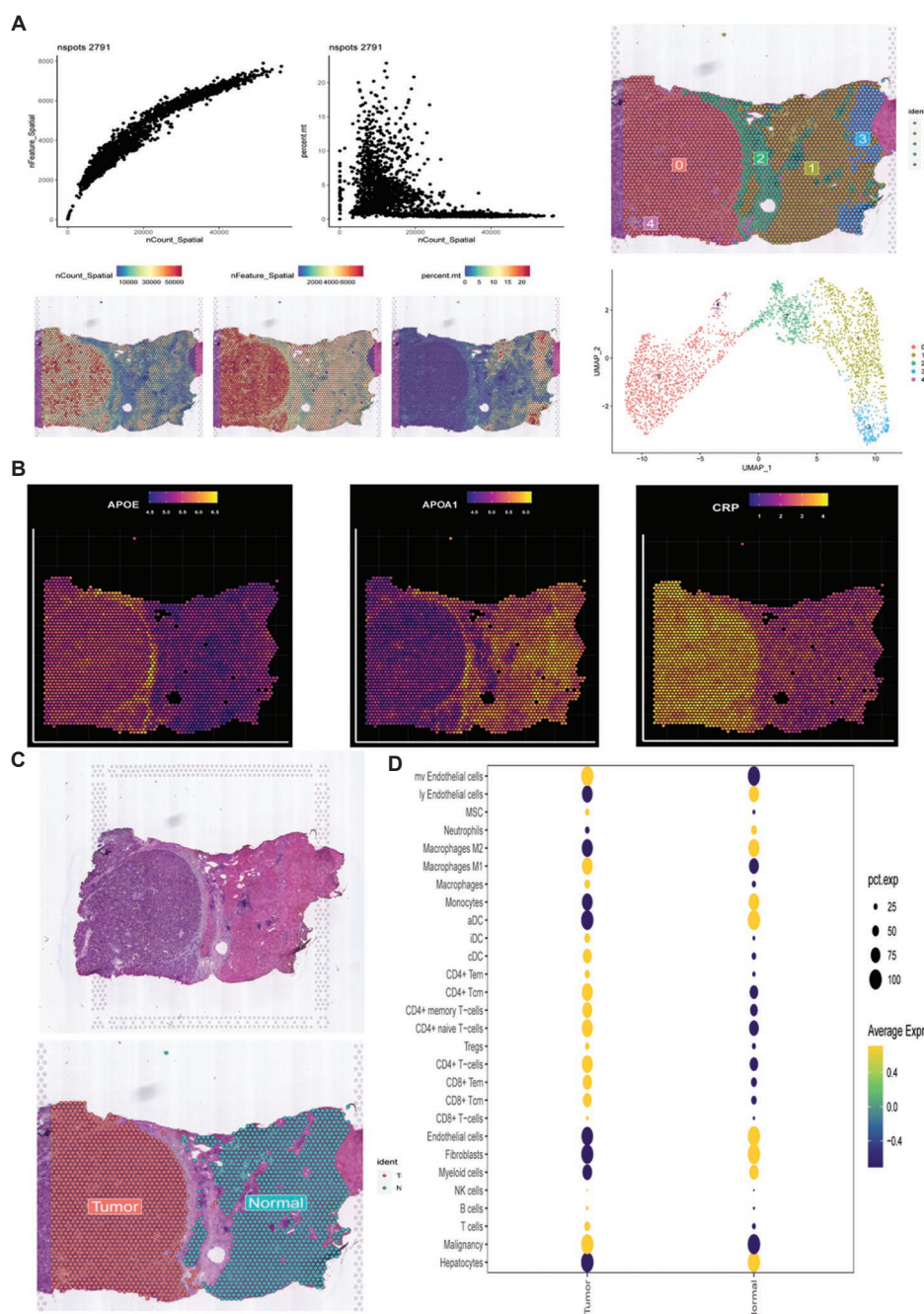


Figure 1. Spatial transcriptomic data quality control and spot cell subpopulation annotation. (A) Quality control and preliminary dimensionality reduction clustering of spatial transcriptomic data. (B) Expression regions of spatially highly variable genes such as *APOE*, *CRP*, and *APOA1*. (C) Spatial annotation of pathological tissue sections. (D) Visualization of cell types based on signature score.

3.2. Identification of a subpopulation of spots rich in TME

We produced six spot subgroups by clustering the tumor area again, and their distribution is displayed in the image, with cluster 4 being enriched in the boundary region of the tumor (Figure 2A). Similarly, we assessed the cellular enrichment of

the subgroups and discovered that cluster 4 was enriched in myCAFs and iCAFs, as well as macrophages, cell types which are associated with cancer spread and immune response (Figure 2B). The heat map depicts the highly variable marker genes in each subpopulation, where cluster 4 exhibited a high expression of myCAF marker genes *ACTA2*, *COL1A1*, and other fibroblast marker genes, and we used violin plots to

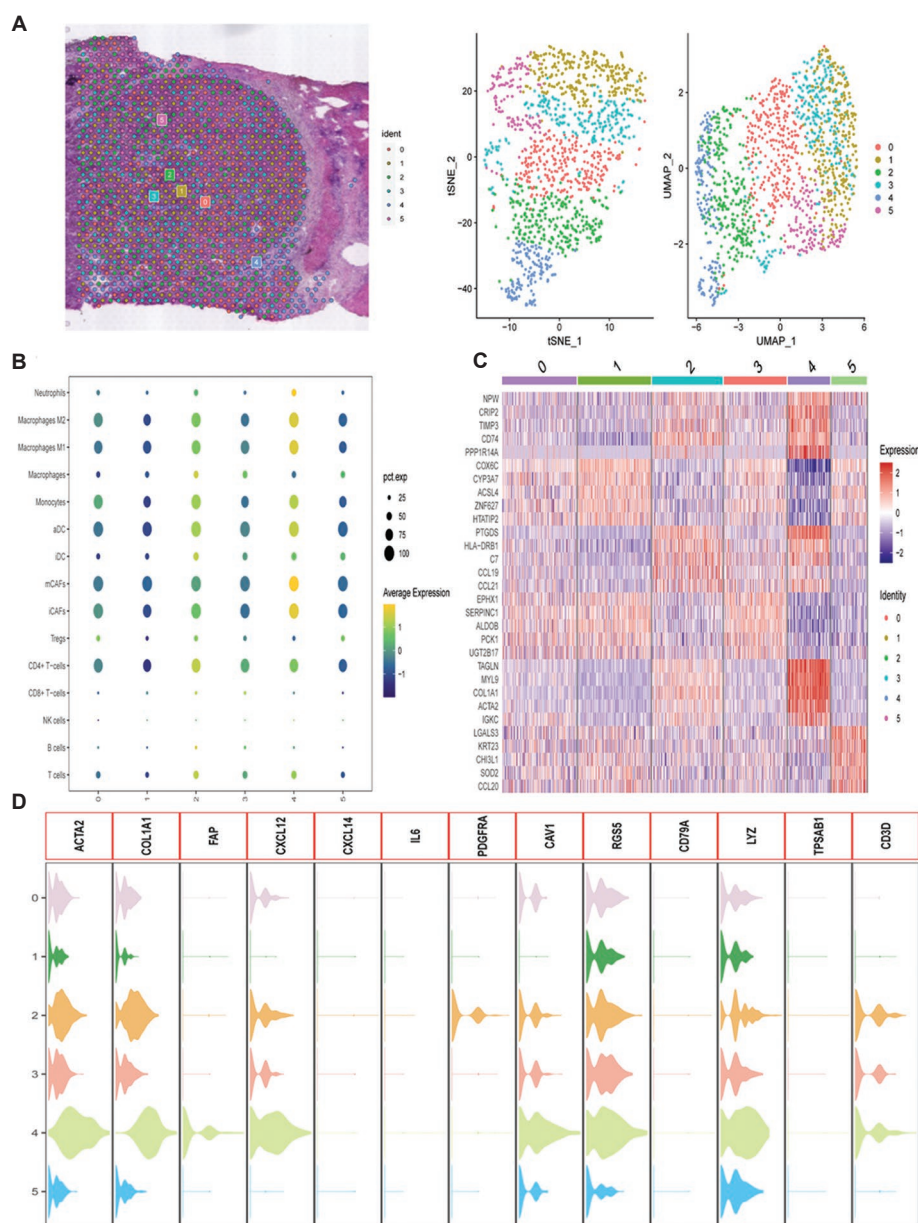


Figure 2. Tumor microenvironment and highly variable genes in tumor subclusters. (A) Results of regional re-clustering of tumor tissues. (B) Map of differences in cell content among different subgroups located in the tumor area. (C) Top five significant marker genes for each subgroup of tumors. (D) Visualization of marker gene expression in myofibroblasts and immune-associated fibroblasts.

better analyze the expression patterns of each marker gene in the subpopulation (Figure 2C and D). Our findings show that cluster 4, which is linked with fibroblasts, may have a role in cancer dissemination and progression.

3.3. Temporally distinct marker gene expression trends revealed by pseudotime series analysis

The developmental trajectory of the subgroup was revealed by pseudotemporal analysis, and we discovered

that the previously identified cluster 4 was at the most advanced stage of development, demonstrating the higher degree of malignancy of cluster 4 and being more consistent with previous findings (Figure 3A and B). Furthermore, we discovered that the myCAF cell marker gene *CDH11* and other genes connected to EMT were all strongly expressed in cluster 4 and steadily increased over time (Figure 3C-E). This is also supported by the heat map results (Figure 3F).

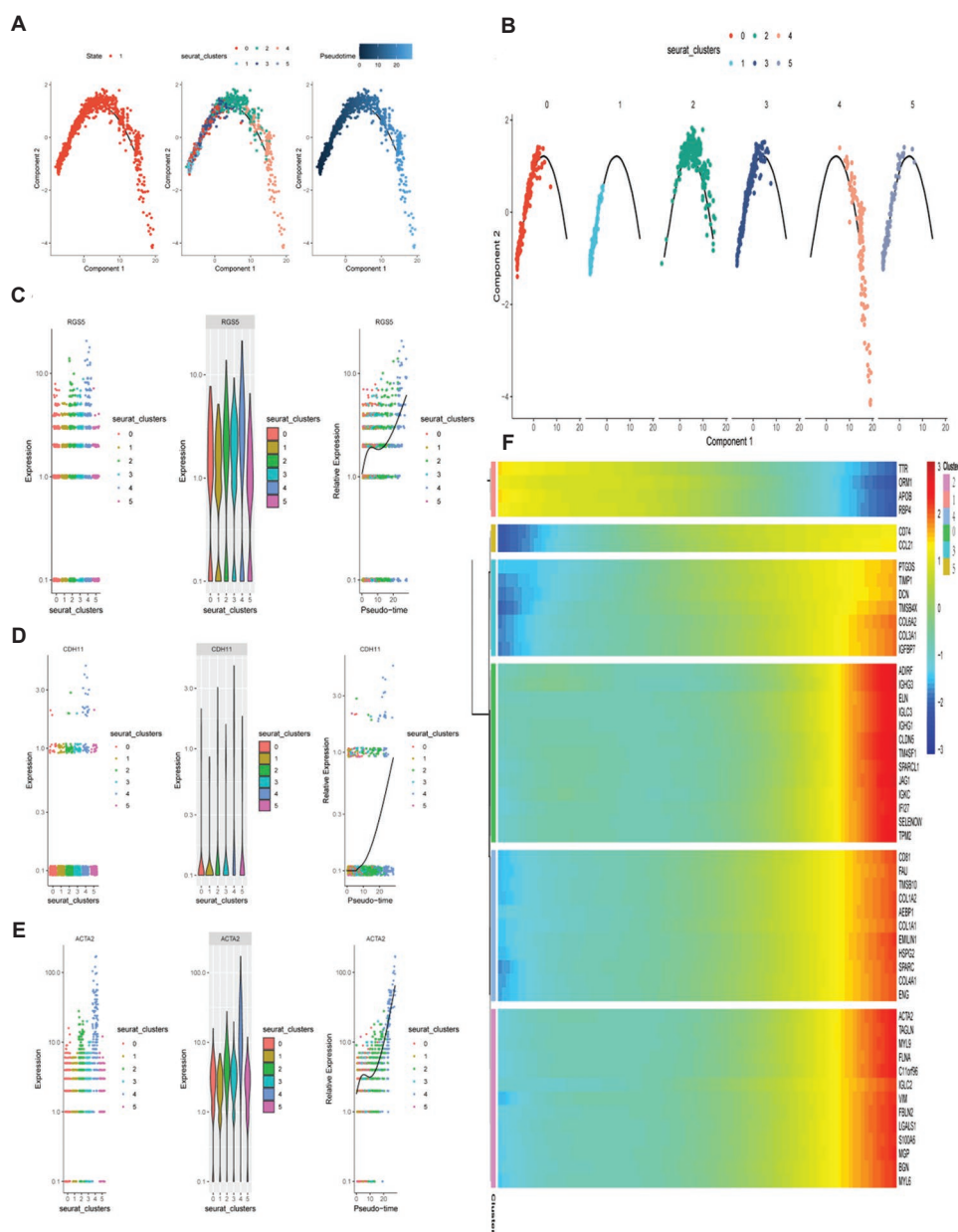


Figure 3. Temporal differentiation trajectory of tumor tissue subclusters. (A and B) The order of the developmental differentiation trajectory of subclusters over time, with cluster 4 positioned at a late stage. (C-E) Trends in the expression of CDH11, RGS5, ACTA2, and other signature proteins in various subgroups over time. (F) Heat map of highly variable gene expression.

3.4. Analysis of pathways and functionally active spatial gene sets

We discovered that oxidative stress was prevalent in the overall region, but the MYC pathway and the angiogenic pathway were more active in cluster 4, indicating that cluster 4 has an important role in tumor spreading (Figure 4A). Furthermore, we represented the route activity as two-dimensional coordinates by substituting

the coordinate information with the original position of the section, and these pathways were likewise enriched in the tumor's marginal envelope region with the maximum overlap with the position of cluster 4 (Figure 4B-D). KEGG analysis revealed four subgroups of gene enrichment pathways where ECM-receptor interaction pathway (Figure 4E), lending credence to our conclusion.

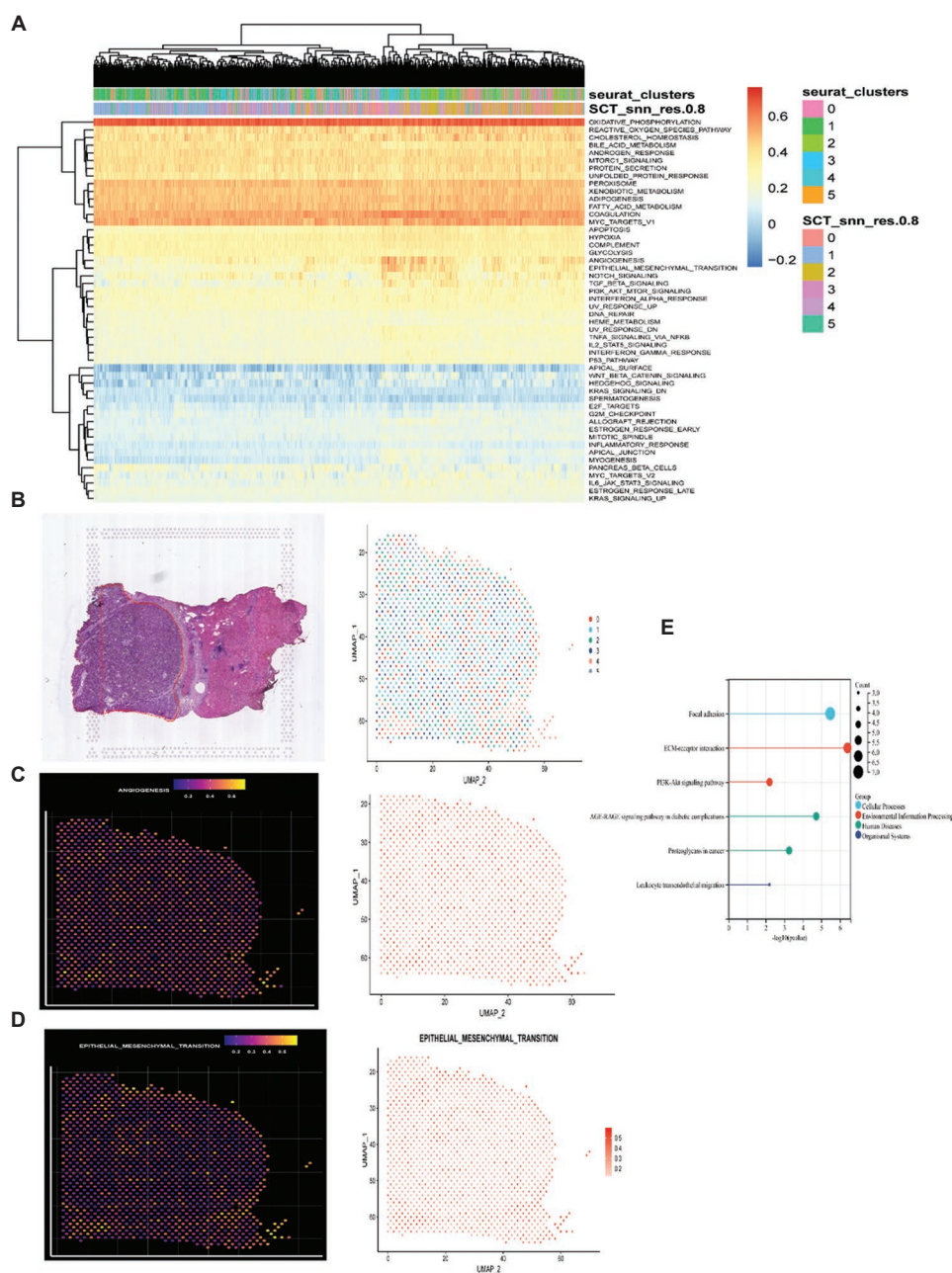


Figure 4. Analysis of gene function and biological pathway enrichment and activity. (A) Spatial gene set variation analysis, where MYC pathway and epithelial-mesenchymal transition pathway were significant in cluster 4. (B-D) Mapping the degree of variation in spatial signaling pathways by transforming spatial tissue slice information through a two-dimensional coordinate system. (E) Analysis of the Kyoto Encyclopedia of Genes and Genomes pathway in cluster 4.

3.5. Integration of different dimensions of histological data to validate cell types and enrichment regions determined by gene sets

By defining log2FC with absolute values larger than 1 and $P < 0.01$, we established a group of genes most substantially related with cluster 4 called STME. We discovered that the genes were more commonly expressed among

fibroblastic tissues by scoring normal sections and tumor sections with unclear tumor areas from the same patient (Figure 5A and B). Furthermore, by evaluating the single-cell dataset, we found that our gene ensemble was enriched in fibroblasts, consistent with the spatial transcriptome findings, and in the immunotherapy dataset, patients receiving PDL1 immunotherapy had lower gene scores (Figure 5C and D).

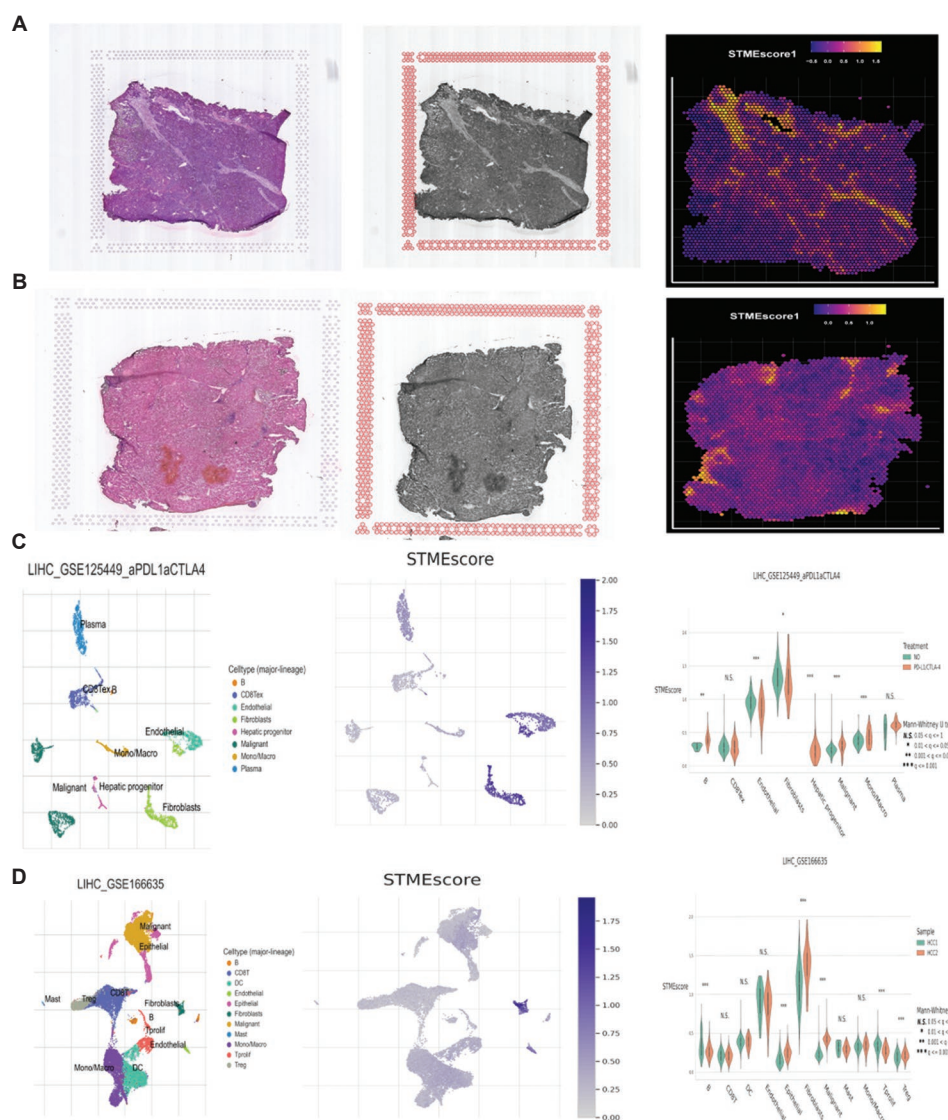


Figure 5. Validation of findings with additional pathology sections and single-cell sequencing data. (A) Visualization of subgroup expression with blurred tumor region boundaries. (B) Visualization of subgroup gene expression in normal tissues. (C and D) Validation of gene scores based on data from immunotherapy cohorts of single cells. Higher STME scores in endothelial and fibroblast subpopulations were detected.

3.6. Identification of prognosis-related STME genes and expression patterns

In the TCGA cohort, we identified genes associated with prognosis through univariate Cox analysis of the STME gene set, and we demonstrated with ROC curves and survival curves that the scores could be used to evaluate the overall survival of patients: the higher the score, the shorter the survival of patients. In addition, the ROC curves indicated a higher efficacy of our gene set (Figure 6A, B, D, and E). *S100A6* was strongly correlated with the TME score in the TCGA cohort (Figure 6F-I), and to confirm the expression pattern of

S100A6 gene, we selected additional sections to visualize its expression region in additional patients, and it was evident that *S100A6* was significantly expressed in the marginal region of the tumor and in the fibroblastic tissue (Figure 6C, J, and K). Therefore, *S100A6* may serve as a marker gene for activated fibroblasts and cancer regions.

3.7. Biological experimental validation and molecular docking analysis

To assess the true expression differences of *S100A6*, we performed immunohistochemical experiments for analysis. Immunohistochemical data showed that *S100A6*

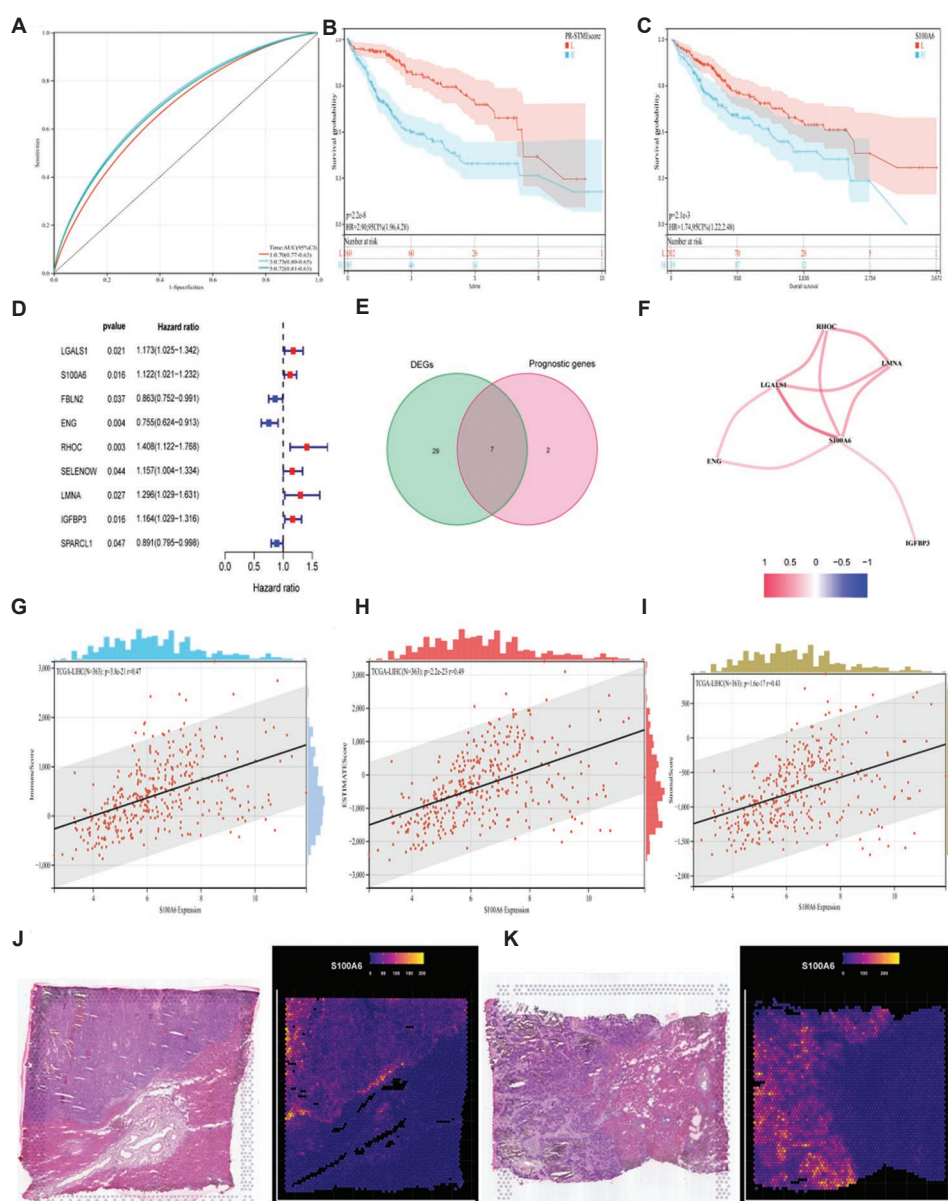


Figure 6. Identification of marker genes that are most significantly associated with the tumor microenvironment and have prognostic roles. (A and B) Scoring of tumor microenvironment-associated gene sets that have a prognostic role. (C-F) The S100A6 gene was differentially expressed across different tissues and had a prognostic effect, with significance in the entire gene set. (G-I) S100A6 had a significant positive relationship with tumor microenvironment score. (J and K) S100A6 was highly expressed in tumor tissue and located in fibroblasts, with a close association with tumor stromal composition.

expression was higher in tumor tissues than in normal tissues and was mostly located in the stromal region (Figure 7A and B). In addition, qRT-PCR results also showed that *S100A6* expression was significantly higher and varied across the three cell lines of hepatocellular carcinoma (Figure 7C). To evaluate the targeting effect of the available first-line therapeutic agents on the biomarkers we screened, we performed molecular docking assays. The hydrophobic pocket of the S100A6 target was successfully occupied by five drug candidates. For S100A6, sorafenib

has a low-binding energy of -7.982 , indicating highly stable binding (Figure 7D and E).

4. Discussion

The physiological status of the TME is intimately related to cancer progression. Despite tumor heterogeneity, comparable characteristics of the TME maturation environment of epithelial-derived malignancies can be detected. Genetic alterations in tumor cells contribute to hypoxia and oxidative stress in the microenvironment,

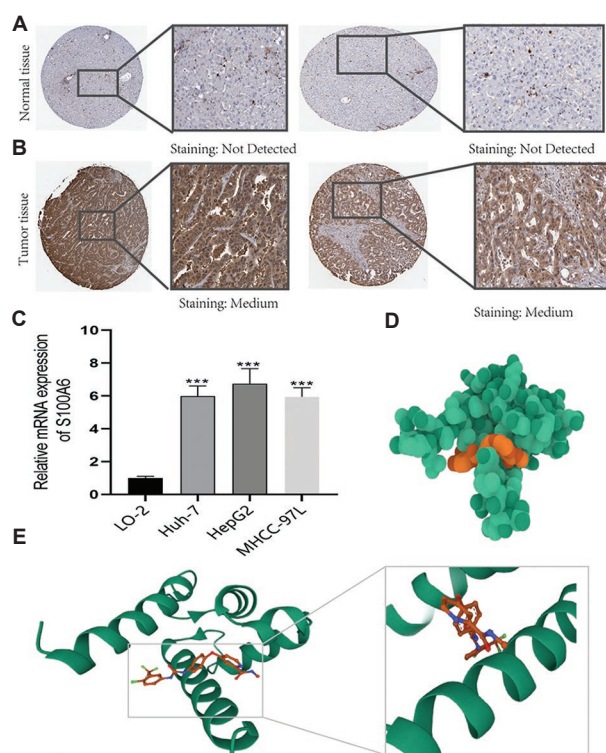


Figure 7. Biological experimental validation and molecular docking. (A and B) Immunohistochemical analysis: The left panel shows the overall view and the right one shows the view in local magnification. (C) Results of qRT-PCR experiments on cell lines. (D and E) Schematic diagrams of molecular docking, ball-and-stick model of S100A6 and sorafenib binding, and molecular model.

causing alterations in the ECM, reactions from surrounding stromal cells (fibroblasts), and immune cell responses, which then stimulate angiogenesis and, eventually, tumor cell spread and invasion.^{21,22} With the development of spatial transcriptomic technology, we can now visualize gene expression patterns on slices as well as the tissue regions of expression to better understand the TME. This is a significant advantage over single-cell and bulk transcriptome sequencing, allowing for the visible recognition of spatial transcriptomic differences.

In this investigation, we included three separate sections of normal tissue from the same patient, as well as tissue with intact tumor envelope boundaries and tissue with ambiguous tumor envelope boundaries. Through a more thorough exploration of gene expression patterns in various disease processes, we were able to define tumor region structure by pinpointing spatially highly variable genes and descending clusters. The tumor tissue areas were infiltrated with immune cells to a greater extent and possessed larger amounts of M1-type macrophages. It has been reported that M1-type macrophages had an inhibitory effect on tumor metastasis,^{23,24} which may be

one of the reasons for the intact tumor envelope in the patient, showing a high immune response and active macrophage function to inhibit cancer invasion and metastasis. In our re-clustering of tumor locations, we discovered a distinct cell subpopulation cluster 4, which expresses the fibroblast marker gene *COL1A1* as well as the myofibroblastic cancer cell marker *ACTA2* and *MYL9*. It also has greater M2 macrophage subpopulation expression. MyoCAFs cells have been shown to connect with immunological inflammatory cells to promote cancer spread,^{25,26} and CAFs have been shown to interact with tumor-associated macrophages to decrease the value added of natural killer cells,²⁷ which are similar with the behavior we discovered. To further investigate the developmental trajectory of each subpopulation, we discovered that cluster 4 was more malignant at the most advanced stage of development and was located in the border region of the section, where the temporal expression trends of CAF marker genes also supported our conclusion that cluster 4 is indicative of a unique genotype and was expressed at the tumor's border stromal region. When we visualized the coordinates of Subgroup 4, we found that they were located at the very edge of the complete tumor envelope, and from Figure 4C and D, we found that the pathways associated with tumor invasion and metastasis were also enriched here and highly overlapped with Subgroup 4. Therefore, we suggest that cluster 4 may play a role aiding in tumor invasion. The integrity of the tumor fiber envelope in the border region of primary hepatocellular carcinoma is closely related to the spatial heterogeneity of tumor cells and the distribution of their surrounding stromal cells and immune cells. An intact fibrous envelope composed mainly of fibroblasts and endothelial cells may serve as a barrier to immune cell infiltration.

GSVA analysis of the spatial transcriptome revealed that the oxidative stress pathway was significantly activated in almost all tumor locations, but only four subclusters possessed active EMT and MYC pathways. Previous studies have revealed complex crosstalk between the MYC pathway and EMT.²⁸⁻³⁰ Results from KEGG analysis and 2D coordinate information also demonstrate the high activity of our gene set in the ECM and cancer-related pathways. We used AddModuleScore to map the cluster 4 gene set to normal tissues and tumor tissues with incomplete borders and discovered that they were strongly expressed in the border stromal region and that single-cell sequencing had superior precision. Our gene set was more highly expressed in the fibroblast subpopulation and scored lower in immunotherapy patients, corroborating our hypothesis that the gene set has higher expression in the fibroblast and stromal regions, although additional research is needed to corroborate this result.

We employed the TCGA cohort for in-depth analysis to determine the marker gene with the highest confidence. Using univariate Cox analysis, we discovered prognosis-related genes, and by considering intersections with TCGA differentially expressed genes and based on gene interactions, we identified the most significant gene, *S100A6* (S100 calcium-binding protein A6), which was shown to be connected with prognosis by survival curve analysis. Previous studies have reported that *S100A6* promotes proliferation and migration in hepatocellular carcinoma and gastric cancer,³¹⁻³⁴ but there are few studies on the relationship between *S100A6* and hepatocellular carcinoma fibroblasts. Previous single-cell sequencing studies demonstrated that the *SMA* (*ACTA2*) and *S100A6* genes, which are stably expressed, can be employed as myofibroblast markers.³⁵ We confirmed this with spatial transcriptomic data analysis and additional sections, and we showed for the 1st time that the *S100A6* expression is localized to the tumor tissue's border stromal region, demonstrating the stability of spatial transcriptome technology and the high accuracy of single-cell transcriptome. Finally, *S100A6* is heavily engaged in the makeup of the complex TME and is strongly connected with CAFs,³⁶ suggesting that it could be used as a therapeutic target to slow down cancer growth.

To further validate our findings, we performed experimental validation of the cell lines and found that the expression of *S100A6* was significantly higher in the hepatocellular carcinoma cell lines, especially in the Hepg2 cell line than in the normal group, signifying that the deeper immunohistochemical staining results in the cell stromal region are consistent with our spatial transcriptomic data. The molecular docking³⁷ assay allowed us to screen the binding ability of the existing first-line therapeutic agent sorafenib³⁸ for our target, suggesting certain implications for clinical translation.

This study has several limitations. For example, the accuracy of the spatial transcriptome analysis cannot be as precise as that of a single-cell analysis, because the former involves 10 – 20 cells inside a single spot. Therefore, multidimensional validation is needed, in addition to applying greater cohort and idle slice data to validate our findings.

5. Conclusion

Our investigation revealed a unique TME profile in hepatocellular carcinoma, as well as the interactions among CAF subclusters using spatial transcriptome combined with single-cell and multidimensional histology, providing promising reference for future studies.

Acknowledgments

The authors would like to thank Dr. Xi for his outstanding teachings on spatial transcriptome technology; in reciprocation, we want to take this opportunity to publicize his public WeChat account “Spatial Transcriptome of Tomatoes” for broader recognition. The first author Hang Yang extends his gratitude to his classmate Qiniang, his instructor Dr. Lei, and his mother, who has been the spiritual support behind his works.

Funding

None.

Conflict of interest

The authors declare that they have no competing interests.

Author contributions

Conceptualization: Hang Yang

Data curation: All authors

Formal analysis: Hang Yang

Investigation: All authors

Methodology: Hang Yang

Writing – original draft: Qi Nian Jiang

Writing – review & editing: All authors

All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

This study is based on open-source data stemming from TCGA and HPA, which are public databases.

Availability of data

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

References

1. Siegel RL, Miller KD, Wagle NS, Jemal A. Cancer statistics, 2023. *CA Cancer J Clin.* 2023;73:17-48.
doi: 10.3322/caac.21763
2. Merlo LM, Pepper JW, Reid BJ, Maley CC. Cancer as an evolutionary and ecological process. *Nat Rev Cancer.* 2006;6:924-935.
doi: 10.1038/nrc2013
3. Chen X, Song E. Turning foes to friends: Targeting cancer-associated fibroblasts. *Nat Rev Drug Discov.* 2019;18:99-115.
doi: 10.1038/s41573-018-0004-1

4. Hinshaw DC, Shevde LA. The tumor microenvironment innately modulates cancer progression. *Cancer Res.* 2019;79:4557-4566.
doi: 10.1158/0008-5472.CAN-18-3962
5. Jin MZ, Jin WL. The updated landscape of tumor microenvironment and drug repurposing. *Signal Transduct Target Ther.* 2020;5:166.
doi: 10.1038/s41392-020-00280-x
6. Chen Z, Zhou L, Liu L, *et al.* Single-cell RNA sequencing highlights the role of inflammatory cancer-associated fibroblasts in bladder urothelial carcinoma. *Nat Commun.* 2020;11:5077.
doi: 10.1038/s41467-020-18916-5
7. Elyada E, Bolisetty M, Laise P, *et al.* Cross-species single-cell analysis of pancreatic ductal adenocarcinoma reveals antigen-presenting cancer-associated fibroblasts. *Cancer Discov.* 2019;9:1102-1123.
doi: 10.1158/2159-8290.CD-19-0094
8. Kamali Zonouzi S, Pezeshki PS, Razi S, Rezaei N. Cancer-associated fibroblasts in colorectal cancer. *Clin Transl Oncol.* 2022;24:757-769.
doi: 10.1007/s12094-021-02734-2
9. Lei Y, Tang R, Xu J, *et al.* Applications of single-cell sequencing in cancer research: Progress and perspectives. *J Hematol Oncol.* 2021;14:91.
doi: 10.1186/s13045-021-01105-2
10. Zhang L, Chen D, Song D, *et al.* Clinical and translational values of spatial transcriptomics. *Signal Transduct Target Ther.* 2022;7:111.
doi: 10.1038/s41392-022-00960-w
11. Peng Z, Ye M, Ding H, Feng Z, Hu K. Spatial transcriptomics atlas reveals the crosstalk between cancer-associated fibroblasts and tumor microenvironment components in colorectal cancer. *J Transl Med.* 2022;20:302.
doi: 10.1186/s12967-022-03510-8
12. Golubnitschaja O, Baban B, Boniolo G, *et al.* Medicine in the early twenty-first century: Paradigm and anticipation-EPMA position paper 2016. *Epma J* 2016;7:23.
doi: 10.1186/s13167-016-0072-4
13. Wu R, Guo W, Qiu X, *et al.* Comprehensive analysis of spatial architecture in primary liver cancer. *Sci Adv.* 2021;7:eabg3750.
doi: 10.1126/sciadv.abg3750
14. Wu Y, Yang S, Ma J, *et al.* Spatiotemporal immune landscape of colorectal cancer liver metastasis at single-cell level. *Cancer Discov.* 2022;12:134-153.
doi: 10.1158/2159-8290.CD-21-0316
15. Sun D, Wang J, Han Y, *et al.* TISCH: A comprehensive web resource enabling interactive single-cell transcriptome visualization of tumor microenvironment. *Nucleic Acids Res.* 2021;49:D1420-D1430.
doi: 10.1093/nar/gkaa1020
16. Hao Y, Hao S, Andersen-Nissen E, Mauck WM 3rd, *et al.* Integrated analysis of multimodal single-cell data. *Cell.* 2021;184:3573-3587.e29.
doi: 10.1016/j.cell.2021.04.048
17. Sun S, Zhu J, Zhou X. Statistical analysis of spatial expression patterns for spatially resolved transcriptomic studies. *Nat Methods.* 2020;17:193-200.
doi: 10.1038/s41592-019-0701-7
18. Qiu X, Hill A, Packer J, Lin D, Ma YA, Trapnell C. Single-cell mRNA quantification and differential analysis with Census. *Nat Methods.* 2017;14:309-315.
doi: 10.1038/nmeth.4150
19. Hänzelmann S, Castelo R, Guinney J. GSEA: Gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics.* 2013;14:7.
doi: 10.1186/1471-2105-14-7
20. Xia Y, Rao L, Yao H, Wang Z, Ning P, Chen X. Engineering macrophages for cancer immunotherapy and drug delivery. *Adv Mater.* 2020;32:e2002054.
doi: 10.1002/adma.202002054
21. Roma-Rodrigues C, Mendes R, Baptista PV, Fernandes AR. Targeting tumor microenvironment for cancer therapy. *Int J Mol Sci.* 2019;20:840.
doi: 10.3390/ijms20040840
22. Wu T, Dai Y. Tumor microenvironment and therapeutic response. *Cancer Lett.* 2017;387:61-68.
doi: 10.1016/j.canlet.2016.01.043
23. Gunassekaran GR, Poongkavithai Vadevoo SM, Baek MC, Lee B. M1 macrophage exosomes engineered to foster M1 polarization and target the IL-4 receptor inhibit tumor growth by reprogramming tumor-associated macrophages into M1-like macrophages. *Biomaterials.* 2021;278:121137.
doi: 10.1016/j.biomaterials.2021.121137
24. Yunna C, Mengru H, Lei W, Weidong C. Macrophage M1/M2 polarization. *Eur J Pharmacol.* 2020;877:173090.
doi: 10.1016/j.ejphar.2020.173090
25. Mellone M, Piotrowska K, Venturi G, *et al.* ATM regulates differentiation of myofibroblastic cancer-associated fibroblasts and can be targeted to overcome immunotherapy resistance. *Cancer Res.* 2022;82:4571-4585.
doi: 10.1158/0008-5472.CAN-22-0435
26. Sun X, He X, Zhang Y, *et al.* Inflammatory cell-derived

- CXCL3 promotes pancreatic cancer metastasis through a novel myofibroblast-hijacked cancer escape mechanism. *Gut*. 2022;71:129-147.
doi: 10.1136/gutjnl-2020-322744
27. Mao X, Xu J, Wang W, *et al.* Crosstalk between cancer-associated fibroblasts and immune cells in the tumor microenvironment: New findings and future perspectives. *Mol Cancer*. 2021;20:131.
doi: 10.1186/s12943-021-01428-1
 28. Jiang Y, Han Q, Zhao H, Zhang J. Promotion of epithelial-mesenchymal transformation by hepatocellular carcinoma-educated macrophages through Wnt2b/ β -catenin/c-Myc signaling and reprogramming glycolysis. *J Exp Clin Cancer Res*. 2021;40:13.
doi: 10.1186/s13046-020-01808-3
 29. Li Z, Liu J, Chen T, *et al.* HMGA1-TRIP13 axis promotes stemness and epithelial mesenchymal transition of perihilar cholangiocarcinoma in a positive feedback loop dependent on c-Myc. *J Exp Clin Cancer Res*. 2021;40:86.
doi: 10.1186/s13046-021-01890-1
 30. Xia P, Zhang H, Xu K, *et al.* MYC-targeted WDR4 promotes proliferation, metastasis, and sorafenib resistance by inducing CCNB1 translation in hepatocellular carcinoma. *Cell Death Dis*. 2021;12:691.
doi: 10.1038/s41419-021-03973-5
 31. Delangre E, Oppliger E, Berkcan S, Gjorgjieva M, Correia de Sousa M, Foti M. S100 proteins in fatty liver disease and hepatocellular carcinoma. *Int J Mol Sci*. 2022;23.
doi: 10.3390/ijms231911030
 32. Li Z, Tang M, Ling B, *et al.* Increased expression of S100A6 promotes cell proliferation and migration in human hepatocellular carcinoma. *J Mol Med (Berl)*. 2014;92:291-303.
doi: 10.1007/s00109-013-1104-3
 33. Song D, Xu B, Shi D, Li S, Cai Y. S100A6 promotes proliferation and migration of HepG2 cells via increased ubiquitin-dependent degradation of p53. *Open Med (Wars)*. 2020;15:317-326.
doi: 10.1515/med-2020-0101
 34. Wang XH, Zhang LH, Zhong XY, *et al.* S100A6 overexpression is associated with poor prognosis and is epigenetically up-regulated in gastric cancer. *Am J Pathol*. 2010;177:586-597.
doi: 10.2353/ajpath.2010.091217
 35. Xia P, He H, Kristine MS, *et al.* Therapeutic effects of recombinant human S100A6 and soluble receptor for advanced glycation end products(sRAGE) on CCl(4)-induced liver fibrosis in mice. *Eur J Pharmacol*. 2018;833:86-93.
doi: 10.1016/j.ejphar.2018.05.030
 36. Krenkel O, Hundertmark J, Ritz TP, Weiskirchen R, Tacke F. Single cell RNA sequencing identifies subsets of hepatic stellate cells and myofibroblasts in liver fibrosis. *Cells*. 2019;8:503.
doi: 10.3390/cells8050503
 37. Dong D, Xu Z, Zhong W, Peng S. Parallelization of molecular docking: A review. *Curr Top Med Chem*. 2018;18:1015-1028.
doi: 10.2174/1568026618666180821145215
 38. Man S, Luo C, Yan M, Zhao G, Ma L, Gao W. Treatment for liver cancer: From sorafenib to natural products. *Eur J Med Chem*. 2021;224:113690.
doi: 10.1016/j.ejmech.2021.113690

Appendix

Table A1. Signature genes of indicated cell types

Cell type	Genes
Hepatocytes	HAMP, CYP1A2, APOE, CYP3A4, C9, ADH4, SLC22A1, CYP2A6, HGFAC, CYP2C8, CYP2E1, RDH16, SPP2, UROC1, AFM, PCK1, F9, CYP4A11, SERPINA11, APOA5, CYP8B1, SLC10A1
Malignancy	SPINK1, GPC3, AKR1B10, TOP2A, REG3A, CAP2, UBE2C, MDK, LCN2, AURKA
T cells	CD3D, CD3E, CD3G
B cells	CD19, MS4A1, CD79A
NK cells	NCAM1, KLRF1, NCR1, KLRC1
Myeloid cells	ITGAX, CD33, CEACAM8, CD68, CD163
Fibroblasts	COL1A2, FAP, PDPN, DCN, COL3A1, COL6A1, COL1A1
Endothelial cells	CLDN5, PECAM1, CD34, FLT1, VWF, ENG, CDH5
CD8+ T-cells	CD8A, CD8B
CD8+ Tcm	CD8A, CD8B, CTSW, GZMK, IL2RB, SH2D1A, GPR171, CRTAM, USP36, TMEM30B
CD8+ Tem	DHX8, GZMH, GZMK, LAG3, ZAP70, COLQ, RGS9, CXCR6, PVRIG, PYHIN1
CD4+ T-cells	CD2, CD3D, CD3E, CD3G, CD5, CD7, CD27, CD40LG, CTLA4, GOLGA4, TRAF1, CD96, ICOS, GPR171, TRAT1, FOXP3, SIRPG, RIC3
Tregs	CCR4, CCR8, CTLA4, GPR25, IL2RA, KCNA2, LAIR2, HS3ST3B1, MCF2L2, FOXP3
CD4+ naive T-cells	CD27, CLC, CTSW, DNAJB1, RBL2, HAUS3, ANKRD55, ZNF394, CHMP7
CD4+ memory T-cells	CD28, CD40LG, CCR4, CTLA4, GPR15, GZMA, GZMK, HMGB2, PDCD1, RBL2, CD226, GPR171, TRAT1, UBASH3A, ARHGAP15
CD4+ Tcm	CD40LG, CCR4, CCR8, CTLA4, DAB1, ERN1, GPR15, DNAJB1, KRT1, POU6F1, TPO, CDC14A, TRADD, DLEC1, ICOS, TRAT1, FXYP7, FBXL8, SIRPG, SLC4A5, ANKRD55, OBSCN
CD4+ Tem	CD40LG, CCR6, KLRB1, LTK, MCF2L2, GALNT8, TRAT1
cDC	CD1A, CD1B, CD1C, CD1E, CD80, CD86, FCER1A, GFRA2, CCL13, CCL17, CLEC10A, CD209, KCNK13
iDC	ALOX15, F13A1, CCL13, CCL17, CCL18, CCL23, CCL24, CD209
aDC	C1QA, C1QB, CD80, FPR3, HLA-DQA1, IL12B, CCL13, CCL17, CCL19, CCL22
Monocytes	AIF1, CYBB, FCAR, FCN1, MNDA, CFP, S100A12, CD163, CLEC5A, TREM1, RETN, FOLR2, MS4A6A, LILRA5
Macrophages	CHIT1, CYP19A1, HK3, MSR1, VSIG4, CLEC5A, ADAMDEC1, DNASE2B
Macrophages M1	ITGAM, CD68, FCGR3A, FCGR2A, FCGR1A, CD80, CD86, HLA-DRB5, TLR4, TNF, IL1A, IL1B, IL12, IL12A, IL12B, IL6, NOS2
Macrophages M2	ITGAM, CD68, CD163, MRC1, IL10, TGFB1, IDO1, IL8, CCL2
Neutrophils	CLC, CSF3R, FCGR3B, FPR2, HBB, CXCR2, S100A12, P2RY13, TREM1
MSC	HTR7, MMP17, PLA2G5, CDKL5, PKD2L1, ADAMTS12, ZNRF4, CTRB2
Lymphatic endothelial cells	FLT4, HYAL2, CLEC1A, SOX18, ROBO4, CXorf36, MYCT1, KANK3
Microvascular endothelial cells	ACVRL1, CETP, RANGAP1, SELE, TIE1, CLDN5, VWF, HYAL2, EIF2B2, TNFSF18, LYVE1, ARHGEF15, CLEC1A, ROBO4, CXorf36, MMRN2, KANK3

Abbreviations: Tcm: Central memory T cell; Tem: Effective memory T cell; Treg: Regulatory T cell; cDC: Conventional dendritic cell; iDC: Interdigitating dendritic cell; aDC: Active dendritic cell; MSC: Mesenchymal stem cell.