RESEARCH

Integration of single-cell and spatial transcriptomics reveals fibroblast subtypes in hepatocellular carcinoma: spatial distribution, differentiation trajectories, and therapeutic potential

Yue Liu[1,2,3], Guoping Dong[2,3], Jie Yu[2,3] and Ping Liang[1,2,3]*

Abstract

Background Cancer-associated fibroblasts (CAFs) are key components of the hepatocellular carcinoma (HCC) tumor microenvironment (TME). regulating tumor proliferation, metastasis, therapy resistance, immune evasion via diverse mechanisms. A deeper understanding of the l diversity of CAFs is essential for predicting patient prognosis and guiding treatment strategies.

Methods We examined the diversity of CAFs in HCC by integrating single-cell, bulk, and spatial transcriptome analyses.

Results Using a training cohort of 88 HCC single-cell RNA sequencing (scRNA-seq) samples and a validation cohort of 94 samples, encompassing over 1.2 million cells, we classified three fibroblast subpopulations in HCC: HLA-DRB1+CAF, MMP11+CAF, and VEGFA+CAF based on highly expressed genes of which, which are primarily located in normal tissue, tumor boundaries, and tumor interiors, respectively. Cell trajectory analysis revealed that VEGFA+CAFs are at the terminal stage of differentiation, which, notably, is tumor-specific. VEGFA+CAFs were significantly associated with patient survival, and the hypoxic microenvironment was found to be a major factor inducing VEGFA+CAFs. Through cellular communication with capillary endothelial cells (CapECs), VEGFA+CAFs promoted intra-tumoral angiogenesis, facilitating tumor progression and metastasis. Additionally, a machine learning model developed using high-expression genes from VEGFA+CAFs demonstrated high accuracy in predicting prognosis and sorafenib response in HCC patients.

Conclusions We characterized three fibroblast subpopulations in HCC and revealed their distinct spatial distributions within the tumor. VEGFA+CAFs, which was induced by hypoxic TME, were associated with poorer prognosis, as they promote tumor angiogenesis through cellular communication with CapECs. Our findings provide novel insights and pave the way for individualized therapy in HCC patients.

Keywords Tumor-associated fibroblasts, Hepatocellular carcinoma, VEGFA+CAF, Hypoxic microenvironment, Angiogen esis

*Correspondence:
Ping Liang
liangping301@126.com
Full list of author information is available at the end of the article

Liu *et al. Journal of Translational Medicine*   (2025) 23:198

Page 2 of 20

Background

Liver cancer is the third leading cause of cancer-related deaths and the sixth most common cancer globally, with HCC accounting for 80–85% of all liver cancer cases [1]. Despite surgical resection, the recurrence rate remains high, and patients with unresectable HCC have limited therapeutic options and poor prognosis [2]. Over 80% of HCC cases are characterized by extensive hepatic fibrosis due to the activation, proliferation, and accumulation of fibroblasts [3]. A prominent feature of the HCC tumor TME is the abundant presence of CAFs, which secrete various cytokines, chemokines, and growth factors that support cancer cell survival [4]. CAF-derived factors not only promote cancer cell survival but also modify the immune landscape by suppressing immune effector cells and recruiting immunosuppressive cells, allowing cancer cells to evade immune surveillance [5]. Fibroblasts play a critical role in TME remodeling, influencing tumor growth, metastasis, treatment resistance, and immune evasion, often likened to the "soil" in which cancer "seeds" thrive [6]. Therefore, understanding the diversityof fibroblasts in the HCC TME is essential.

Recent advances in scRNA-seq and spatial transcriptomics (ST) have overcome technical barriers to studying cellular heterogeneity in complex tissues like cancer [7, 8], providing valuable insights into fibroblast diversity in HCC. CAFs in primary liver cancer have been classified into eight subtypes, with DAB2+and SPP1+tumor-associated macrophages (TAMs) shown to promote immune barrier formation by enhancing FAP+CAF function through TGF-β, PDGF, and ADM signaling [9]. In HCC, six CAF subtypes have been identified based on gene expression and function, with POSTN+CAFs recruiting SPP1+macrophages and elevating SPP1 expression through the IL-6/STAT3 pathway [10]. Additionally, five CAF subpopulations have been classified, with CD36+CAFs shown to recruit CD33+myeloid-derived suppressor cells (MDSCs) in an MIF and CD74 dependent manner, enhancing MDSC-mediated immunosuppression and promoting tumor growth [11]. However, these studies did not account for the potential impact of mural cells that is composed of pericytes and smooth muscle cells in mesenchymal cells on fibroblast behavior, highlighting the need for standardized approaches to exclude mural cell effects for a comprehensive, high-resolution assessment of CAFs.

In this study, we characterized three distinct fibroblast subtypes in HCC, revealing their functional diversity and spatial distribution. Inference of cell differentiation trajectories highlighted the critical role of the TME in driving fibroblast differentiation. Additionally, we analyzed the complex biological functions resulting from intercellular communication patterns, paving the way for the development of therapeutic approaches targeting CAF in HCC.

Methods

Data collection

A total of 89 HCC-related samples from the Xue et al. study [12] were used as the discovery cohort, including 10 adjacent liver (AL) samples and 79 HCC samples. The raw FASTQ data (BioProject ID: PRJCA007744) is available in the Chinese National Center for Biological Information (CNCB) (https://www.cncb.ac.cn/). Additionally, scRNA-seq data for 94 HCC and AL samples were obtained as a validation cohort from the Gene Expression Omnibus (GEO) database (GSE149614, GSE151530, GSE189903, GSE202642, GSE156625) [11, 13–16]. For ST, samples from five HCC patients—including adjacent tumor, tumor, and tumor margin regions—were sourced from Wu et al. [17], totaling 15 slides. RNA-seq and microarray data were collected from The Cancer Genome Atlas (TCGA-LIHC, n=424), GEO (GSE14520, n=488) [18], the International Cancer Genome Consortium (ICGC-LIRI-JP, n=389), and the National Omics Data Encyclopedia (NODE, OEP000321, n=316).

Single-cell data processing

Raw FASTQ data from the discovery cohort were filtered and aligned using Starsolo (v2.7.9a) with the human reference genome GRCh38. All validation cohort scRNA-seq data were directly merged after downloading from GEO. Given the large cell number in the training set, preliminary filtering, dimension reduction, and clustering were performed using scanpy (v1.6) [19] in Python. Filtering criteria included: (1) 500–6000 detected genes, (2) UMI counts between 1000 and 30,000 per cell, (3) mitochondrial gene content below 10%. We selected 2000 highly variable genes (HVGs) using highly_variable_genes, calculated the top 50 principal components (PCs) via pca, regressed out mitochondrial gene effects, and scaled each gene to unit variance. Nearest neighbor graphs were constructed with the neighbors function, and clustering was performed using the louvain function (resolution=0.1). Dimension reduction and visualization were achieved using uniform manifold approximation and projection (UMAP). For each clustered subpopulation, dimension reduction and clustering were further refined using Seurat (v4.3.0) in R. For the validation cohort, Seurat [20] was used to analyze merged gene expression matrices with parameters consistent with those described above. To mitigate batch effects between samples and datasets, Harmony (v1.2.0) was applied to both discovery and validation cohorts.

Liu *et al. Journal of Translational Medicine*   (2025) 23:198

Page 3 of 20

Identification of fibroblasts

Mesenchymal cells were characterized based on established markers (LUM, COL1A1, and RGS5). The expression levels of fibroblast and mural cell markers were calculated using the addmodulescore function. Fibroblasts were classified as cells with a fibroblast score at least 0.1 higher than the mural cell score, while mural cells were defined as cells with a mural cell score exceeding the fibroblast score by 0.1. Cells with a score difference of less than 0.1 were categorized as undetermined.

Functional enrichment analysis

To understand the functional characteristics of the different cell subpopulations, we first identified highly expressed genes for each subpopulation using the Find-AllMarkers function (adjusted $P < 0.05$, $log2FC > 0.25$). Functional enrichment analysis of the top 30 genes for each subpopulation was then performed using the GO and KEGG databases through the R package clusterProfiler (v4.10.0) [21]. An FDR of less than 0.05 was considered significant. Additionally, the R package progeny (v1.24.0) [22] was used to evaluate the function of 14 pathways within the cells. For single-cell data, each cell was functionally scored using the addmodulescore function. For bulk RNA-seq data, GSVA (v1.52.3) [23] was used to assign functional scores to samples based on predefined gene sets.

Cell preference analysis

To assess the prevalence of distinct cell types across different tissues, we used the Ro/e, which represents the ratio of observed to expected cell counts. An Ro/e greater than 1 indicates enrichment of a specific cell cluster within a tissue, while an Ro/e less than 1 indicates depletion. Additionally, the odds ratio (OR) algorithm was used to further analyze cell preference. Here, an OR greater than 1.5 suggests a significant enrichment of cells in the specified sample category, while an OR less than 0.5 indicates substantial depletion.

Bulk deconvolution analysis

To assess the infiltration of different cell subpopulations in bulk RNA-seq and microarray data, we estimated cell abundance in each sample using CIBERSORTx [24]. This was conducted in two steps. First, due to CIBER-SORTx limitations, we randomly sampled each cell type, reserving 500 cells per type to create a single-cell count expression matrix and derive a signature matrix. Next, the constructed signature matrices and bulk mixture files were used for CIBERSORTx analyses.

Survival analysis

To link single-cell data with patient phenotypes from bulk sequencing, we utilized the R package Scissor (v2.0.0) [25]. Scissor + cells were associated with poorer patient survival, while Scissor − cells correlated with better prognosis.

Additionally, the bulk sequencing matrix was transformed into a cell-type abundance matrix using CIBERSORTx, which integrated single-cell and bulk sequencing data. The abundance of fibroblast subpopulations was used as a variable to assess the relationship between these subpopulations and patient survival. We performed univariate Cox regression analyses with the R package survival (v3.5.7), and Kaplan–Meier survival curves were generated for selected cell subgroups. Log-rank and Cox p-values of less than 0.05 indicated statistically significant associations. Optimal grouping thresholds and survival analyses for each cohort were calculated using the R package survminer (v0.4.9).

To evaluate whether specific gene expression levels in fibroblast subpopulations could serve as a prognosis predictor, we applied the GSVA method to generate a score based on bulk sequencing gene sets and assessed its correlation with patient survival.

CAFs differentiation trajectory analysis

To investigate potential cell differentiation trajectory, we used the R packages Monocle2 (v2.26.0) [26] and slingshot (v2.7.0) [27] for cell trajectory analysis. The fitGAM function in slingshot identified genes associated with pseudotime, enabling us to track gene expression changes over the inferred differentiation timeline.

Spatial transcriptomics (ST)analysis

We used the Seurat package for dimensionality reduction and clustering analyses on the downloaded ST data. Unlike single-cell analyses, ST data were not subjected to quality control. To assess the spatial distribution of fibroblast subpopulations, we scored highly expressed genes of CAFs using the addmodulescore function. The spatial distribution of these subpopulations was further validated by integrating single-cell transcriptomic and ST data with the R package celltrek (v0.0.94) [28], which generated sparse plots via a random forest model to simulate CAF distribution.

For spatial colocalization analysis of VEGFA + CAF and CapEc, we first scored the top 50 highly expressed genes of these cell types for each Visium sample using the AddModuleScore function, estimating their abundance at each spot. Spatial coordinates were extracted, scaled, and distances to the six nearest spots were calculated using the knn function from the dbscan package

Liu *et al. Journal of Translational Medicine*   (2025) 23:198

Page 4 of 20

(v1.1-12). Spots scoring above the 75th percentile for VEGFA + CAF were designated as starting points, while those above the 75th percentile for CapEc were defined as endpoints. Colocalized spots were identified when the distance between starting and endpoint spots was less than six. For each starting spot, the CapEc abundance of colocalized spots was normalized to obtain a standardized neighbor enrichment score.

Cell communication analysis

Cellular communication between fibroblasts and endothelial cells was initially analyzed using CellChat (v2.1.2) [29]. First, the overall communication strength and frequency between all endothelial cells and fibroblast subpopulations were calculated. Key cell types interacting with VEGFA + CAFs were identified, along with the main senders and receivers within angiogenesis-related pathways.

To infer specific signaling interactions between endothelial cells and fibroblasts and assess their impact on downstream target genes, we used the R package NicheNet (v2.1.0) [30]. Here, fibroblasts served as signal senders, and endothelial cells as receivers. The top 30 ligands, receptors, and target genes ranked by aupr_corrected—were visualized in a heatmap. Functional enrichment of target and receiver genes was conducted using the clusterProfiler package, with GO, KEGG, databases as references. Pathways with a corrected p-value < 0.05 were deemed significantly enriched.

Machine learning model construction

The mime R package (v0.12) [31] was used to determine whether gene expression levels in fibroblast subpopulations could predict HCC patient prognosis and inform personalized treatment strategies. First, patient prognosis was predicted using 101 combinations of 10 machine learning algorithms applied to the top 15 highly expressed genes in VEGFA + CAFs. One-way Cox regression (unicox.filter.for.candi = T) was used to filter genes independently correlated with patient prognosis. Seven additional machine learning models were then employed to predict HCC patients' responses to the angiogenesis inhibitor sorafenib.

Statistical analysis

Statistical analyses and visualizations were performed in R (v4.3.0). Differences between groups were compared using the Wilcoxon test. Survival analyses were conducted using the Kaplan–Meier method, with statistical significance determined by the log-rank test. Cox regression assessed correlations between variables and patient overall survival (OS) or recurrence-free survival (RFS),

using the median or optimal cut-off as thresholds. Statistical significance was set at P < 0.05. Correlation analyses, given the non-normal data distribution, were conducted using the Spearman method.

Results

Isolation of fibroblasts in HCC

To explore the diversity of CAFs in the HCC tumor microenvironment, we analyzed a cohort of 89 HCC-associated samples, as reported by Xue et al. After excluding one outlier sample, the cohort consisted of 78 HCC and 10 AL cases (Supplementary Fig. 1A). Following quality control (Supplementary Fig. 1B), 587,954 cells were obtained, with 54,125 cells from AL and 533,829 from HCC samples. After data normalization, a k-nearest-neighbor overlap of cells from individual patients was used to assess batch effects in the scRNA-seq data. Harmony correction significantly minimized batch effects between samples (Supplementary Fig. 1C). Through further clustering and visualization, we characterized seven cell types (Fig. 1A, B): B cells (marked by CD79A, MS4A1, n = 15,138), T/NK cells (marked by CD3D, GNLY, n = 293,829), endothelial cells (marked by PECAM1, VWF, n = 61,041), mesenchymal cells (marked by RGS5, COL1A1, n = 28,773), myeloid cells (marked by CD68, CD14, n = 106,695), epithelial cells (marked by EPCAM, ALB, APOA2, n = 69,574) and plasma cells (marked by IGHG1, MZB1, n = 12,904).

To isolate fibroblasts, we focused on mesenchymal cell subpopulations, which, upon reclustering, were divided into seven subpopulations (Supplementary Fig. 1D). Given the important role of mural cells within mesenchymal cells, we aimed to distinguish fibroblasts from mural cells. We initially analyzed expression levels of RGS5, MYH11, DCN, and LUM, as they are highly expressed in pericytes, smooth muscle cells, and fibroblasts, respectively. Results indicated that DCN and LUM were highly expressed in subclusters 2 and 5, RGS5 in subclusters 1, 3, 4, and 6, and MYH11 in subcluster 0 (Supplementary Fig. 1E). We also evaluated marker genes of fibroblasts from mural cells accompanied with those previously reported for fibroblast subpopulations in mesenchymal cell subpopulations. However, effective differentiation was not achieved (Supplementary Fig. 1F, G). To improve classification, we synthesized findings from prior studies to generate specific gene markers for fibroblasts and mural cells (Table S1), then examined the expression of these markers in our dataset (Fig. 1C). While some cells remained undetermined, these markers allowed us to distinguish between fibroblasts and mural cells within the mesenchymal cell population (Fig. 1D, E). Validation on

Liu *et al. Journal of Translational Medicine*   (2025) 23:198

Page 5 of 20
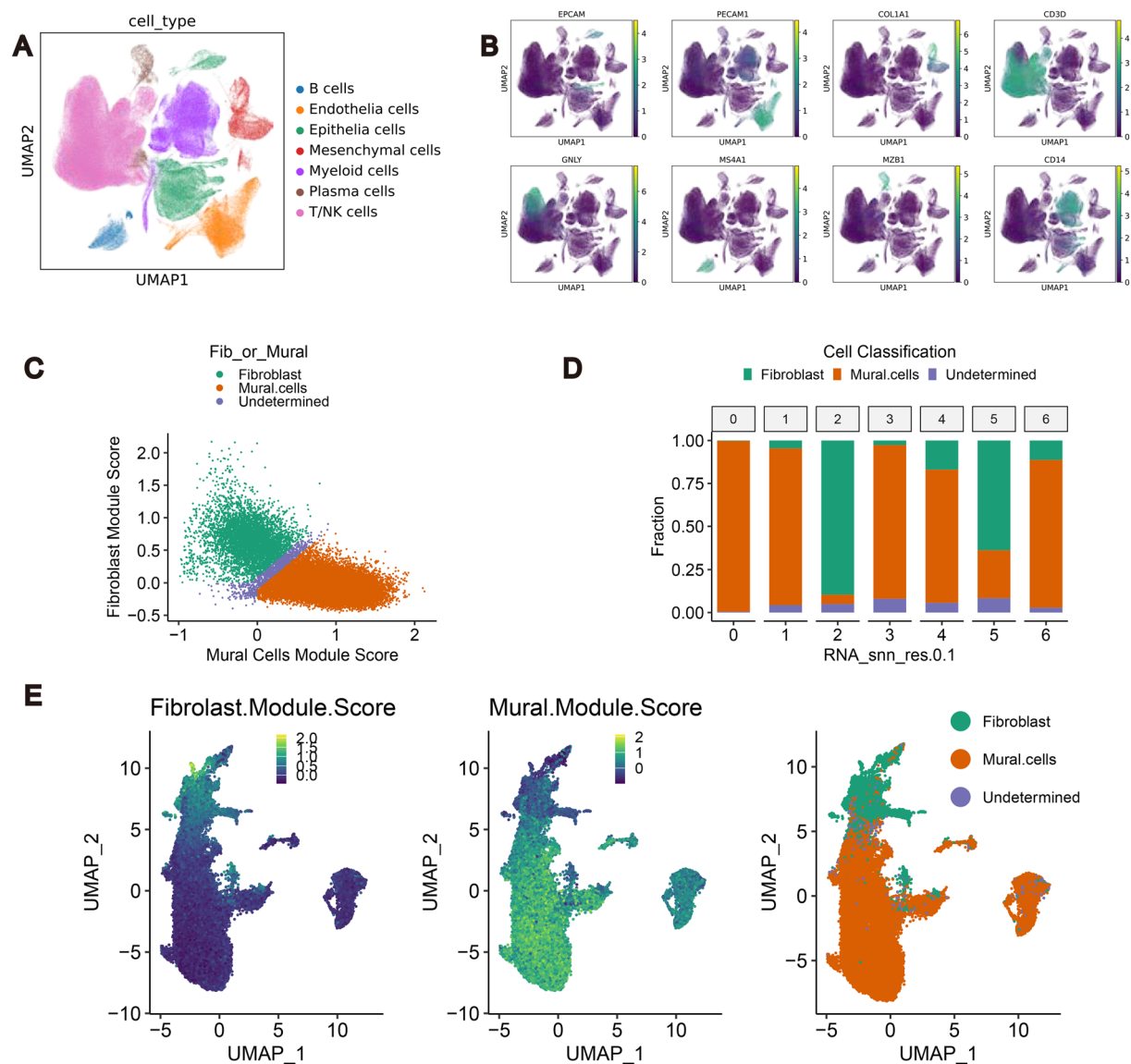
Fig. 1 Isolation of fibroblasts in HCC. **A** UMAP shows cell clustering of cell types in the discovery cohort. **B** UMAP plot showing highly expressed genes in cell types. **C** Scatterplot showing the classification of fibroblasts and mural cells based on canonical gene markers. **D** Bar graph showing the percentage of fibroblasts/mural cells as well as undetermined cells in each mesenchymal subpopulation based on the classification results. **E** UMAP plot showing the results of fibroblasts and mural cells demarcation in the discovery cohort using the consensus gene signature

additional cohorts consistently showed similar results (Supplementary Fig. 1H–N).

In summary, we identified a broadly applicable approach for distinguishing fibroblasts from mural cells in HCC-associated scRNA-seq data, a critical step for analyzing fibroblast heterogeneity in the HCC TME.

Subpopulation and functional enrichment analysis of fibroblasts

In the discovery cohort, fibroblast subpopulations were classified and named based on their highly expressed genes: HLA-DRB1 + CAF, MMP11 + CAF, and VEGFA + CAF (Fig. 2A, Table S2). These three fibroblast types could be distinguishable by the expression levels of HLA-DRB1, MMP11, and VEGFA (Fig. 2B, C). In the validation cohort, fibroblasts were classified into four subpopulations, with one group showing high
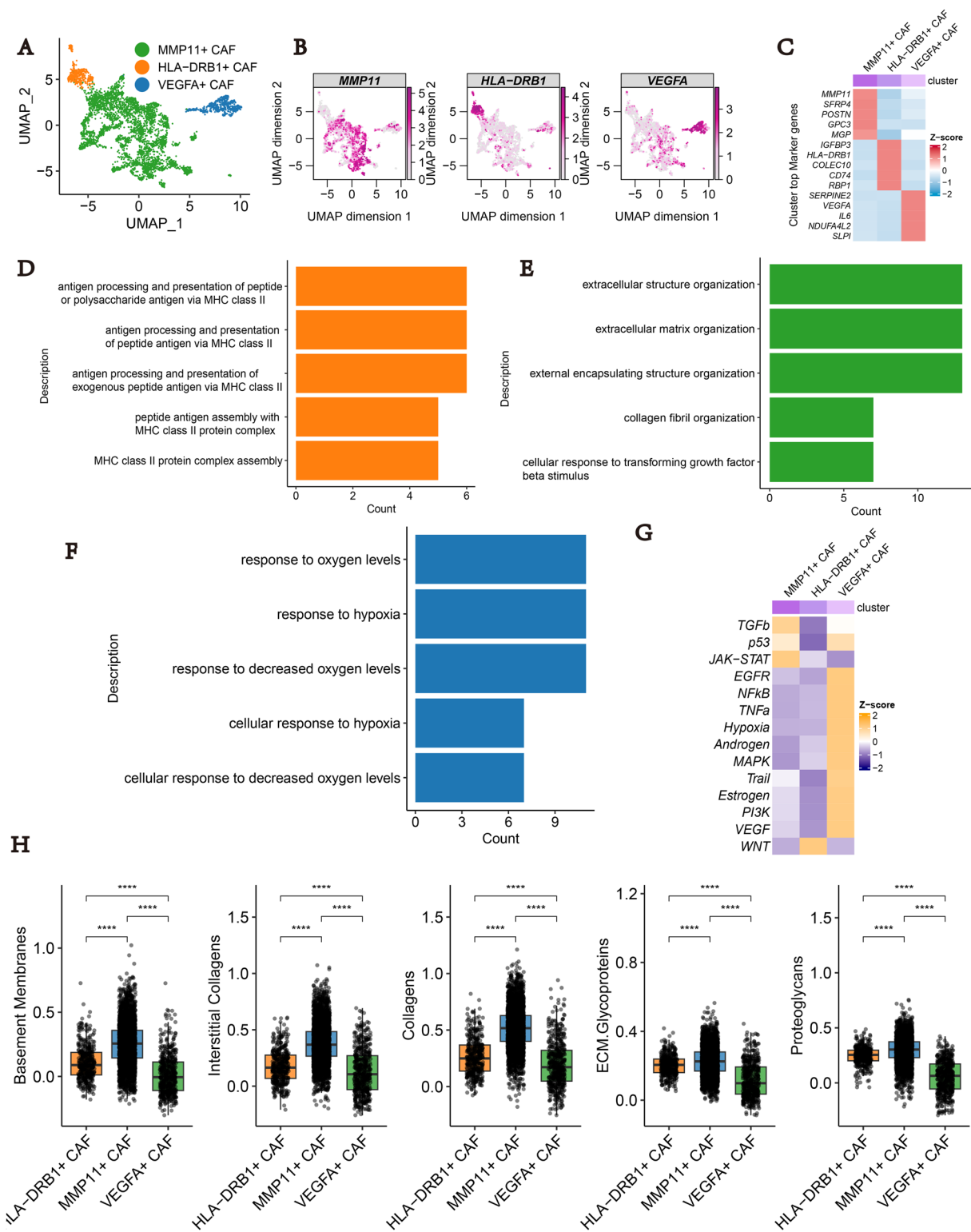
Fig. 2 Subpopulation and functional enrichment analysis of fibroblasts. **A** UMAP shows that fibroblasts can be clustered into three subpopulations. **B** UMAP plot showing the expression of three typical genes (MMP11/HLA-DRB1/VEGFA) in three fibroblast subpopulations. **C** Heatmap showing top 5 highly expressed genes in each subpopulation of fibroblasts. **D**–**F** Bar plot showing top 5 terms or pathways significantly enriched for HLA-DRB1 + CAF, MMP11 + CAF, VEGFA + CAF. **G** Heatmap showing the different subtypes of fibroblasts' progeny pathway scores. **H** Box line plots showing the scores of basement membrane, interstitial collagen, collagen, ECM glycoproteins, and proteoglycans in comparison for three subpopulations of fibroblasts. *P < 0.05; **P < 0.01; ***P < 0.001; ****P < 0.0001

Liu *et al. Journal of Translational Medicine*   (2025) 23:198

Page 7 of 20

HLA-DRB1 expression. However, MMP11 and VEGFA were expressed in the same subpopulations, likely due to the limited number of fibroblasts in the validation cohort (n=702) (Supplementary Fig. 2A–C).

To elucidate the functions of these fibroblast subtypes, we conducted functional enrichment analyses based on their highly expressed genes. The HLA-DRB1+CAFs were enriched in pathways related to antigen presentation (Fig. 2D), including antigen processing and presentation of peptide, polysaccharide antigen via MHC class II, MHC class II protein complex assembly, and so forth. These cells exhibited high expression of HLA-DRB1, COLEC10, and CD74, classifying them as classical antigen-presenting CAFs (apCAFs) [32]. MMP11+CAFs were significantly enriched in pathways related to extracellular matrix (ECM) organization and collagen fibril formation, such as extracellular structure organization, collagen fibril organization and cellular response to transforming growth factor beta stimulu, with high expression of ECM remodeling genes like MMP11 and POSTN, aligning them with matrix CAFs (mCAFs) [33] (Fig. 2E). VEGFA+CAFs were enriched in hypoxia-response pathways, including response to oxygen levels, response to hypoxia and cellular response to hypoxia. These cells showed high expression of SERPINE2, VEGFA, and IL-6, which are associated with angiogenesis, suggesting a role in promoting angiogenesis within hypoxic environments (Fig. 2F). Pathway scoring using Progeny further demonstrated that MMP11+CAFs scored highest on the TGF-β pathway, while VEGFA+CAFs showed higher scores on the hypoxia and VEGF pathways, consistent with the functional enrichment findings (Fig. 2G).

The generation and remodeling of ECM are fundamental functions of fibroblasts across tissues; however, these functions vary significantly among fibroblast subpopulations. We examined whether each subpopulation upregulated genes linked to specific ECM components, including basement membrane, interstitial collagen, ECM glycoproteins, and proteoglycans. Results showed that MMP11+CAFs upregulated numerous genes related to basement membrane, interstitial collagen, ECM glycoproteins, and proteoglycans. HLA-DRB1+CAFs also regulated these pathways. Interestingly, VEGFA+CAFs only

upregulated genes related to ECM glycoproteins and proteoglycans, with a relatively lower proportion, suggesting that VEGFA+CAFs may have lost most of their ECM-related functions. (Supplementary Fig. 2D). To further clarify these distinctions, we calculated module scores for each ECM category and compared expression across fibroblast subpopulations (Fig. 2H). MMP11+CAFs exhibited the highest ECM-generation-related functions, including basement membrane, interstitial collagen, collagen, ECM glycoproteins, and proteoglycans, followed by HLA-DRB1+CAFs, while VEGFA+CAFs demonstrated the lowest activity.

Taken together, these data reveal three fibroblast subpopulations and their functions in the HCC tumor microenvironment, which may differentially regulate the maintenance/remodeling of the extracellular matrix.
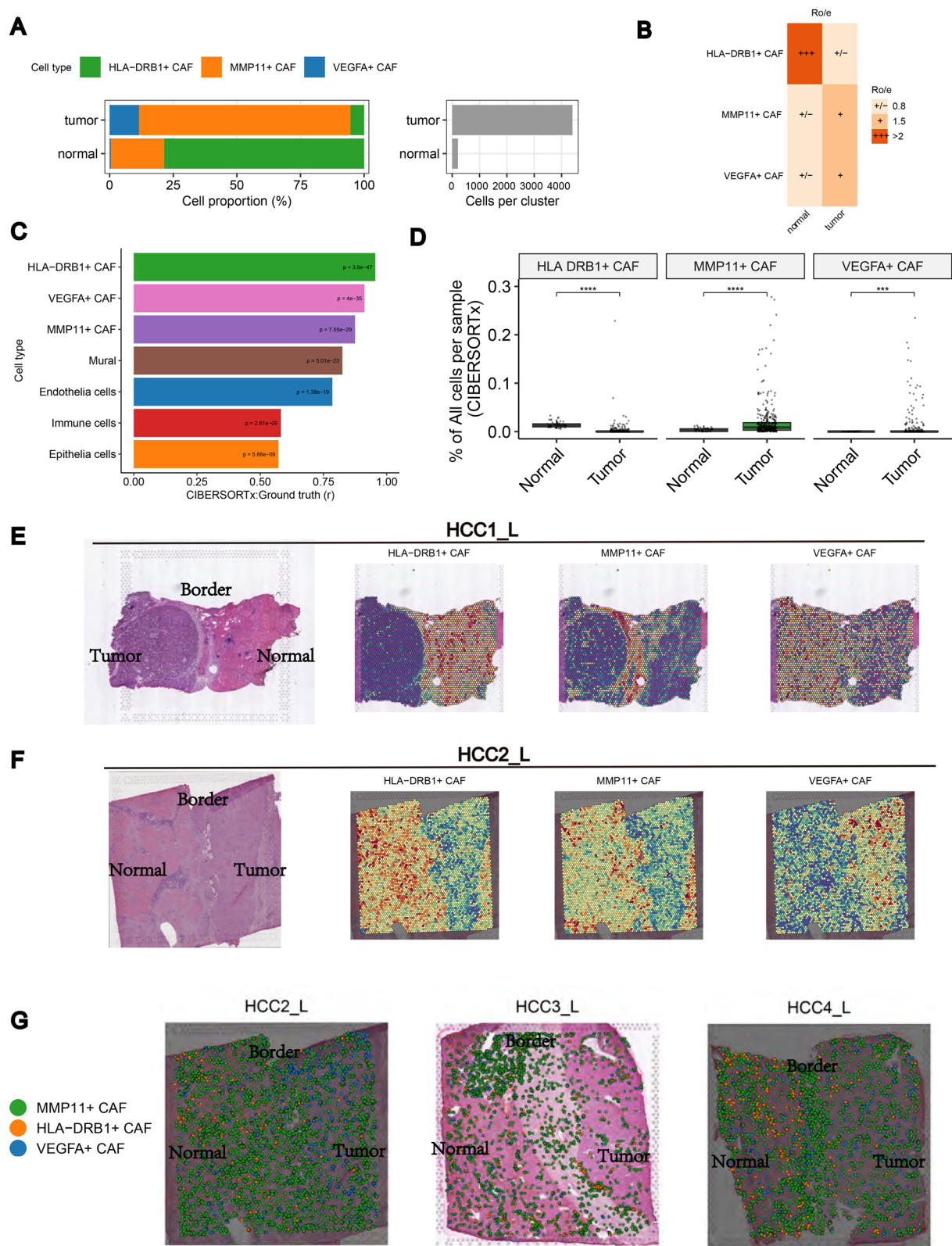
Spatial distribution of the three fibroblast subtypes

We analyzed the infiltration patterns of fibroblast subtypes across different tissue types. Results showed that HLA-DRB1+CAF and MMP11+CAF were predominant in normal tissues, while VEGFA+CAF and MMP11+CAF were more frequent in tumor tissues (Fig. 3A). Notably, HLA-DRB1+CAF was almost absent in tumor tissues, whereas VEGFA+CAF was concentrated within tumor areas, suggesting that the tumor microenvironment may influence fibroblast differentiation. Ro/e analysis supported these findings (Fig. 3B). The relative percentages of the three fibroblast types in normal and tumor tissues are provided in Supplementary Fig. 3A. These results indicate that HLA-DRB1+CAF is scarce in tumor tissues, while MMP11+CAF is less abundant in normal tissues. VEGFA+CAF did not show a statistically significant difference between normal and tumor tissues, likely due to the low count in normal tissues (n=1). Scoring of all cells using addmodulescore further confirmed that cells in normal tissues had high scores for HLA-DRB1+CAF, whereas tumor cells had higher scores for both MMP11+CAF and VEGFA+CAF (Supplementary Fig. 3B).

To validate these fibroblast phenotypes in a larger cohort, we used CIBERSORTx to create cell abundance matrices that included fibroblast subpopulations,

(See figure on next page.)

Fig. 3 Spatial distribution of the three fibroblast subtypes. **A** Bar graphs show the percentage of the three fibroblasts among various sample types. **B** RO/E showing the abundance of the three fibroblasts to be enriched in AL and tumour tissues **C** Accuracy validation using a pseudobulk dataset generated from a single-cell transcriptome. Bar graphs show the correlation between CIBERSORTx estimates and true abundance for each cell type (Pearson r). **D** Box plot showing the comparison of the abundance of the three fibroblasts in normal and tumour tissues based on the deconvolution results of CIBERSORTx. **E** Spatial distribution of the three fibroblasts in the HCC1_L slide. **F** Spatial distribution of the three fibroblasts in the HCC2_L slide. **G** CellTrek deconvolution based on the HCC sections in the distribution of the three fibroblast subtypes in,HCC3_L,HCC4_L. *P < 0.05; **P < 0.01; ***P < 0.001; ****P < 0.0001

Fig. 3 (See legend on previous page.)

Liu *et al. Journal of Translational Medicine*   (2025) 23:198

Page 9 of 20

endothelial cells, mural cells, epithelial cells, and immune cells. We evaluated accuracy using pseudobulk datasets, generated by pooling single-cell transcriptomes to serve as reference samples. Results showed a strong correlation between CIBERSORTx estimations and the pseudobulk dataset for HLA-DRB1+CAF ($R^2=0.91$, p < 0.001), MMP11+CAF ($R^2=0.77$, p < 0.001), and VEGFA+CAF ($R^2=0.83$, p < 0.001) (Fig. 3C, Supplementary Fig. 3C–E). Analysis of the TCGA RNA-seq LIHC dataset using CIBERSORTx confirmed that HLA-DRB1+CAF is frequent in normal tissues, whereas MMP11+CAF and VEGFA+CAF are more common in tumor tissues (Fig. 3D).

We investigated the spatial distribution of the three fibroblast subtypes within tumors. On the HCC1-L slide, addmodulescore scoring across all spots revealed that MMP11+CAF was mostly located at the tumor border, HLA-DRB1+CAF was mainly found in normal tissues, and VEGFA+CAF was concentrated within the tumor interior. (Fig. 3E). Similar distribution patterns were observed in other tumor border slides (Fig. 3F). The distributions of fibroblast types across tumor and normal slides corroborated these findings (Supplementary Fig. 3F, G). Using Celltrek to integrate single-cell and spatial transcriptomic data, we simulated the spatial distribution of these fibroblast subtypes, showing MMP11+CAFs predominantly at the tumor border, HLA-DRB1+CAFs in normal tissues, and VEGFA+CAFs in tumor interiors (Fig. 3G).

In conclusion, HLA-DRB1+CAF is primarily distributed in normal tissues, MMP11+CAF is concentrated at the tumor periphery, and VEGFA+CAF is located within the tumor interior. These distribution patterns may reflect the distinct biological roles of each fibroblast subtype.

Cell-state transition trajectory inference of CAFs

To explore the dynamic processes of CAF subtypes at the single-cell level, we inferred cellular trajectories of CAF subtypes using Slingshot (Fig. 4A–C). Results revealed that HLA-DRB1+CAFs represent the progenitor state, which progressed into MMP11+CAFs, with VEGFA+CAFs occupying the terminal differentiation stage. In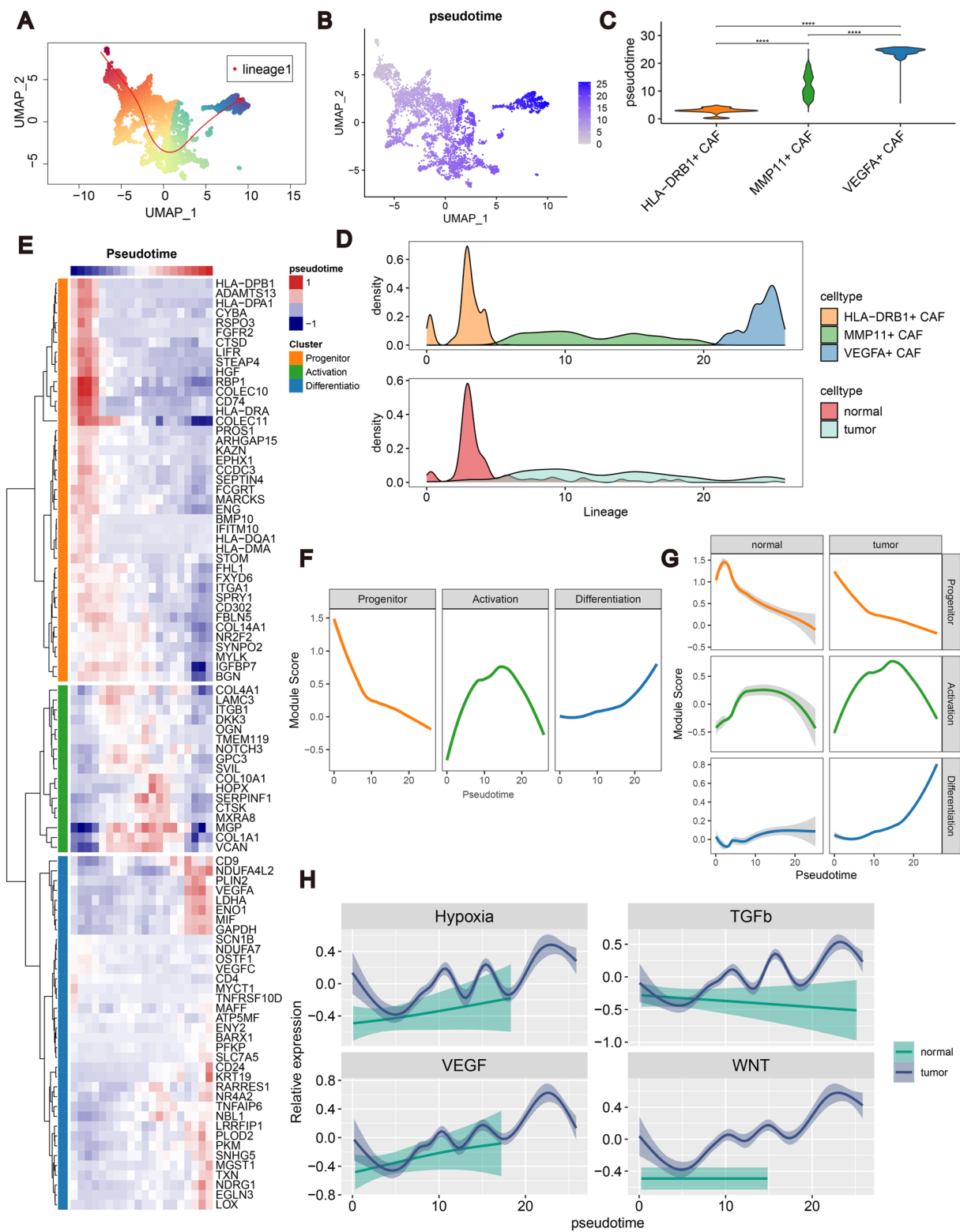terestingly, fibroblasts from normal tissues predominantly appeared at the early, pre-differentiation stage, while fibroblasts from tumor tissues were observed at later stages, suggesting a transition from normal fibroblasts to tumor-associated fibroblasts during tumorigenesis and progression (Fig. 4D). We repeated this analysis using Monocle2, which yielded similar results, thereby enhancing the robustness of our findings (Supplementary Fig. 4A, B).

We also examined gene expression patterns associated with CAF state transitions. Differentially expressed genes were clustered into three modules based on pseudotime-associated expression patterns, labeled as progenitor, activation, and differentiation modules (Fig. 4E). To investigate the functions of these genes at different differentiation stages, we performed KEGG pathway enrichment analysis (Supplementary Fig. 4C–E). The progenitor module was enriched in pathways related to antigen processing and presentation; the activation module was enriched in pathways involving focal adhesion and ECM-receptor interactions, both linked to ECM; and the differentiation module was associated with the HIF-1 signaling pathway, suggesting a close link to hypoxia. These functional analyses suggest that the three pseudotime-related modules align with the identified fibroblast subpopulations. In terms of expression across the tumor timeline, we observed that the progenitor, activation, and differentiation modules were highly expressed in the early, mid, and late stages, respectively. Notably, analyzing tumor and normal tissues separately showed that the progenitor and activation modules were present in both normal and tumor tissues, while the differentiation module was specific to tumor tissues, suggesting a tumor-specific feature (Fig. 4F, G, Supplementary Fig. 4F).

Furthermore, we analyzed pathway expression along pseudotime and found that Hypoxia, TGF-β, VEGF, and WNT pathway activity increased over time in tumor tissues (Fig. 4H), whereas these levels remained stable or decreased in normal tissues. This may correlate with the high presence of VEGFA+CAF in tumor tissues, as these pathways are enriched in this subtype. Conversely, some MMP11+CAF-enriched pathways did

(See figure on next page.)

Fig. 4 Cell-state transition trajectory inference of CAFs. **A** UMAP plot demonstrating possible differentiation trajectories between the three fibroblasts. **B** UMAP plot demonstrating the distribution of fibroblasts in pseudotime. **C** Violin plot demonstrating the comparison of the three fibroblasts in pseudotime. **D** Distribution of the three fibroblasts as well as fibroblasts in normal and tumour tissues in pseudotime. **E** Heatmap demonstrating genes in fibroblasts during their differentiation over time and clustering analysis dividing them into three modules: progenitor, activation and differentiation. **F** Trends in gene expression of the three modules over time. **G** Trends in gene expression of the three modules over time in normal and tumour tissues. **H** Hypoxia, TGFb, VEGF, WNT pathway scores over time in normal and tumour tissues

Fig. 4 (See legend on previous page.)

Liu *et al. Journal of Translational Medicine*   (2025) 23:198

Page 11 of 20

not exhibit significant differences between normal and tumor tissues (Supplementary Fig. 4G).

Overall, our cell trajectory analysis of CAF subtypes classified pseudotime-related genes into three modules: progenitor, activation, and differentiation. Importantly, we found that the differentiation module appears tumor-specific, corresponding to the VEGFA + CAF subtype.

VEGFA + CAF was associated with poor prognosis

We used Scissor, developed by Sun et al., to investigate the relationship between fibroblast subpopulations and OS in HCC patients. Based on TCGA bulk RNA-seq LIHC and survival data, we identified Scissor-positive (Scissor +) and Scissor-negative (Scissor −) fibroblast cells. Scissor + cells were associated with shorter OS, whereas Scissor − cells correlated with longer OS (Fig. 5A).

Next, we assessed the proportion of Scissor + cells within each fibroblast subpopulation. Results indicated that Scissor + cells were most prevalent in VEGFA + CAFs, while Scissor − cells were more abundant in MMP11 + CAFs (Fig. 5B, Supplementary Fig. 5A). Overall, Scissor + cells showed an association with VEGFA + CAF and, to a lesser extent, MMP11 + CAF, whereas HLA-DRB1 + CAF was more commonly associated with Scissor − cells (Supplementary Fig. 5B). Validation in the discovery cohort confirmed that patients with a high percentage of Scissor + cells had lower OS (Fig. 5C). Further analysis revealed that Scissor + cells highly expressed VEGFA (Fig. 5D). Additionally, a dot plot of the top ten genes showed that genes highly expressed in Scissor + cells were predominantly found in VEGFA + CAF, while genes highly expressed in Scissor − cells were primarily in HLA-DRB1 + CAF (Table S3). This suggests that VEGFA + CAF is the dominant cell type in Scissor + cells and a key factor impacting OS in HCC patients (Supplementary Fig. 5C).

To further investigate the correlation between fibroblast subpopulations and patient prognosis in a larger cohort, we used CIBERSORTx to quantify the abundance of the three fibroblast subpopulations in the TCGA LIHC dataset. Results from one-way Cox regression analyses indicated that VEGFA + CAFs were significantly associated with poorer OS (HR = 2.17, P = 0.001), whereas HLA-DRB1 + CAFs correlated with a favorable prognosis (HR = 0.62, P = 0.012). No statistically significant association was found between OS and MMP11 + CAFs (HR = 1.38, P = 0.224) (Fig. 5E). Kaplan–Meier survival curve analysis confirmed that patients with a higher proportion of VEGFA + CAFs had a significantly worse prognosis (P < 0.0001) (Fig. 5F). These findings were further validated in two external datasets, showing a significant

correlation between the ratio of VEGFA + CAFs and both OS and RFS (Fig. 5G).

In conclusion, our findings indicate that VEGFA + CAF is a dominant cell type among Scissor + cells and plays a pivotal role in influencing OS in HCC patients. These results were further validated in two external datasets, reinforcing the significance of VEGFA + CAF in HCC prognosis.

Cell communication between VEGFA + CAF and CapECs contribute to angiogenesis.in HCC

The studies above demonstrated a strong correlation between the presence of VEGFA + CAF and poorer prognosis, alongside high VEGFA expression, which plays a key role in angiogenesis. Based on these observations, we propose that VEGFA + CAF may contribute to poor prognosis in HCC patients by promoting angiogenesis. To explore this, we re-dimensioned, clustered, and annotated endothelial cells, identifying six cell types: arterial endothelial cells (ArtECs), venous endothelial cells (VenECs), CapECs, liver1 sinusoidal endothelial cells (LSECs), lymphatic endothelial cells (LECs), and proliferating endothelial cells (pECs) (Fig. 6A, B, Table S4). We then analyzed cell subtype enrichment preferences and infiltration ratios across different sample types (Fig. 6C). In tumor tissues, CapECs and pECs were enriched along with VEGFA + CAF and MMP11 + CAF, while HLA-DRB1 + CAF, LSEC, and LECs showed greater enrichment in normal samples.

To analyze the spatial colocalization of VEGFA + CAF and CapEc, we first scored the top 50 highly expressed genes of these cell types for each Visium sample using the addmodulescore function, estimating their abundance at each spot. Spots with VEGFA + CAF scores above the 75th percentile were designated as starting points, while those with CapEc scores above the 75th percentile were defined as endpoints. Colocalized spots were identified when the distance between starting and endpoint spots was less than six. For each starting spot, the CapEc abundance of colocalized spots was normalized to obtain a standardized neighbor enrichment score. The results showed significant colocalization of VEGFA + CAF and CapEc in HCC tumor samples, suggesting that their spatial proximity may provide a physiological basis for potential cell–cell communication (Fig. 6D, E).

Using CellChat, we examined cellular communication patterns between fibroblasts and endothelial cells, revealing complex communication between these cell types (Supplementary Fig. 6A). We examined the communication pattern of VEGFA + CAFs and CapECs across all pathways, with a specific focus on the VEGF pathway. Here, VEGFA + CAFs demonstrated the strongest outgoing interaction, while CapECs showed the strongest
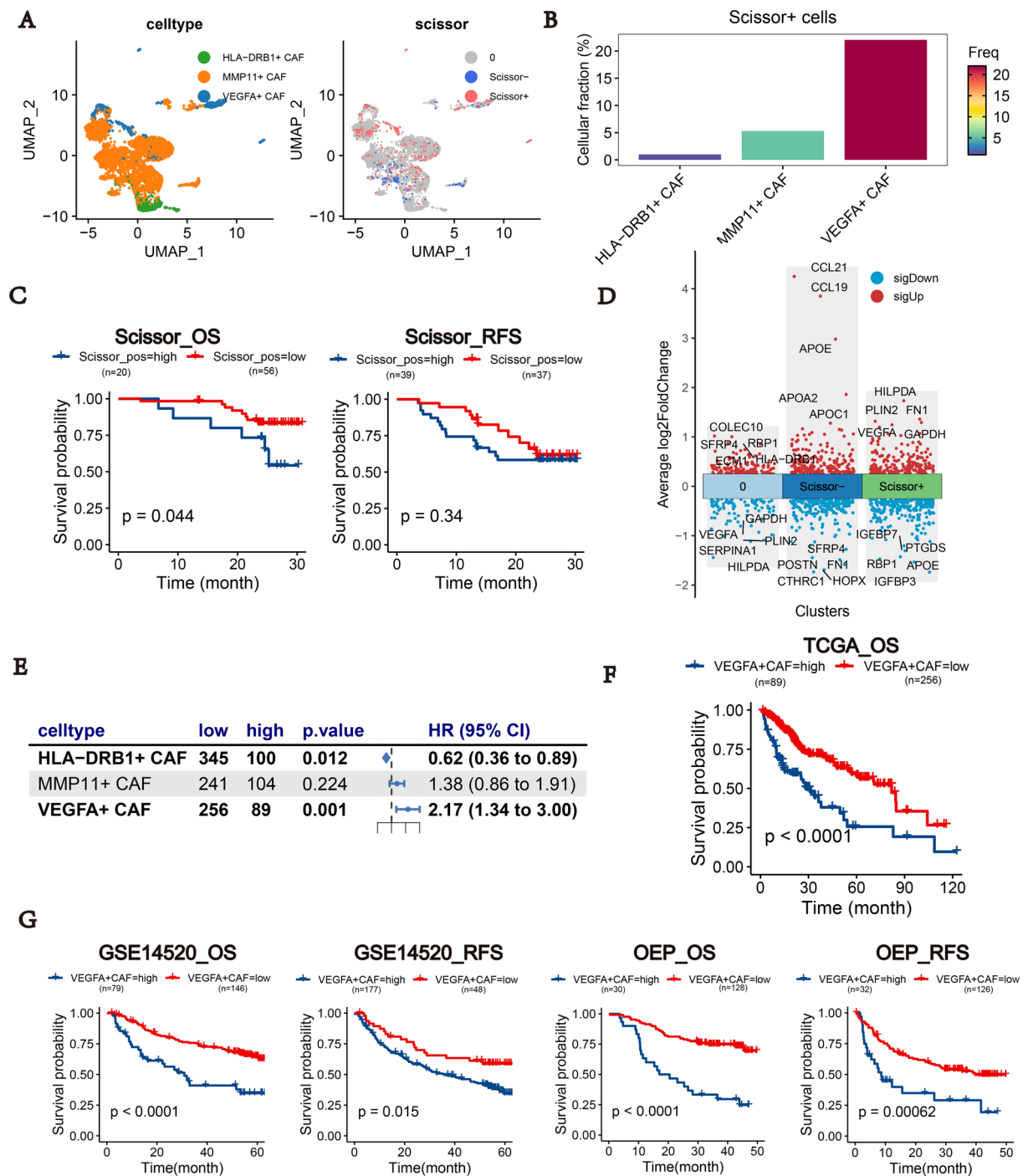
Fig. 5 VEGFA + CAF was associated with poor prognosis. **A** UMAP plot demonstrating the distribution of Scisso + as well as Scissor − cells among the three fibroblasts. **B** Bar graph demonstrating the percentage of Scisso + cells in the fibroblast subpopulation. **C** Kaplan–Meier survival curve demonstrating the differences in OS as well as RFS in patients with different Scisso + as well as Scissor − cell abundance. **D** Demonstration of survival-independent cells, Scisso + as well as Scissor − cell subpopulations of highly and lowly expressed genes. **E** Forest plot demonstrating the relationship between the three fibroblast abundances and patient OS (in TCGA-LIHC). **F** Kaplan–Meier survival curve demonstrating the differences in OS in patients with different VEGFA + CAF abundances. **G** Kaplan–Meier survival curve demonstrating the differences in OS in patients with different VEGFA + CAF abundances in the two external validation sets, VEGFA + CAF was not only associated with poorer OS but also poorer RFS

Liu *et al. Journal of Translational Medicine*   (2025) 23:198

Page 13 of 20

incoming interaction, suggesting a high probability of interaction via the VEGF pathway (Fig. 6F). Heatmaps further supported this hypothesis (Fig. 6G, Supplementary Fig. 6B–D), indicating that this cellular interaction is likely mediated by the VEGFA-VEGFR1 ligand-receptor pair (Supplementary Fig. 6E).

To assess the downstream effects of cellular interactions between fibroblasts and CapECs, we employed NicheNet (Supplementary Fig. 6F). VEGFA, the top-ranked ligand, was most highly expressed in VEGFA + CAFs and had the greatest impact on fibroblast-CapEC communication. Enrichment analysis of receptor and target genes revealed that receptor genes were enriched in pathways related to endothelium development and endothelial cell differentiation, while target genes were linked to epithelial cell proliferation and vasculogenesis. This suggests that communication between VEGFA + CAF and CapECs may enhance the angiogenic capacity of CapECs, thereby facilitating tumor progression and metastasis (Supplementary Fig. 6G, H).

In conclusion, our findings suggest that VEGFA + CAFs may interact with CapECs through the VEGF pathway, promoting angiogenesis and potentially driving tumor progression.

Predicting prognosis and making therapeutic decision choices for HCC patients by machine learning using highly expressed genes in VEGFA + CAF

Based on the findings from our analyses, we propose that VEGFA + CAF could serve as a novel biomarker for prognosis and treatment selection. We selected the 15 most highly expressed genes of VEGFA + CAF as marker genes and scored patients using GSVA. Results demonstrated that in the TCGA ($P < 0.0001$), OEP ($P < 0.0001$), ICGC ($P = 0.035$), and GSE14520 ($P < 0.0001$) cohorts, patients with high VEGFA + CAF scores had significantly lower OS than those with low scores. Similarly, in the OEP ($P = 0.0034$) and GSE14520 ($P < 0.0001$) cohorts, RFS was also significantly lower in patients with high scores (Fig. 7A). Additionally, in the TCGA and GSE14520 cohorts, patients with tumor stages III or IV had significantly higher VEGFA + CAF scores compared to those with stages I or II, suggesting that high VEGFA + CAF expression may serve as a biomarker for

tumor progression and poor prognosis in HCC patients (Supplementary Fig. 7A).

The top 15 genes expressed by VEGFA + CAF were used to predict HCC prognosis through 101 combinations of 10 machine learning algorithms in mime. The TCGA dataset served as the training set, while the ICGC, GSE14520, and OEP datasets were used as test sets. Results indicated that the StepCox [forward] + Enet [α = 0.3] combination had the highest predictive accuracy, with a C-index of 0.693 across all cohorts and 0.69 in the test cohort. Notably, the C-index for the test cohort ICGC (0.71) exceeded that of the discovery cohort (0.7), indicating strong model performance (Fig. 7B, C). When patients were classified by median risk score from the model, high-risk patients showed lower OS rates across all four cohorts (Fig. 7D). The model demonstrated robust efficacy in predicting one-year and three-year survival, with AUCs of 0.787 and 0.723, respectively (Supplementary Fig. 7B, C). Cox regression analyses further confirmed that the risk score was an independent predictor of OS in all four cohorts (Supplementary Fig. 7D).

Given the role of VEGFA + CAF in promoting angiogenesis, we explored whether a machine learning model based on VEGFA + CAF's top genes could assist in personalizing sorafenib treatment for HCC patients. Sorafenib is an anticancer drug effective for unresectable HCC, functions by inhibiting tumor growth and angiogenesis [34]. The GSE109211 dataset, a dataset that includes RNA seq data from HCC patients, along with information on their response to sorafenib treatment was split into a discovery cohort (comprising 70% of the samples) and a validation cohort (comprising 30% of the samples). Seven machine learning models were created using the top 15 genes of VEGFA + CAF, and their predictive efficacy was evaluated in the validation cohort. Results showed that the LogiBoost model had the highest predictive power for identifying HCC patients likely to respond to sorafenib, with an AUC of 0.94 (Fig. 7E, Supplementary Fig. 7E). These findings highlight the LogiBoost model's potential utility in predicting responses to sorafenib therapy in clinical settings.

In conclusion, machine learning models based on VEGFA + CAF's top genes offer a powerful tool for

(See figure on next page.)

Fig. 6 Cell communication between VEGFA + CAF and CapECs contribute to angiogenesis.in HCC. **A** UMAP plot demonstrating subpopulations of endothelial cells. **B** Heatmaps illustrating the top five most highly expressed genes in the various subpopulations of endothelial cells. **C** OR plots demonstrating the tissue distribution preferences of different fibroblasts as well as endothelial cell subpopulations. **D** VEGFA + CAF and CapEC scores, colocalized spots, and abundance distribution in the HCCT_5_3 sample. **E** VEGFA + CAF and CapEC scores, colocalized spots, and abundance distribution in the HCCT_5_2 sample. **F** Scatter plots demonstrating the activity of all pathways, as well as the VEGF pathway alone. **G** Heatmap demonstrating the strength of outgoing and incoming interactions of VEGF pathways between endothelial cells and fibroblast subpopulations
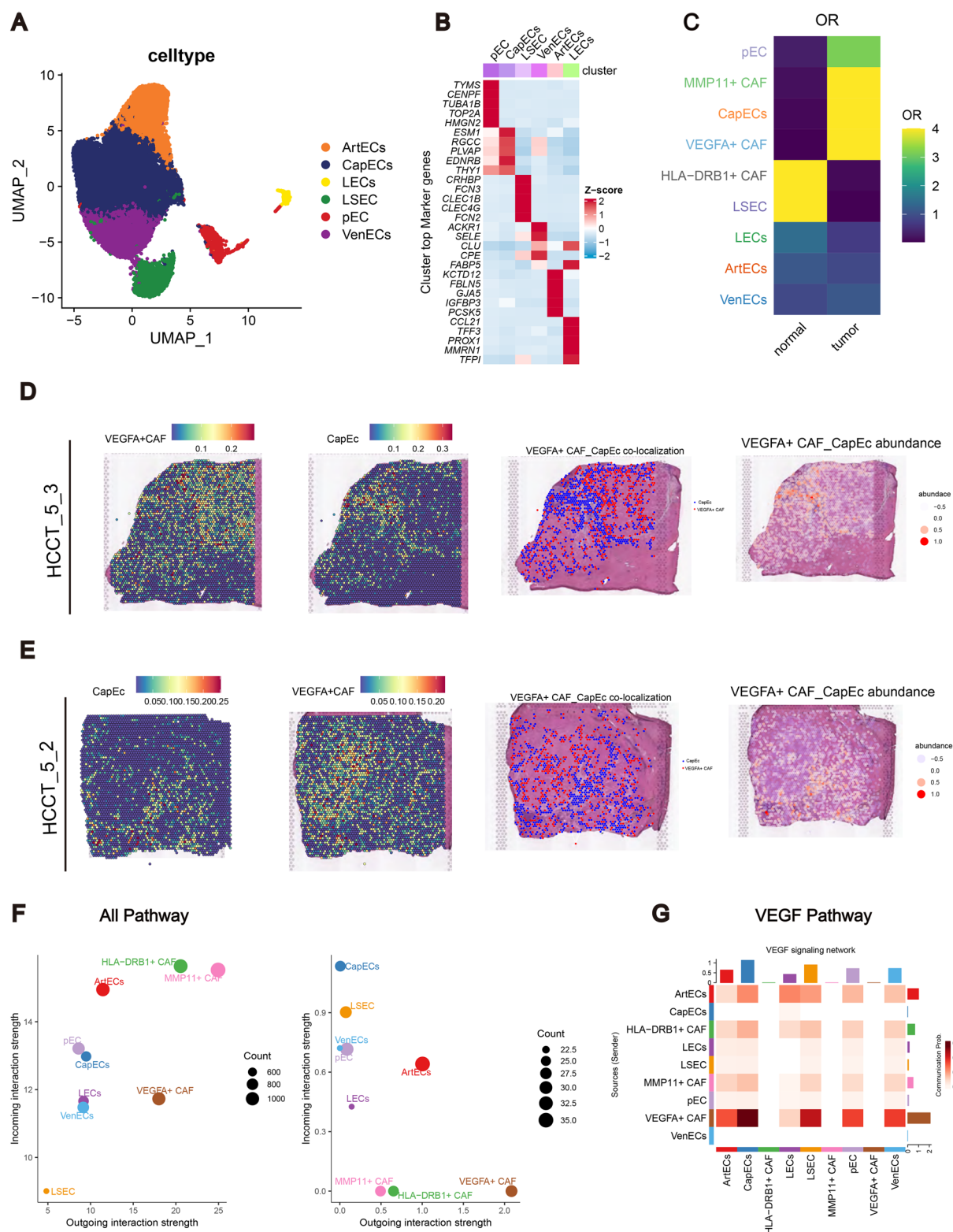
Fig. 6 (See legend on previous page.)

predicting HCC patient prognosis and supporting personalized sorafenib therapy.

Discussion

The growth of solid tumors is largely dependent on a remodeled stroma composed of CAFs and ECM, which play a pivotal role in the formation of an immunosuppressive TME, tumorigenesis, progression, metastasis and treatment resistance [35]. Previous studies of CAFs in HCC have typically been limited by small sample sizes and the potential interference of mural cells in mesenchymal cells [9–11]. In this study, we employed multiple cohorts and comprehensive analyses of scRNA-seq, ST, and bulk RNA-seq to systematically establish a process for isolating fibroblasts from mesenchymal cells, characterized three fibroblast subpopulations, analyze their locations, biological functions, and clarify their value in predicting the prognosis of patients with HCC and in developing personalized treatment plans. It is noteworthy that the communication between VEGFA + CAF and CapECs appears to be of crucial importance during the progression of HCC and the process of metastasis.
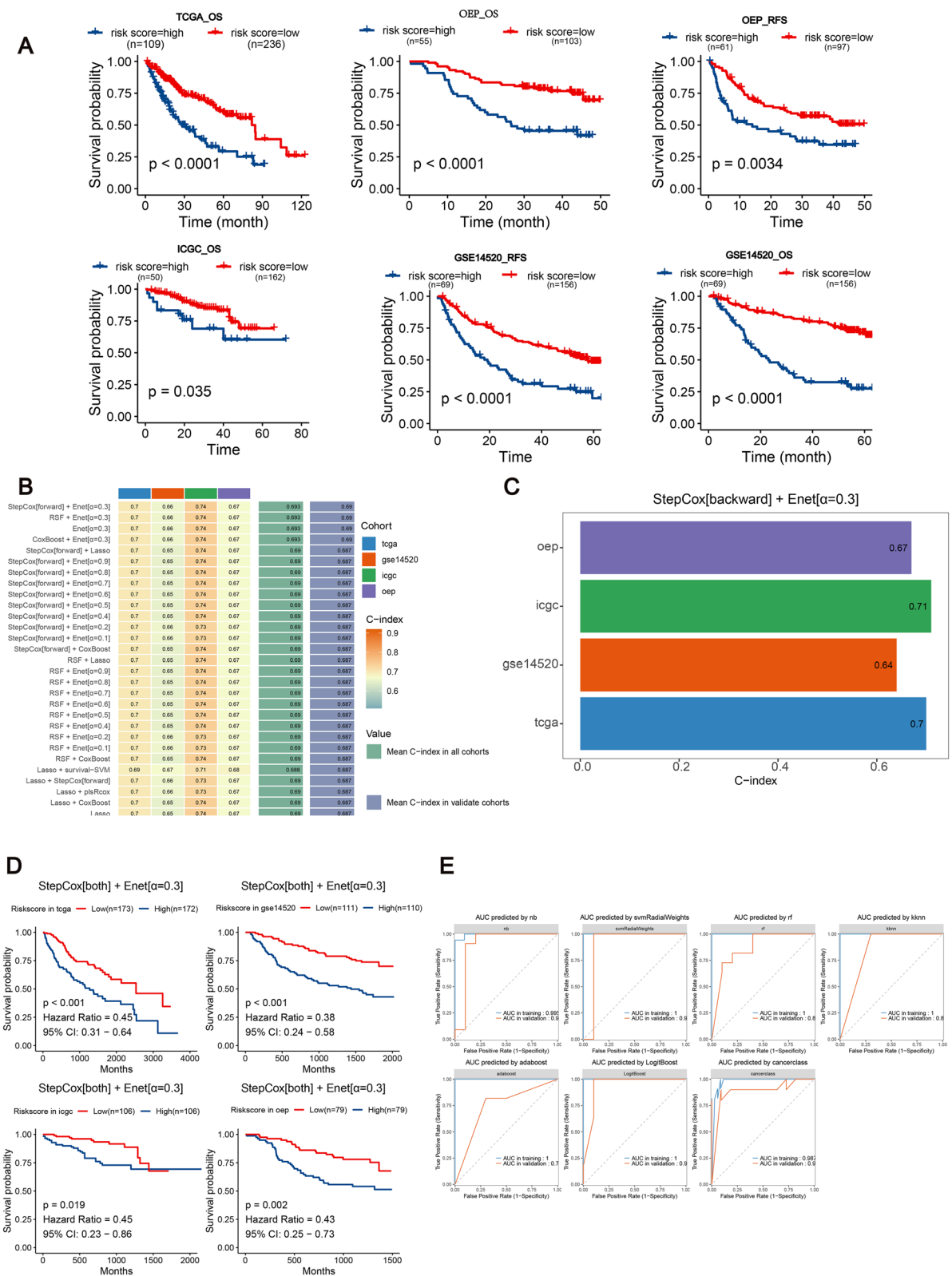
Fibroblasts in the HCC tumor microenvironment are highly heterogeneous and the heterogeneity confers different roles on them. In this study, we characterized a subpopulation of fibroblasts with high expression of MHC class II related genes, which is similar to the previously described apCAF [32, 36]. Although it is mainly enriched for the functions of Antigen processing and presentation, the perception of its role in cancer is often contrary. Ela Elyada et al. have shown that in pancreatic ductal adenocarcinoma, as the absence of co-stimulatory molecules on antigen-presenting cells leads to T-cell unresponsiveness and regulatory T-cell (Treg) induction, the activation of CD4 + T-cells by apCAF and the promotion of their immune-regulatory function directly leads to tumor immunosuppression and a poorer prognosis[32]. In contrast, Dimitra Kerdidani et al. reported that apCAFs in lung cancer have the unique ability to activate effector T cells and have tumor suppressor functions [37]. In the present study, HLA-DRB1 + CAFs were predominantly enriched in AL and were associated with a better prognosis.

MMP11 + CAFs were significantly enriched in tumors and were involved in the maintenance and remodeling of the ECM, similar to mCAFs described in previous studies [10, 11, 33]. MMP is part of a large family of zinc-dependent protein hydrolase metalloenzymes and is well known for its role in ECM degradation [38]. MMP11 is a zinc-dependent protein hydrolase and a zinc-dependent protein hydrolase that remodels the ECM by degrading ECM proteins, plays an important role in the invasion and metastasis of solid malignant tumors, enabling tumor cells to modify ECM components and release cytokines that promote protease-dependent tumor progression [39]. Interestingly, we found that MMP11 + CAF were mainly enriched around the tumor, forming a barrier, which is consistent with the presence of peritumor on pathological sections, that are present around the tumor in 10–76% of HCC patients [40]. Previous studies have shown that the presence of these peritumor may play the role of a physical barrier, preventing immune cells from migrating into the core of the tumor, leading to immune rejection and resulting in patients' non-response to immunotherapy and poor prognosis [9, 10, 41].

VEGFA + CAF was significantly enriched, within tumors, in contrast to MMP11 + CAF. Enrichment analysis demonstrated that terms it mainly enriched in is response to hypoxia, which may be related to the hypoxic immune microenvironment within the tumor. Ernestina Marianna De Francesco, et al. demonstrated that, in the hypoxic microenvironment, HIF-1α/GPER signaling was elevated and mediated the expression of VEGF in CAF, thereby promoting angiogenesis within the tumour [42], which is consistent with our findings. It is noteworthy that evidence indicates that the synthesis and remodeling of the ECM by CAFs generates solid stress, which leads to vascular and lymphatic compression, reduced perfusion rates and increased hypoxia in the TME [43, 44]. This suggests that the hypoxic TME plays an important role in the induction and maintenance of VEGFA + CAFs. The pseudotime analysis indicated that VEGFA + CAF is at the terminal stage of differentiation and that the differentiation module corresponding to VEGFA + CAF is tumor-specific. This further suggests that the differentiation of VEGFA + CAF cannot be separated from the

(See figure on next page.)

Fig. 7 Predicting prognosis and making therapeutic decision choices for HCC patients by machine learning using highly expressed genes in VEGFA + CAF. **A** Kaplan–Meier survival curve between VEGFA + CAF scores and patients' OS and RFS in four large cohort bulk seq datasets. **B** Using VEGFA + CAF highly expressed genes, 101 combinations of 10 machine learning predictions were used to predict the OS of patients with HCC, and the top 20 patients with the highest average C-index in the four datasets were selected for presentation. The top20 with the highest average C-index in the four datasets are shown. **C** C-index of the model with the best predictive performance in the four datasets. **D** Risk scores generated by machine learning are associated with poorer prognosis in HCC patients. **E** Demonstration of the prediction of response or non-response to sorafenib in patients with HCC using 7 machine learning using VEGFA + CAF highly expressed genes (in auc form)

Fig. 7 (See legend on previous page.)

Liu *et al. Journal of Translational Medicine*   (2025) 23:198

Page 17 of 20

TME. In addition to VEGFA, VEGFA + CAF also exhibits high expression of IL-6. It has been demonstrated that IL-6 produced by CAFs induces tumor angiogenesis by stimulating adjacent stromal fibroblasts [45, 46].

Furthermore, our findings revealed that, similar to VEGFA + CAF, CapECs were also enriched in the tumor. Additionally, cell communication analysis indicated that there was a robust cellular communication between the two, which was primarily constituted by the VEGF pathway. This suggests that VEGFA + CAF may facilitate angiogenesis, as well as tumor progression and metastasis, through cellular communication with CapECs. Furthermore, the NicheNet results indicated that the primary function of the downstream target genes of CapECs following the receipt of VEGFA ligands was vasculogenesis, thereby reinforcing the notion that VEGFA + CAF plays a pivotal role in angiogenesis.

The relationship between VEGFA + CAF and patient prognosis was validated using multiple datasets, as well as VEGFA + CAF abundance and VEGFA + CAF score. The results demonstrated that VEGFA + CAF was an independent risk factor for patient prognosis across multiple datasets. Importantly, the abundance of VEGFA + CAF also increased with tumor stage progression, highlighting its potential role in driving tumor progression and metastasis. These findings underline the critical interplay between VEGFA + CAF and tumor dynamics, suggesting that VEGFA + CAFs may serve as both a driver and a marker of tumor progression. Furthermore, we constructed a machine learning model to predict the prognosis and response to sorafenib in HCC patients with good accuracy. This provides a valuable tool for accurate diagnosis and personalized treatment of HCC patients. However, one limitation of this study is the preliminary use of the LogiBoost model to predict therapy response. While the model served as a proof-of-concept to demonstrate the potential application of VEGFA + CAF-associated features in clinical settings, further exploration and validation are required to enhance its clinical relevance.

In conclusion, our study revealed three fibroblast subpopulations in HCC. Specifically, we determined that VEGFA + CAF were predominantly distributed within the tumor and were significantly associated with patient survival. The hypoxic microenvironment may be the main factor inducing VEGFA + CAF, which promotes vasculogenesis and tumor proliferation through cellular communication with CapECs and facilitates tumor progression and metastasis. A machine learning model for predicting the prognosis and response to sorafenib in HCC patients using the highly expressed gene of VEGFA + CAF has high accuracy and provides a powerful tool for the precise diagnosis and personalized treatment.

Abbreviations

CAFs	Cancer-associated fibroblasts
HCC	Hepatocellular carcinoma
TME	Tumor microenvironment
scRNA-seq	Single-cell RNA sequencing
CapECs	Capillary endothelial cells
ST	Spatial transcriptomics
TAMs	Tumor-associated macrophages
MDSC	Myeloid-derived suppressor cells
AL	Adjacent liver
CNCB	Chinese National Center for Biological Information
GEO	Gene Expression Omnibus
TCGA	The Cancer Genome Atlas
ICGC	International Cancer Genome Consortium
NODE	National Omics Data Encyclopedia
OS	Overall survival
RFS	Recurrence-free survival
apCAF	Antigen-presenting CAFs
ECM	Extracellular matrix
mCAF	Matrix CAFs
ArtECs	Arterial endothelial cells
VenECs	Venous endothelial cells
LSECs	Liver1 sinusoidal endothelial cells
LECs	Lymphatic endothelial cells
pECs	Proliferating endothelial cells
Treg	Regulatory T-cell

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12967-025-06192-0.

Supplementary Material 1: Fig. 1. Isolation of fibroblasts in HCC. (A) Bar graph showing the main demographic data of the discovery cohort (B) Gene, Counts, and mitochondria-related genes per cell subpopulation of the discovery cohort after quality control (C) Box plot showing the Intra-Sample KNN overlap of the raw data and the adjusted by harmony (D) UMAP showing cell clustering of mesenchymal cells. (E) UMAP showing the expression of canonical markers pericytes (RGS5), fibroblasts (DCN/LUM) and smooth muscle cells (MYH11) in subpopulations of mesenchymal cells (F) Dot plot showing the expression of recognised genes in fibroblasts, mural cells, and smooth muscle cells in the mesenchymal subpopulation.(G) Dot plot showing the widely recognised marker of fibroblast in the fibroblast subpopulations of the discovery cohort (H) UMAP showing the cellular clustering of the cell types in the validation cohort (I) Gene, Counts, and mitochondria-related genes in each cell subpopulation of the validation cohort following cell quality control. subpopulation of gene, Counts, and mitochondria-related gene occupancy (J) UMAP shows the cell clustering of the validation cohort of mesenchymal cells. (K) UMAP showing the expression of canonical markers pericytes (RGS5), fibroblasts (DCN/LUM), and smooth muscle cells (MYH11) in the validation cohort MSC subpopulations (L) Scatterplot showing the classification of fibroblasts and mural cells based on the canonical gene markers in the validation cohort. (M) Bar graph showing the percentage of fibroblasts/mural cells as well as undetermined cells in each mesenchymal subpopulation in the validation cohort based on the classification results.(N) UMAP plot showing the results of the demarcation of fibroblasts and mural cells in the validation cohort using the consensus gene signature.

Supplementary Material 2: Fig. 2. Subpopulation and functional enrichment analysis of fibroblasts. (A) UMAP shows that fibroblasts in the validation cohort can be clustered into four subpopulations. (B) Dot plot showing the expression of highly expressed genes of VEGFA + CAF, MMP11 + CAF, and HLA-DRB1 + CAF, in fibroblast subpopulations in the validation cohort. (C) UMAP plot showing the expression of VEGFA, MMP1, and HLA-DRB1 in the fibroblast subpopulation of the validation cohort (D) Bar graph showing the proportion of different matrix body components differentially expressed in each subpopulation.

Supplementary Material 3: Fig. 3. Spatial distribution of three fibroblast subpopulations. (A) Bar graph showing the proportion of the three

Liu *et al. Journal of Translational Medicine*   (2025) 23:198

Page 18 of 20

fibroblasts to the total number of fibroblasts in AL and tumour tissues (B) Bar graph showing the comparison of the scores of the three fibroblasts in the AL and tumour tissues in TCGA. (C-E)Scatter plots showing the linear relationship between CIBERSORTx estimates and true abundance for each fibroblast subpopulation. (F) In HCC2_T and HCC5_3_T slides the spatial distribution of MMP11 + CAF, VEGFA + CAF (G) in HCC1_N and HCC3_N slides the spatial distribution of HLA-DRB1 + CAF, MMP11 + CAF. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; ****, $P < 0.0001$.

Supplementary Material 4: Fig. 4. Cell-state transition trajectory inference of CAFs. (A) UMAP plot demonstrating the order of differentiation of the three fibroblasts analysed using Monocel2, and pseudotemporal distribution (B) Violin plot demonstrating the pesudotime comparison of the three fibroblasts calculated using Monocle2. (C-E) Bar plot showing top 5 terms or pathways significantly enriched for the three modules of progenitor, activation and differentiation (F) Box line plot showing the expression of the three modules in cells grouped by tissue type and pseudotime tertile 3. (G) Androgen, EGFR, Estrogen, JAK-STAT, MAPK, NFκB, p53, P13K, TNFa, Trail pathway scores in paracancerous and tumour tissues over time.

Supplementary Material 5: Fig. 5. VEGFA + CAF is associated with poorer prognosis in HCC patients. (A) Bar graph showing the percentage of Scisso- cells to the fibroblast subpopulation (B) Bar graph showing the overall relationship of the three types of fibroblasts to OS in patients (C) Dot plot showing the expression of highly expressed genes in the three types of cells presenting as Scisso + as well as Scissor- cells.

Supplementary Material 6: Fig. 6. Analysis of cellular communication between VEGFA + CAF and endothelial cells. (A) Number and strength of cellular communication between endothelial and fibroblast subpopulations (B) Heatmap showing for endothelial and fibroblast subpopulations the effect on the VEGF pathway (C) Strength of cellular communication between VEGFA + CAF and other cell types in term of VEGF pathway. (D) Heatmap showing for cell communication between endothelial and fibroblasts in term of VEGF pathway. (E) Demonstration of the effect on the VEGF pathway for the ligand-receptor pair, VEGFA-VEGFR1, of VEGFA + CAF and other cell types (F) Combined heatmap illustrating the results of the NicheNet analysis of CapECS and fibroblasts. The initial section of the combined plot illustrates the Pearson's coefficient of the fibroblast ligand, with elevated coefficients signifying a robust capacity of the ligand to regulate the CapECS target genes. The subsequent section depicts the expression of the ligand across distinct fibroblast subtypes, while the fourth section elucidates the regulatory potential of the target genes.(G)Bar plot showing top 5 terms or pathways significantly enriched for reciver genes inCapEcs (H) Bar plot showing top 5 terms or pathways significantly enriched for target genes in CapEcs.

Supplementary Material 7: Fig. 7. Predicting prognosis and making therapeutic decision choices for HCC patients by machine learning using highly expressed genes in VEGFA + CAF. (A Box plots demonstrating VEGFA + CAF scores in HCC patients with different staging in three cohortsTCGA,GSE14520,ICGC. (B) AUC curves demonstrating the predictive efficacy of using machine learning to predict 1- and 3-year survival in HCC patients in four datasets (C) Bar charts demonstrating the AUC values of using machine learning to predict 1- and 3-year survival in HCC patients in four datasets (D) Forest plots demonstrating the VEGFA + CAF risk scores were significantly associated with patients' OS in all four datasets (E) AUC values for predicting HCC response with sorafenib using seven machine learning in the training and test sets are shown using histograms. *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$; ****, $p < 0.0001$.

Supplementary Material 8: Table 1. Differential genes in endothelial cells subtypes. Supplementary Table 2.Differential genes in fibroblast subtypes. Supplementary Table 3.Differential genes in scissor subtypes. Supplementary Table 4.Differential genes in endothelial cells subtypes.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]School of Medicine, Nankai University, Tianjin 300071, China. [2]Department of Ultrasound, Fifth Medical Center of Chinese, PLA General Hospital, Beijing 100853, China. [3]Department of Interventional Ultrasound, First Medical Center of Chinese, PLA General Hospital, Beijing 100853, China.

References

1. Torimura T, Iwamoto H. Treatment and the prognosis of hepatocellular carcinoma in Asia. Liver Int. 2022;42:2042–54. https://doi.org/10.1111/liv.15130.
2. Gordan JD, Kennedy EB, Abou-Alfa GK, Beg MS, Brower ST, Gade TP, Goff L, Gupta S, Guy J, Harris WP, et al. Systemic therapy for advanced hepatocellular carcinoma: ASCO Guideline. J Clin Oncol. 2020;38:4317–45. https://doi.org/10.1200/jco.20.02672.
3. Affo S, Yu LX, Schwabe RF. The role of cancer-associated fibroblasts and fibrosis in liver cancer. Annu Rev Pathol. 2017;12:153–86. https://doi.org/10.1146/annurev-pathol-052016-100322.
4. Yin Z, Dong C, Jiang K, Xu Z, Li R, Guo K, Shao S, Wang L. Heterogeneity of cancer-associated fibroblasts and roles in the progression, prognosis, and therapy of hepatocellular carcinoma. J Hematol Oncol. 2019;12:101. https://doi.org/10.1186/s13045-019-0782-x.
5. Song M, He J, Pan QZ, Yang J, Zhao J, Zhang YJ, Huang Y, Tang Y, Wang Q, He J, et al. Cancer-associated fibroblast-mediated cellular crosstalk supports hepatocellular carcinoma progression. Hepatology. 2021;73:1717–35. https://doi.org/10.1002/hep.31792.
6. Affo S, Filliol A, Gores GJ, Schwabe RF. Fibroblasts in liver cancer: functions and therapeutic translation. Lancet Gastroenterol Hepatol. 2023;8:748–59. https://doi.org/10.1016/s2468-1253(23)00111-5.

Liu *et al. Journal of Translational Medicine*   (2025) 23:198

Page 19 of 20

7. Rao A, Barkley D, França GS, Yanai I. Exploring tissue architecture using spatial transcriptomics. Nature. 2021;596:211–20. https://doi.org/10.1038/s41586-021-03634-9.

8. Tian H, Liu H, Zhu Y, Xing D, Wang B. The trends of single-cell analysis: a global study. Biomed Res Int. 2020;2020:7425397. https://doi.org/10.1155/2020/7425397.

9. Long F, Zhong W, Zhao F, Xu Y, Hu X, Jia G, Huang L, Yi K, Wang N, Si H, et al. DAB2 (+) macrophages support FAP (+) fibroblasts in shaping tumor barrier and inducing poor clinical outcomes in liver cancer. Theranostics. 2024;14:4822–43. https://doi.org/10.7150/thno.99046.

10. Wang H, Liang Y, Liu Z, Zhang R, Chao J, Wang M, Liu M, Qiao L, Xuan Z, Zhao H, Lu L. POSTN(+) cancer-associated fibroblasts determine the efficacy of immunotherapy in hepatocellular carcinoma. J Immunother Cancer. 2024;12: e008721. https://doi.org/10.1136/jitc-2023-008721.

11. Zhu GQ, Tang Z, Huang R, Qu WF, Fang Y, Yang R, Tao CY, Gao J, Wu XL, Sun HX, et al. CD36(+) cancer-associated fibroblasts provide immunosuppressive microenvironment for hepatocellular carcinoma via secretion of macrophage migration inhibitory factor. Cell Discov. 2023;9:25. https://doi.org/10.1038/s41421-023-00529-z.

12. Xue R, Zhang Q, Cao Q, Kong R, Xiang X, Liu H, Feng M, Wang F, Cheng J, Li Z, et al. Liver tumour immune microenvironment subtypes and neutrophil heterogeneity. Nature. 2022;612:141–7. https://doi.org/10.1038/s41586-022-05400-x.

13. Lu Y, Yang A, Quan C, Pan Y, Zhang H, Li Y, Gao C, Lu H, Wang X, Cao P, et al. A single-cell atlas of the multicellular ecosystem of primary and metastatic hepatocellular carcinoma. Nat Commun. 2022;13:4594. https://doi.org/10.1038/s41467-022-32283-3.

14. Ma L, Heinrich S, Wang L, Keggenhoff FL, Khatib S, Forgues M, Kelly M, Hewitt SM, Saif A, Hernandez JM, et al. Multiregional single-cell dissection of tumor and immune cells reveals stable lock-and-key features in liver cancer. Nat Commun. 2022;13:7533. https://doi.org/10.1038/s41467-022-35291-5.

15. Ma L, Wang L, Khatib SA, Chang CW, Heinrich S, Dominguez DA, Forgues M, Candia J, Hernandez MO, Kelly M, et al. Single-cell atlas of tumor cell evolution in response to therapy in hepatocellular carcinoma and intrahepatic cholangiocarcinoma. J Hepatol. 2021;75:1397–408. https://doi.org/10.1016/j.jhep.2021.06.028.

16. Sharma A, Seow JJW, Dutertre CA, Pai R, Blériot C, Mishra A, Wong RMM, Singh GSN, Sudhagar S, Khalilnezhad S, et al. Onco-fetal reprogramming of endothelial cells drives immunosuppressive macrophages in hepatocellular carcinoma. Cell. 2020;183:377-394.e321. https://doi.org/10.1016/j.cell.2020.08.040.

17. Wu R, Guo W, Qiu X, Wang S, Sui C, Lian Q, Wu J, Shan Y, Yang Z, Yang S, et al. Comprehensive analysis of spatial architecture in primary liver cancer. Sci Adv. 2021;7: eabg3750. https://doi.org/10.1126/sciadv.abg3750.

18. Roessler S, Jia HL, Budhu A, Forgues M, Ye QH, Lee JS, Thorgeirsson SS, Sun Z, Tang ZY, Qin LX, Wang XW. A unique metastasis gene signature enables prediction of tumor relapse in early-stage hepatocellular carcinoma patients. Cancer Res. 2010;70:10202–12. https://doi.org/10.1158/0008-5472.Can-10-2607.

19. Wolf FA, Angerer P, Theis FJ. SCANPY: large-scale single-cell gene expression data analysis. Genome Biol. 2018;19:15. https://doi.org/10.1186/s13059-017-1382-0.

20. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat Biotechnol. 2018;36:411–20. https://doi.org/10.1038/nbt.4096.

21. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, Feng T, Zhou L, Tang W, Zhan L, et al. clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. Innovation (Camb). 2021;2:100141. https://doi.org/10.1016/j.xinn.2021.100141.

22. Schubert M, Klinger B, Klünemann M, Sieber A, Uhlitz F, Sauer S, Garnett MJ, Blüthgen N, Saez-Rodriguez J. Perturbation-response genes reveal signaling footprints in cancer gene expression. Nat Commun. 2018;9:20. https://doi.org/10.1038/s41467-017-02391-6.

23. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinformatics. 2013;14:7. https://doi.org/10.1186/1471-2105-14-7.

24. Newman AM, Steen CB, Liu CL, Gentles AJ, Chaudhuri AA, Scherer F, Khodadoust MS, Esfahani MS, Luca BA, Steiner D, et al. Determining cell type abundance and expression from bulk tissues with digital cytometry. Nat Biotechnol. 2019;37:773–82. https://doi.org/10.1038/s41587-019-0114-2.

25. Sun D, Guan X, Moran AE, Wu LY, Qian DZ, Schedin P, Dai MS, Danilov AV, Alumkal JJ, Adey AC, et al. Identifying phenotype-associated subpopulations by integrating bulk and single-cell sequencing data. Nat Biotechnol. 2022;40:527–38. https://doi.org/10.1038/s41587-021-01091-3.

26. Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, Lennon NJ, Livak KJ, Mikkelsen TS, Rinn JL. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat Biotechnol. 2014;32:381–6. https://doi.org/10.1038/nbt.2859.

27. Street K, Risso D, Fletcher RB, Das D, Ngai J, Yosef N, Purdom E, Dudoit S. Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. BMC Genomics. 2018;19:477. https://doi.org/10.1186/s12864-018-4772-0.

28. Wei R, He S, Bai S, Sei E, Hu M, Thompson A, Chen K, Krishnamurthy S, Navin NE. Spatial charting of single-cell transcriptomes in tissues. Nat Biotechnol. 2022;40:1190–9. https://doi.org/10.1038/s41587-022-01233-1.

29. Jin S, Guerrero-Juarez CF, Zhang L, Chang I, Ramos R, Kuan CH, Myung P, Plikus MV, Nie Q. Inference and analysis of cell–cell communication using Cell Chat. Nat Commun. 2021;12:1088. https://doi.org/10.1038/s41467-021-21246-9.

30. Browaeys R, Saelens W, Saeys Y. NicheNet: modeling intercellular communication by linking ligands to target genes. Nat Methods. 2020;17:159–62. https://doi.org/10.1038/s41592-019-0667-5.

31. Liu H, Zhang W, Zhang Y, Adegboro AA, Fasoranti DO, Dai L, Pan Z, Liu H, Xiong Y, Li W, et al. Mime: a flexible machine-learning framework to construct and visualize models for clinical characteristics prediction and feature selection. Comput Struct Biotechnol J. 2024;23:2798–810. https://doi.org/10.1016/j.csbj.2024.06.035.

32. Elyada E, Bolisetty M, Laise P, Flynn WF, Courtois ET, Burkhart RA, Teinor JA, Belleau P, Biffi G, Lucito MS, et al. Cross-species single-cell analysis of pancreatic ductal adenocarcinoma reveals antigen-presenting cancer-associated fibroblasts. Cancer Discov. 2019;9:1102–23. https://doi.org/10.1158/2159-8290.Cd-19-0094.

33. Bartoschek M, Oskolkov N, Bocci M, Lövrot J, Larsson C, Sommarin M, Madsen CD, Lindgren D, Pekar G, Karlsson G, et al. Spatially and functionally distinct subclasses of breast cancer-associated fibroblasts revealed by single cell RNA sequencing. Nat Commun. 2018;9:5150. https://doi.org/10.1038/s41467-018-07582-3.

34. Abdelgalil AA, Alkahtani HM, Al-Jenoobi FI. Sorafenib. Profiles Drug Subst Excip Relat Methodol. 2019;44:239–66. https://doi.org/10.1016/bs.podrm.2018.11.003.

35. Chhabra Y, Weeraratna AT. Fibroblasts in cancer: unity in heterogeneity. Cell. 2023;186:1580–609. https://doi.org/10.1016/j.cell.2023.03.016.

36. Sebastian A, Hum NR, Martin KA, Gilmore SF, Peran I, Byers SW, Wheeler EK, Coleman MA, Loots GG. Single-cell transcriptomic analysis of tumor-derived fibroblasts and normal tissue-resident fibroblasts reveals fibroblast heterogeneity in breast cancer. Cancers (Basel). 2020;12:1307. https://doi.org/10.3390/cancers12051307.

37. Kerdidani D, Aerakis E, Verrou KM, Angelidis I, Douka K, Maniou MA, Stamoulis P, Goudevenou K, Prados A, Tzaferis C, et al. Lung tumor MHCII immunity depends on in situ antigen presentation by fibroblasts. J Exp Med. 2022;219. https://doi.org/10.1084/jem.20210815.

38. Motrescu ER, Rio MC. Cancer cells, adipocytes and matrix metalloproteinase 11: a vicious tumor progression cycle. Biol Chem. 2008;389:1037–41. https://doi.org/10.1515/bc.2008.110.

39. Zhang X, Huang S, Guo J, Zhou L, You L, Zhang T, Zhao Y. Insights into the distinct roles of MMP-11 in tumor biology and future therapeutics (Review). Int J Oncol. 2016;48:1783–93. https://doi.org/10.3892/ijo.2016.3400.

40. Wu L, Yan J, Bai Y, Chen F, Zou X, Xu J, Huang A, Hou L, Zhong Y, Jing Z, et al. An invasive zone in human liver cancer identified by Stereo-seq promotes hepatocyte-tumor cell crosstalk, local immunosuppression and tumor progression. Cell Res. 2023;33:585–603. https://doi.org/10.1038/s41422-023-00831-1.

41. Liu Y, Xun Z, Ma K, Liang S, Li X, Zhou S, Sun L, Liu Y, Du Y, Guo X, et al. Identification of a tumour immune barrier in the HCC microenvironment

Liu *et al. Journal of Translational Medicine*   (2025) 23:198

Page 20 of 20

that determines the efficacy of immunotherapy. J Hepatol. 2023;78:770–82. https://doi.org/10.1016/j.jhep.2023.01.011.

42. De Francesco EM, Lappano R, Santolla MF, Marsico S, Caruso A, Maggiolini M. HIF-1α/GPER signaling mediates the expression of VEGF induced by hypoxia in breast cancer associated fibroblasts (CAFs). Breast Cancer Res. 2013;15:R64. https://doi.org/10.1186/bcr3458.

43. Ide T, Kitajima Y, Miyoshi A, Ohtsuka T, Mitsuno M, Ohtaka K, Koga Y, Miyazaki K. Tumor-stromal cell interaction under hypoxia increases the invasiveness of pancreatic cancer cells through the hepatocyte growth factor/c-Met pathway. Int J Cancer. 2006;119:2750–9. https://doi.org/10.1002/ijc.22178.

44. Ziani L, Buart S, Chouaib S, Thiery J. Hypoxia increases melanoma-associated fibroblasts immunosuppressive potential and inhibitory effect on T cell-mediated cytotoxicity. Oncoimmunology. 2021;10:1950953. https://doi.org/10.1080/2162402x.2021.1950953.

45. Nagasaki T, Hara M, Nakanishi H, Takahashi H, Sato M, Takeyama H. Interleukin-6 released by colon cancer-associated fibroblasts is critical for tumour angiogenesis: anti-interleukin-6 receptor antibody suppressed angiogenesis and inhibited tumour–stroma interaction. Br J Cancer. 2014;110:469–78. https://doi.org/10.1038/bjc.2013.748.

46. Nilsson MB, Langley RR, Fidler IJ. Interleukin-6, secreted by human ovarian carcinoma cells, is a potent proangiogenic cytokine. Cancer Res. 2005;65:10794–800. https://doi.org/10.1158/0008-5472.Can-05-0623.

Publisher's Note