

RESEARCH

Open Access



Pan-cancer spatially resolved single-cell analysis reveals the crosstalk between cancer-associated fibroblasts and tumor microenvironment

Chenxi Ma¹, Chengzhe Yang^{2,3}, Ai Peng¹, Tianyong Sun¹, Xiaoli Ji⁴, Jun Mi¹, Li Wei¹, Song Shen¹ and Qiang Feng^{1,5*}

Abstract

Cancer-associated fibroblasts (CAFs) are a heterogeneous cell population that plays a crucial role in remodeling the tumor microenvironment (TME). Here, through the integrated analysis of spatial and single-cell transcriptomics data across six common cancer types, we identified four distinct functional subgroups of CAFs and described their spatial distribution characteristics. Additionally, the analysis of single-cell RNA sequencing (scRNA-seq) data from three additional common cancer types and two newly generated scRNA-seq datasets of rare cancer types, namely epithelial-myoeplithelial carcinoma (EMC) and mucoepidermoid carcinoma (MEC), expanded our understanding of CAF heterogeneity. Cell–cell interaction analysis conducted within the spatial context highlighted the pivotal roles of matrix CAFs (mCAFs) in tumor angiogenesis and inflammatory CAFs (iCAFs) in shaping the immunosuppressive microenvironment. In patients with breast cancer (BRCA) undergoing anti-PD-1 immunotherapy, iCAFs demonstrated heightened capacity in facilitating cancer cell proliferation, promoting epithelial-mesenchymal transition (EMT), and contributing to the establishment of an immunosuppressive microenvironment. Furthermore, a scoring system based on iCAFs showed a significant correlation with immune therapy response in melanoma patients. Lastly, we provided a web interface (<https://chenxisd.shinyapps.io/pancaf/>) for the research community to investigate CAFs in the context of pan-cancer.

Keywords Spatial transcriptomics, Single-cell RNA sequencing, Pan-cancer analysis, Cancer-associated fibroblasts, Tumor microenvironment, Tumor immunotherapy

*Correspondence:

Qiang Feng
fengqiang@sdu.edu.cn

¹ Department of Human Microbiome and Periodontology and Implantology and Orthodontics, School and Hospital of Stomatology, Cheeloo College of Medicine, Shandong University and Shandong Key Laboratory of Oral Tissue Regeneration and Shandong Engineering Laboratory for Dental Materials and Oral Tissue Regeneration and Shandong Provincial Clinical Research Center for Oral Diseases, Jinan 250012, China

² Department of Oral and Maxillofacial Surgery, Qilu Hospital of Shandong University, Jinan, Shandong, China

³ Institute of Stomatology, Shandong University, Jinan, Shandong, China

⁴ Department of Stomatology, Central Hospital Affiliated to Shandong First Medical University, No.105 Jiefang Road, Jinan, Shandong, China

⁵ State Key Laboratory of Microbial Technology, Shandong University, Qingdao 266237, China



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Introduction

Tumors display extensive heterogeneity, with cancer cells engaging in reciprocal interactions with their microenvironment, forming a complex ecosystem [1]. Cancer-associated fibroblasts (CAFs), as one of the most prominent and abundant cell populations in the tumor microenvironment (TME) [2], have garnered significant attention in recent years. CAFs' intricate interactions with stromal components and immune cells play a crucial role in orchestrating TME reorganization, encompassing processes such as angiogenesis, extracellular matrix (ECM) remodeling, and immune evasion [3–5]. At present, the crucial role of CAFs has been largely overlooked by most therapies, including immunotherapy and chemotherapy [3]. Our current understanding of the interplay between CAFs and components of TME is insufficient to support the development of reliable treatment strategies. Further research is needed to deepen our understanding of these interactions and pave the way for effective therapeutic interventions.

In recent years, the application of single-cell transcriptomics has unraveled the heterogeneity of CAFs within many cancer types, such as bladder carcinoma (BC) [6], head and neck squamous cell carcinoma (HNSCC) [7], papillary thyroid carcinoma (PTC) [8], and lung cancer (LC) [9]. Furthermore, two recent unbiased studies based on single-cell RNA sequencing (scRNA-seq) explored the heterogeneity and plasticity of CAFs from a pan-cancer perspective and revealed the conservation of CAF phenotypes across cancer types [10, 11]. Although scRNA-seq provides an unprecedented opportunity to systematically dissect the heterogeneity of CAFs, the loss of spatial information during tissue dissociation hinders the study of the crosstalk between CAFs and TME. Recently developed spatial transcriptomics (ST) can obtain whole-transcriptome data within tissue sections, thereby preserving the spatial position information of cells [12]. Therefore, orthogonal integration of scRNA-seq data and ST data will help determine the spatial distribution characteristics of CAFs and further dissect the cellular communication between CAFs and TME.

In this study, we have delineated the landscape of CAFs in six common cancer types and described the unique functional features of these subtypes. We also analyzed

scRNA-seq data of three additional common tumors and two newly sequenced rare tumors to expand our understanding of CAF heterogeneity. A spatial single-cell transcriptomic atlas spanning six tumors, including 744,289 cells, generated by integrating scRNA-seq data and ST data was used to describe the spatial distribution characteristics of CAFs and to characterize the complex interactions between CAFs and TME. Notably, a score generated based on inflammatory CAFs (iCAFs) showed a significant correlation with the response of melanoma patients to immunotherapy. In summary, our integrated data resources provide novel insights and guidance for the development of therapeutic strategies targeting CAFs in TME.

Results

Construction of a pan-cancer spatial single-cell transcriptome atlas

To establish a spatial single-cell landscape in pan-cancer, we acquired scRNA-seq data from 69 samples of 56 patients diagnosed with one of the six prevalent cancer types, along with ST data from 22 tissue slices of 22 patients (Fig. 1a and b; Table S1 and S2). Among them, the ST data of 10 tissue slices had corresponding scRNA-seq data from the same patient (Fig. 1c; Table S1 and S2). The data we collected included six types of cancer: BRCA, colorectal cancer (CRC), liver hepatocellular carcinoma (LIHC), ovarian cancer (OVCA), prostate adenocarcinoma (PRAD), and uterine corpus endometrial carcinoma (UCEC) (Fig. 1a; Table S1 and S2). After strict quality control and filtration, a total of 163,919 cells in the scRNA-seq data and 59,529 spots in ST data were retained for downstream analysis (Fig. 1d and S1a). In the scRNA-seq dataset, the median number of unique molecular identifiers (UMIs) per cell was 3955, and the median number of genes per cell was 1425 (Figure S1b and c). For ST analysis, the median number of UMIs per spot was 11,139 and the median number of genes per spot was 3,863 (Figure S1d and e). To minimize the batch effect between different scRNA-seq datasets, we independently analyzed each dataset. Taking CRC as an example, we used graph-based clustering and identified seven major clusters based on typical markers of different cell types (Table S3), including epithelial cells,

(See figure on next page.)

Fig. 1 A pan-cancer spatial single-cell transcriptome atlas. **a** Schematic depicting the study design. The cancer types included in this pan-cancer study were displayed in the first image on the left, created by Figdraw. **b** The number of samples in the pan-cancer analysis of scRNA-seq and ST. **c** Pie chart showing the proportion of ST sections that have corresponding scRNA-seq data from the same patient compared to those without such corresponding scRNA-seq data. **d** The number of cells in the pan-cancer analysis of scRNA-seq and ST. **e** Uniform Manifold Approximation and Projection (UMAP) plots showing the major cell types in CRC. **f** Bubble heatmap showing the expression of marker genes for the major cell types in CRC. **g** Spatial cell charting of CRC using CellTrek

fibroblasts, endothelial cells, T&NK, B cells, myeloid cells and mast cells (Fig. 1e and f). CopyKAT was used to estimate the single-cell copy number variation (CNV) landscape of tumors, in order to distinguish malignant epithelium from non-malignant epithelium (Fig. 1e). The myeloid cells were further divided into monocytes, macrophages and dendritic cells (Fig. 1e). The CD8+T cells, CD4+T cells, regulatory T cells (Treg cells) and natural killer cells (NK cells) were identified from the T&NK cluster (Fig. 1e). Similarly, cells from the other 5 types of cancer were clustered into roughly the same subgroups (Figure S2, S3, S4 and S5). Of note, we detected neutrophils in the scRNA-seq data from the inDrop platform, which were not detected in the scRNA-seq data from the 10×Genomics platform (Figure S2). Neutrophils are very fragile and have low RNA content [13], which may be the main reason for the capture failure.

CellTrek is a computational toolkit that enables direct mapping of individual cells back to their spatial coordinates in tissue sections based on scRNA-seq and ST data [14]. Unlike ST deconvolution methods, this approach transferred ST coordinates to single cells, thereby achieving single-cell resolution [14]. We applied it to quality-controlled scRNA-seq and ST data in pan-cancer to reconstruct spatial single-cell atlases. Even without corresponding scRNA-seq data from the same patient, ST datasets were still largely covered by scRNA-seq datasets based on the co-embedding results (Figure S6). Due to some cells being repeatedly mapped, we ultimately obtained a pan-cancer spatial single-cell transcriptomic atlas containing 744,289 cells (Fig. 1d and g).

CAF heterogeneity in pan-cancer

To compare the similarity of the main cell lineages of different cancer types, we constructed a phylogenetic tree (Figure S7a). Compared with the biased distribution of epithelial cells, fibroblasts from different cancer types clustered together (Figure S7a), indicating that fibroblasts had similar transcriptional features in different cancer types. Interestingly, NK cells and B cells originating from UCEC demonstrated unique features (Figure S7a), implying that the TMEs across diverse cancer types could have potentially exerted distinct effects on immune cell phenotypes. We subsequently investigated the heterogeneity of fibroblasts in scRNA-seq datasets of the 6 cancer types (Fig. 2a). The reclustering of the fibroblast cluster identified four CAF subtypes, as well as pericytes and smooth muscle cells (SMCs) (Fig. 2a). After applying Harmony for batch correction, all cells with local inverse Simpson's Index (LISI) greater than 1 indicate that no obvious batch effects were observed (Figure S8). CFD+ fibroblasts showed high expression of chemokines (CCL11, CXCL12, and CXCL14) (Fig. 2b; Table S4),

similar to the previously reported iCAFs in various types of tumors such as BC [6] and PTC [8]. GO enrichment analysis of its marker genes showed their association with the response to mechanical stimulation, reactive oxygen species, epithelial cell proliferation, immune system, and cell migration (Fig. 2c). POSTN+ fibroblasts showed high expression levels of several ECM remodeling genes (MMP11, CTHRC1, COL1A1, COL1A2, COL3A1, COL10A1, and COL11A1) and enriched signatures of ECM (Fig. 2b and c; Table S4), which were consistent with the previously reported matrix CAFs (mCAFs) in cervical squamous cell carcinoma (CESC) [15]. Interestingly, a cluster of cells was related to the response to hypoxia and canonical glycolysis (Fig. 2c), resembling the reported metabolic CAFs (meCAFs) in pancreatic ductal adenocarcinoma (PDAC) [16]. Notably, we also found a cluster of cells that exhibited higher expression of a set of cell cycle-related genes (CENPE, NUSAP1, PTTG1, STMN1, TOP2A, and TUBA1B) (Fig. 2b; Table S4), which was consistent with proliferative CAFs (pCAFs) in a previous pan-cancer study of CAFs [10]. Immunofluorescence on tissue microarrays from BRCA patients further substantiated the existence of the four CAF subtypes (Figure S9). Next, we further investigated the heterogeneity of CAFs using the AUCell algorithm, based on the functional features of CAFs summarized by Lavie et al. [17] (Fig. 2d; Table S5). iCAFs exhibited the highest activity in immune-related functions, including complement activation, chemokine production, and inflammatory response (Fig. 2d). Additionally, the biological processes of angiogenesis, wound healing, regulation of ECM organization and collagen biosynthetic process were all enriched in mCAFs (Fig. 2d). As expected, meCAFs exhibited a high level of glycolytic activity (Fig. 2d). Interestingly, in addition to the cell cycle, pCAFs were also involved in IFN- γ production and muscle contraction (Fig. 2d).

Although iCAFs and mCAFs were the major CAFs celltypes across 6 cancer types, different subtypes of CAFs still exhibited significant cancer preferences (Fig. 2e and f). iCAFs were enriched in BRCA and CRC, whereas meCAFs were enriched in LIHC and OVCA (Fig. 2f). The other two subtypes, especially pCAFs, were enriched in OVCA (Fig. 2f). To investigate the presence of these fibroblast subtypes in other common cancer types, we obtained and analyzed publicly available scRNA-seq data from non-small cell lung cancer (NSCLC) [18] and melanoma [19] (Figure S10a and b). Moreover, we conducted scRNA-seq on tumor and adjacent non-tumor tissues from a patient with HNSCC and integrated it with previously published scRNA-seq data of the same cancer type [7] (Figure S10c; Table S6). The findings indicated that both iCAFs and mCAFs were observed

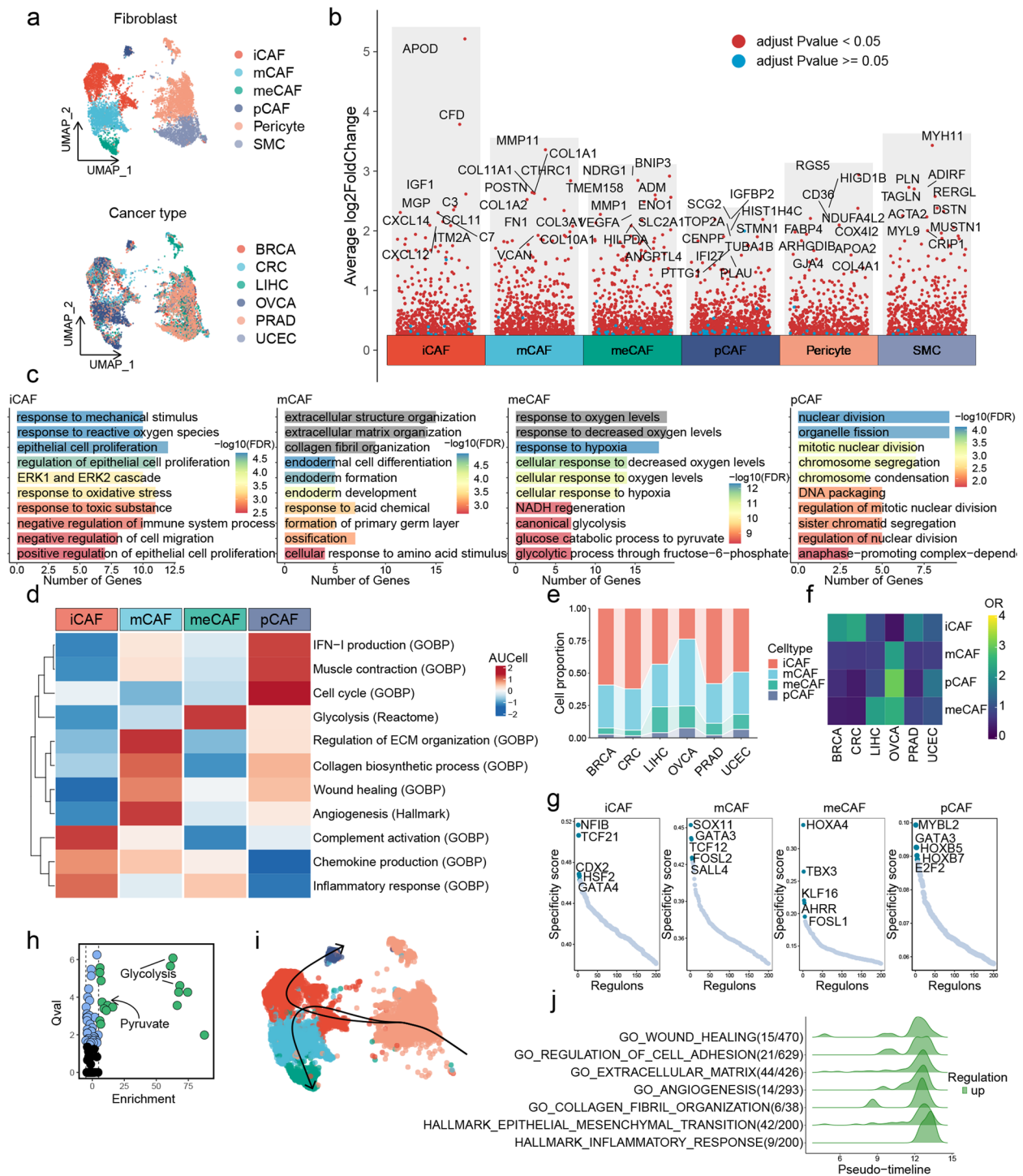


Fig. 2 CAF heterogeneity in pan-cancer. **a** UMAP plots showing the integration of fibroblasts across six different cancer types by Harmony. **b** Differential expression analysis showing the upregulated genes for each fibroblast subtype. An adjusted p value < 0.05 is indicated in red, while an adjusted p value ≥ 0.05 is indicated in blue. **c** GO enrichment analysis of upregulated genes in each CAF subtype. **d** Heatmap showing pathway activities scored by AUCell in each CAF subtype. **e** Proportion of CAF subtypes across multiple cancer types. **f** Heatmap showing the ORs of CAF subtypes in each cancer type. **g** Scatter plot showing the RSSs in each CAF subtype. The top 5 regulons are highlighted. **h** SCAP analysis of metabolic pathways in meCAFs. **i** Slingshot trajectory analysis of CAFs. **j** GeneSwitches analysis of pathway activity changes in the transition pathway from pericytes to iCAFs

in all three types of cancer (Figure S10a-c). Next, we performed scRNA-seq on three samples derived from two patients with two rare types of tumors that have not been previously studied by scRNA-seq, including one tumor tissue from epithelial-myoepithelial carcinoma (EMC), one tumor tissue from mucoepidermoid carcinoma (MEC), and one adjacent non-tumor tissue from MEC (Figure S10d and e; Table S6). iCAFs and mCAFs were also both identified in these two types of tumors (Figure S10d and e). To investigate the key differences between fibroblasts derived from tumor tissue and adjacent non-tumor tissue, we conducted differential expression analysis. The results revealed a significant upregulation of multiple marker genes of mCAFs in fibroblasts derived from tumor tissues of HNSCC patients, as compared to fibroblasts derived from adjacent non-tumor tissues (Figure S10f). Functional enrichment analysis results indicated that the upregulated genes were related to the ECM, which was also observed in MEC (Figure S10g-i).

Via SCENIC analysis, we determined essential motifs within the four subtypes of CAFs. The regulatory protein CDX2 [20] associated with inflammation and the regulatory protein TCF12 [21] related to ECM remodeling were enriched in iCAFs and mCAFs, respectively (Fig. 2g; Table S7). Additionally, we observed that mCAFs were enriched for the KLF16 [22] (Fig. 2g; Table S7), which was a known regulator of metabolism. Lastly, Cell-cycle-related regulons (MYBL2 and E2F2) were highly enriched in pCAF (Fig. 2g; Table S7). The metabolic correlation of mCAFs prompted us to perform SCPA analysis to study their metabolic pathway activity. As expected, glycolysis and pyruvate were enriched in the top metabolic pathway of mCAFs (Fig. 2h; Table S8). Previous studies have reported that CAFs provide energy to cancer cells through glycolysis in hypoxic TME [23, 24]. This reverse Warburg effect may be caused by mCAFs. In order to further investigate metabolic reprogramming of CAFs, we conducted scMetabolism analysis and identified various metabolic rewiring mechanisms related to tumor growth in different CAF subgroups [24]. mCAFs exhibited higher activity in fatty acid biosynthesis, while the TCA cycle was enriched in pCAFs (Figure S11). Besides glycolysis, metabolism of alanine, aspartate, and glutamate was also enriched in mCAFs (Figure S11).

The complexity of CAF cellular characteristics can be attributed to their highly heterogeneous origins [17, 25]. In addition to transformation from tissue-resident fibroblasts, pericytes are also an important source for the formation of CAFs [17, 25]. With Slingshot analysis, a potential transition pathway from pericytes to iCAFs was suggested (Fig. 2i). Compared to other subtypes of

CAF, iCAFs exhibit the lowest level of transcriptional homogeneity (Figure S7b), which may be attributed to their complex origins. Previous studies have indicated that the transition from pericytes to fibroblasts is closely associated with cancer invasion and metastasis [26]. GeneSwitches analysis identified multiple biological processes that were activated along the pathway from pericytes to iCAFs, including wound healing, regulation of cell adhesion, ECM, angiogenesis, collagen fibril organization, epithelial-mesenchymal transition (EMT), and inflammatory response (Fig. 2j). While our data suggests that CAFs derived from pericytes are iCAFs, further investigation is necessary to explore its possibility and underlying mechanisms.

Spatial distribution characteristics of CAFs

To determine the spatial distribution characteristics of CAFs, we added their cell subpopulation annotation information into the CellTrek object. As the cell ratio of mCAFs and pCAFs is very low, we first focused our analysis on iCAFs and mCAFs. Taking one tissue section each from OVCA (OVCA1) and CRC (CRC1) as examples, we observed a spatially exclusive phenomenon between the high-density areas of iCAFs and mCAFs (Fig. 3a-d), suggesting that the activation state of CAFs is related to their location within the TME. To dissect the spatial expression dynamics from high-density areas of iCAFs to high-density areas of mCAFs, we conducted spatial trajectory analysis in OVCA1 and CRC1. The results demonstrated a gradual change in the proportions of cells along the trajectory, accompanied by a gradual increase in features such as collagen biosynthetic process, regulation of ECM organization, wound healing, and angiogenesis (Figure S12a-f). In addition, our slingshot analysis revealed a potential transition path from iCAFs to mCAFs (Fig. 2i), which is consistent with previously reported lineage plasticity among CAF subpopulations [17]. Overall, these results suggest that the state of CAFs could potentially be influenced by the specific TME.

Robust rank aggregation (RRA) is an algorithm that integrates ranks to obtain a comprehensive ranking list [27]. We computed the spatial k-distance between all cells and the subpopulations of fibroblasts in each tissue section, sorted them from closest to farthest, and integrated them using the RRA algorithm to obtain a comprehensive ranking of all cells. Upon analysis, the four subtypes of CAFs exhibited the minimum spatial k-distance between them, whereas there were no notable differences in the ordering of immune cell subtypes (Fig. 3e-l). To further investigate the microenvironmental characteristics surrounding different CAF subtypes, cells within the top 10% of spatial k-distance from the fibroblasts were defined as "CAF-proximal cells", with all others classified

as "CAF-distal cells" (Fig. 3m). Then, through paired t-tests, we compared the proportion of cells between CAF-proximal and CAF-distal cells. As anticipated, there was an enrichment of other types of CAF subtypes surrounding each type of CAF subtype (Fig. 3m). Additionally, a higher density of pericytes was observed in the vicinity of CAFs (Fig. 3m), which serve as an important source of CAFs. Endothelial cells exhibited an increased abundance in proximity to iCAFs and mCAFs (Fig. 3m). This observation may be explained by the angiogenic effect of mCAFs and the potential transformational relationship between iCAFs and mCAFs. Notably, the proportion of epithelial cells decreased around all four CAF subtypes (Fig. 3m), implying that these subtypes are located farther away from the epithelial area. Moreover, a decrease in the proportion of certain immune cell subtypes was observed in the vicinity of CAFs, including a reduction in neutrophils and Tregs proportions around iCAFs, a decrease in Tregs proportions near mCAFs, and a lower proportion of B cells near pCAFs (Fig. 3m).

Effect of CAFs on TME through paracrine signaling

Given the high angiogenic activity of mCAFs, we employed Spatalk to explore the interplay between mCAFs and endothelial cells within the spatial context. Angiogenesis is a complex molecular process involving endothelial cell activation, proliferation, and migration to form new blood vessels and vasculature [28, 29]. Intriguingly, the ligands identified in mCAFs were found to play significant roles in various endothelial cell functions, including migration, proliferation, apoptosis, chemotaxis, differentiation, and development (Fig. 4a and S13). Our analysis further revealed a series of ligand-receptor interactions (LRIs) associated with angiogenesis, such as VEGFA-(FLT1+KDR+NRP1+ITGB1+ITGA9), VEGFB-(FLT1+NRP1), VEGFC-(FLT1+KDR+ITGB1), PGF-(NRP1+FLT1), and THBS1-(ITGA6+ITGB1+LRP5) [30, 31] (Fig. 4d and S13; Table S9). Collectively, these findings suggest that mCAFs exert their pro-angiogenic effects by modulating endothelial cell function through paracrine signaling.

In addition to their impact on angiogenesis within TME, CAFs also regulate immune cell responses to promote tumor growth and immune escape [17]. Ligands derived from iCAFs to macrophages were found to significantly enrich various macrophage functions, including differentiation, cytokine production, chemotaxis, migration, and activation (Fig. 4b and S14). Considering the close association between M2 macrophage polarization and tumor progression, our investigation focused on the influence of iCAFs on M2 macrophage polarization. We collected relevant ligands based on previous findings [32] (Table S10) and identified a series of LRIs involved in this process, such as TGFB1-(CXCR4+ITGAV+TGFB2+TGFB1+ITGB5+SDC2+SMAD3+ITGB8), TGFB2-(TGFB2+TGFB1+ACVR1), TGFB3-(ITGB1+TGFB2+ITGB5+TGFB1), CSF1-(CSF1R+SIRPA), IL34-CSF1R, and IL10-(IL10RA+IL10RB) [32, 33] (Fig. 4e and S14; Table S9).

CD8+T cells play a pivotal role in anti-tumor immunity, yet the mechanisms underlying the interaction between iCAFs and CD8+T cells remain elusive. Ligands from iCAFs that bind to CD8+T cells were enriched in various T cell-related functions, including migration, activation, proliferation, chemotaxis, differentiation, costimulation, apoptotic process, homeostasis, cytokine production, establishment of T cell polarity, T cell mediated immunity and cytotoxicity (Fig. 4c and S15). Notably, iCAFs may induce CD8+T cells apoptosis and impair their anti-tumor functions by interacting with CD8+T cells PTPRC receptors via Galectin-1 (LGALS1) [34, 35] (Fig. 4f and S15; Table S9). Additionally, iCAFs can suppress the activation and proliferation of CD8+T cells through macrophage migration inhibitory factor (MIF)-CXCR4 interaction [36] (Fig. 4f and S15; Table S9). Furthermore, iCAFs may secrete TGFB1 to inhibit the activation and proliferation of CD8+T cells [37] (Fig. 4f and S15; Table S9). Apart from their interactions with CD8+T cells and macrophages, iCAFs exhibited complex interplays with other immune cell populations, including B cells, dendritic cells, mast cells, neutrophils,

(See figure on next page.)

Fig. 3 Spatial distribution characteristics of CAFs. **a** Spatial cell charting of CAFs in OVCA1 using CellTrek. **b** Density plots showing high-density regions of iCAFs and mCAFs in OVCA1. **c** Spatial cell charting of CAFs in CRC1 using CellTrek. **d** Density plots showing high-density regions of iCAFs and mCAFs in CRC1. **e** Heatmap showing the average k-distance from different cell types to iCAFs in each tissue slice. The columns were scaled. **f** Integrated ranking of cell types based on proximity to iCAFs using RRA algorithm across 22 tissue slices. **g** Heatmap showing the average k-distance from different cell types to mCAFs in each tissue slice. The columns were scaled. **h** Integrated ranking of cell types based on proximity to mCAFs using RRA algorithm across 22 tissue slices. **i** Heatmap showing the average k-distance from different cell types to mCAFs in each tissue slice. The columns were scaled. **j** Integrated ranking of cell types based on proximity to mCAFs using RRA algorithm across 22 tissue slices. **k** Heatmap showing the average k-distance from different cell types to pCAFs in each tissue slice. The columns were scaled. **l** Integrated ranking of cell types based on proximity to pCAFs using RRA algorithm across 22 tissue slices. **m** Left: Circular plot showing the proportions of proximal and distal regions of CAFs. Right: heatmap showing the enrichment of various cell types in the proximal and distal regions for each CAF subpopulation. The paired t-test was used to compare the differences in cell proportions between the proximal and distal regions for each CAF subpopulation. Red color represents enrichment of a cell type in the proximal region of CAFs, while blue color represents enrichment of a cell type in the distal region of CAFs. Only p values < 0.05 are shown

spatial proximity (Fig. 4h). By analyzing datasets derived from The Cancer Genome Atlas (TCGA) database, we observed a significant positive correlation between LGALS1 and PDCD1 in five tumor types, with the exception of UCEC (Figure S17a). Notably, we identified nuclear factor of activated T cells 1 and 2 (NFATC1 and NFATC2) in the intracellular signaling network triggered by LGALS1-PTPRC interaction (Figure S18a), which are considered as key transcription factors (TFs) leading to CD8+T cells exhaustion [39, 40]. To investigate the association between NFATC1/2 and CD8+T cells exhaustion, we analyzed CD8+T cells from pan-cancer scRNA-seq data. CD8+T cells were re-clustering into thirteen subpopulations, and Slingshot analysis identified six distinct lineages (Figure S18b and c). The C6 cluster, marked by high expression of naive markers including CCR7 and TCF7, was deemed as the trajectory starting point (Figure S18d). The C8 cluster was characterized by upregulation of T cell exhaustion markers (HAVCR2, TIGIT, LAG3, PDCD1, CXCL13, and LAYN) and identified as exhausted CD8+T (Tex) cells (Figure S18d). Confirming our expectations, GeneSwitches analysis revealed the activation of NFATC2 following immune checkpoint gene induction in the T cell exhaustion trajectory (lineage 1) (Figure S18e). Moreover, our analysis of the TCGA datasets also revealed a significant positive correlation between NFATC2 and PDCD1 across six distinct tumor types (Figure S17b). Collectively, these findings expand our understanding of the role of iCAFs in mediating CD8+T cells exhaustion.

Anti-PD1 treatment influences the communication between iCAFs and TME

To investigate the effect of anti-PD1 therapy on iCAFs, we analyzed publicly available scRNA-seq data from 31 paired pre- and on-treatment samples of BRCA patients receiving pembrolizumab [41]. Interestingly, we obtained CAF subtyping results consistent with those in pan-cancer after further subclustering of fibroblasts (Fig. 5a and b). We then stratified the samples based on T-cell clonal expansion and treatment time point and compared the changes in cell proportions. Due to the absence

of clonal expansion information for two patients, they were excluded from this analysis. During treatment, patients with clonal expansion had lower proportions of cancer cells compared to those without, potentially due to an increase number of T cells with cytotoxic activity (Fig. 5e). Moreover, the proportion of iCAFs was consistently lower in patients with clonal expansion compared to those without, both pre- and on-treatment (Fig. 5c-e). Notably, for both clonal expansion and non-clonal expansion patients, the proportion of iCAFs did not change during treatment compared to pre-treatment (Figure S19a and b).

While the proportion of iCAFs remained unchanged after anti-PD-1 treatment, it is possible that their transcriptional profiles underwent changes. To explore this possibility, we conducted differential expression analysis. Interestingly, we found that Chitinase-3-Like-1 (CHI3L1) was significantly upregulated (Figure S19c), which is a known regulator promoting M2 macrophage polarization [42]. Consistently, AUCell analysis revealed an enhanced ability of iCAFs to promote M2 macrophage polarization during the treatment compared to before (Figure S19e). It is noteworthy that the differentially expressed genes (DEGs) of iCAFs between pre-treatment and on-treatment were enriched in the TNF α signaling via NF- κ B, epithelial cell proliferation, and EMT both before and during the treatment (Figure S19d). Similarly, the AUCell scores of epithelial cell proliferation and EMT in iCAFs were significantly increased during the treatment (Figure S19e). Additionally, the AUCell analysis results showed that anti-PD1 treatment also enhanced the complement activation feature of iCAFs (Figure S19e). Overall, these findings suggest that anti-PD1 treatment influences the communication between iCAFs and other cells, including promoting epithelial cell proliferation, EMT, and M2 macrophage polarization.

We next sought to determine the differences in communication between iCAFs and immune cells before and during anti-PD-1 therapy. We further categorized myeloid cells into monocytes, macrophages, LAMP3+ dendritic cells (LAMP3+DCs), classical type 1 dendritic cells (cDC1s), and classical type 2 dendritic cells (cDC2s), and T cells into CD4+T cells, CD8+T cells, and Tregs, and

(See figure on next page.)

Fig. 4 Effect of CAFs on TME through paracrine signaling. **a** GO enrichment of ligands from mCAFs to endothelial cells. **b** GO enrichment of ligands from iCAFs to macrophages. **c** GO enrichment of ligands from iCAFs to CD8+T cells. **d** Integrated ranking of LRIs based on number of LRIs from mCAFs to endothelial cells using RRA algorithm across 22 tissue slices. **e** Integrated ranking of LRIs based on number of LRIs from iCAFs to macrophages using RRA algorithm across 22 tissue slices. **f** Integrated ranking of LRIs based on number of LRIs from iCAFs to CD8+T cells using RRA algorithm across 22 tissue slices. **g** Spatial distribution of the LGALS1-PTPRC interaction on two spatial transcriptomics tissue slices (BRCA0 and LIHC1). **h** Representative immunofluorescence images of CFD (red) and CD8 (green) in tissues from three patients with BRCA and three patients with LIHC. Scale bar represents 20 μ m

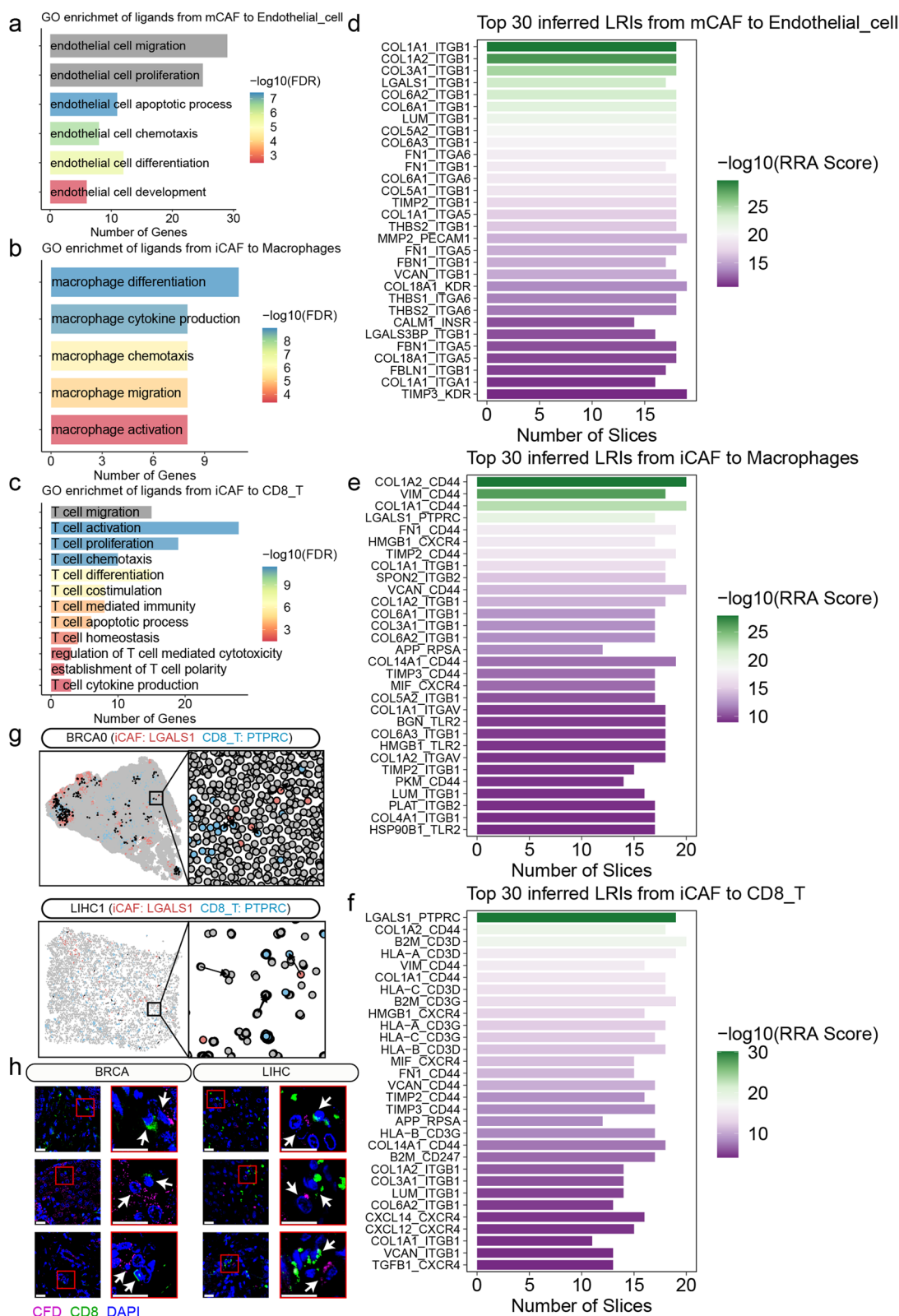


Fig. 4 (See legend on previous page.)

conducted CellChat analysis between iCAFs and immune cells (Figure S19f-i). Notably, we found anti-PD1 treatment enhanced iCAFs' ability to promote the formation of an immunosuppressive microenvironment. Compared to pre-treatment, iCAFs secreted MIF and laminins during treatment to suppress the activation, proliferation, and migration of CD8+ T cells [36, 43–45] (Fig. 5f and g). Although iCAFs downregulated TGFB3 during treatment compared to pre-treatment, they may still promote monocyte survival and differentiation into tumor-associated macrophages by overexpressing macrophage colony-stimulating factor-1 (CSF-1) [46–48] (Fig. 5g and h). Through the CXCL12/CXCR4 axis, iCAFs may reduce CD8+ T cell infiltration, promote CD8+ T cell dysfunction, and increase the number of Tregs [49] (Fig. 5g).

iCAF score correlate with immunotherapy response

Given the complex interplay between iCAFs and immune cells in TME, we hypothesized that the gene expression features of iCAFs are associated with immune checkpoint blockade (ICB) response. Using ssGSEA algorithm, we constructed an iCAF score using the top ten marker genes of iCAFs (Table S3) and applied it to different melanoma immunotherapy cohorts. In all cohorts, patients with high iCAF scores displayed prolonged overall survival (OS) (Log-rank test $P < 0.0001$ for the Gide anti-PD-1 cohort; $P = 0.00026$ for the Gide anti-CTLA-4 cohort and $P = 0.001$ for the Nathanson cohort; Fig. 6a) and progression-free survival (PFS) (Log-rank test $P < 0.0001$ for the Gide anti-PD-1 cohort and $P < 0.0001$ for the Gide anti-CTLA-4 cohort; Fig. 6a). Next, we divided the melanoma patients into high and low iCAF score groups based on the median and compared the percentage of ICB responders between the two groups. The results showed that patients with high iCAF scores had higher percentages of responders to anti-PD-1 treatment (Fisher's exact test $P = 0.0169$; Fig. 6b) and anti-CTLA-4 treatment (Fisher's exact test $P = 0.04146$; Fig. 6c). Moreover, consistent with these findings, both anti-PD-1 and anti-CTLA-4 responders had higher iCAF scores than non-responders (Figure S20a). These findings indicate that the iCAF score is a valuable tool in predicting patient survival and response to ICB therapy.

Tumor mutational burden (TMB) serves as a widely recognized biomarker for immunotherapy and is generally associated with patients' response to ICB [50, 51]. Therefore, we conducted a comprehensive analysis using data from TCGA database focusing on melanoma patients. Surprisingly, we found that melanoma patients with a high iCAF score exhibited significantly lower TMB compared to those with a low iCAF score (Figure S20b and c), suggesting the presence of additional mechanisms driving anti-tumor immune responses in high iCAF score melanoma patients. A previous study reported that TMB is not associated with the response to immunotherapy in melanoma patients [52]. Notably, apart from HYDIN and ADGRV1 mutations being more frequent in low iCAF score melanoma patients, there were no significant differences in the prevalence of other common mutations between high and low iCAF score groups (Figure S20b). Interestingly, we observed that the burden of CNVs at the arm level showed no significant difference between high and low iCAF score patients (Figure S20d). However, when examining CNVs at the focal level, we found that high iCAF score patients exhibited a lower burden of gain and loss of CNVs (Figure S20d). This pattern closely resembles the immune-rich tumor phenotype previously reported in LIHC [53] and CRC [54]. Based on these intriguing findings, we focused our analysis on the immune landscape of melanoma patients and found several key features associated with high iCAF score melanoma patients. Specifically, we observed higher expression of immune checkpoint genes (PDCD1, CTLA4, and LAG3) and a higher frequency of CNV amplifications in these patients (Fig. 6d and S20e). Notably, the majority of immune modulators showed elevated expression in high iCAF score patients (Fig. 6d and S20e), implying the presence of more complex interactions within the TME of these patients. The immune scores calculated by the ESTIMATE algorithm and previously reported immune response scores, including immune score (Roh) [55], cytolytic activity (CYT) [56], tertiary lymphoid structures signature (TLS) [57], IFN γ [58], expanded immune [58], and T cell inflamed [58] (Table S11), were also found to be higher in patients with high iCAF scores (Fig. 6e). Furthermore, immune cells and multiple inflammatory

(See figure on next page.)

Fig. 5 Anti-PD1 treatment influences the communication between iCAFs and TME. **a** UMAP plot showing the fibroblasts subpopulations in BRCA immunotherapy cohort. **b** Heatmap showing the expression of marker genes in fibroblast subpopulations. **c** UMAP plots showing the temporal alterations of fibroblasts subpopulations. **d** Boxplot showing the differences in cell proportions between patients with and without clonal expansion before anti-PD1 treatment. Statistical analysis was performed using unpaired t-tests; * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. **e** Boxplot showing the differences in cell proportions between patients with and without clonal expansion on anti-PD1 treatment. Statistical analysis was performed using unpaired t-tests; * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. **f** Differential cell–cell interaction signaling pathway alterations in iCAFs during anti-PD-1 treatment compared to pre-treatment. **g** Upregulated LRIs in iCAFs during anti-PD-1 treatment compared to pre-treatment. **h** Downregulated LRIs in iCAFs during anti-PD-1 treatment compared to pre-treatment

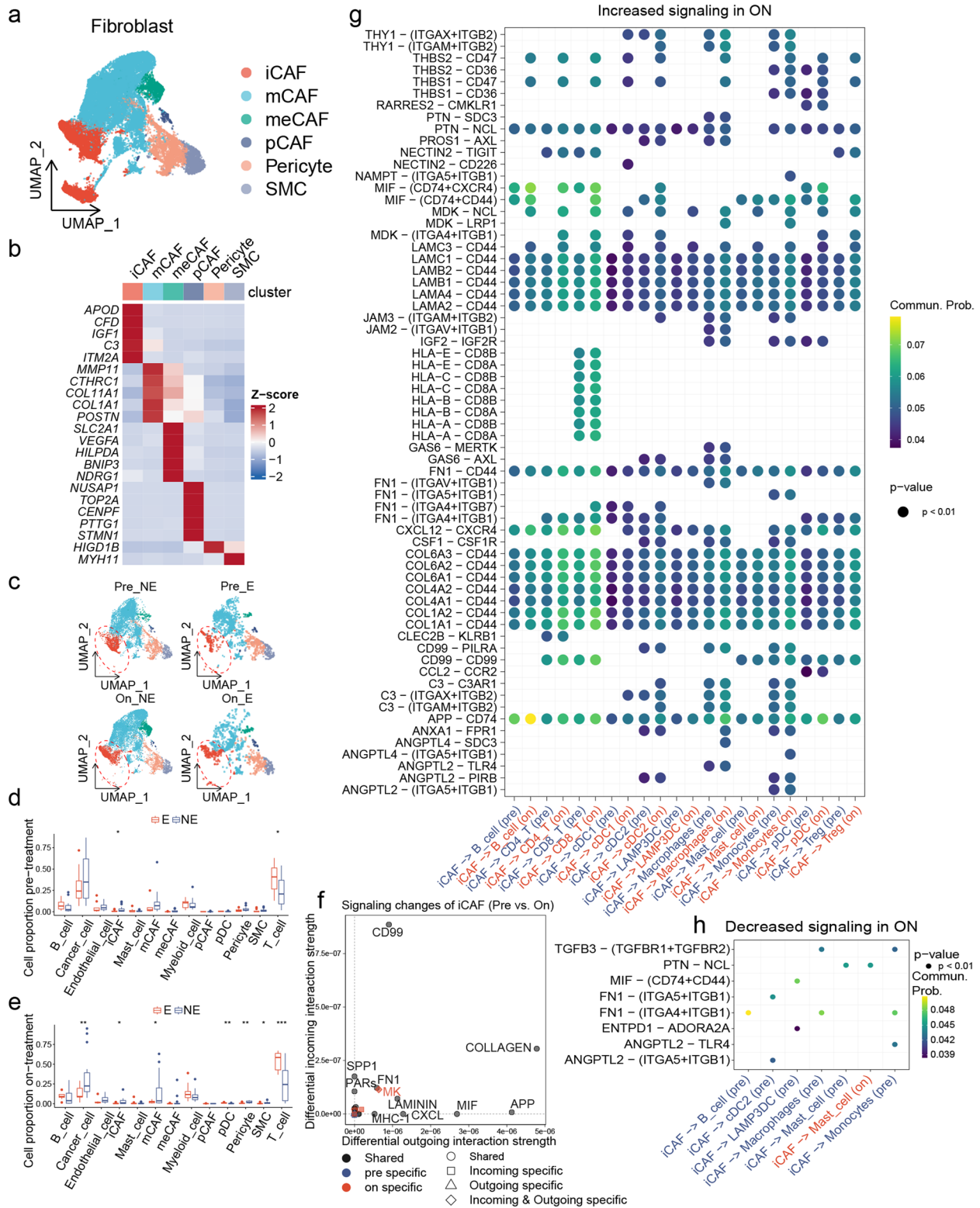


Fig. 5 (See legend on previous page.)

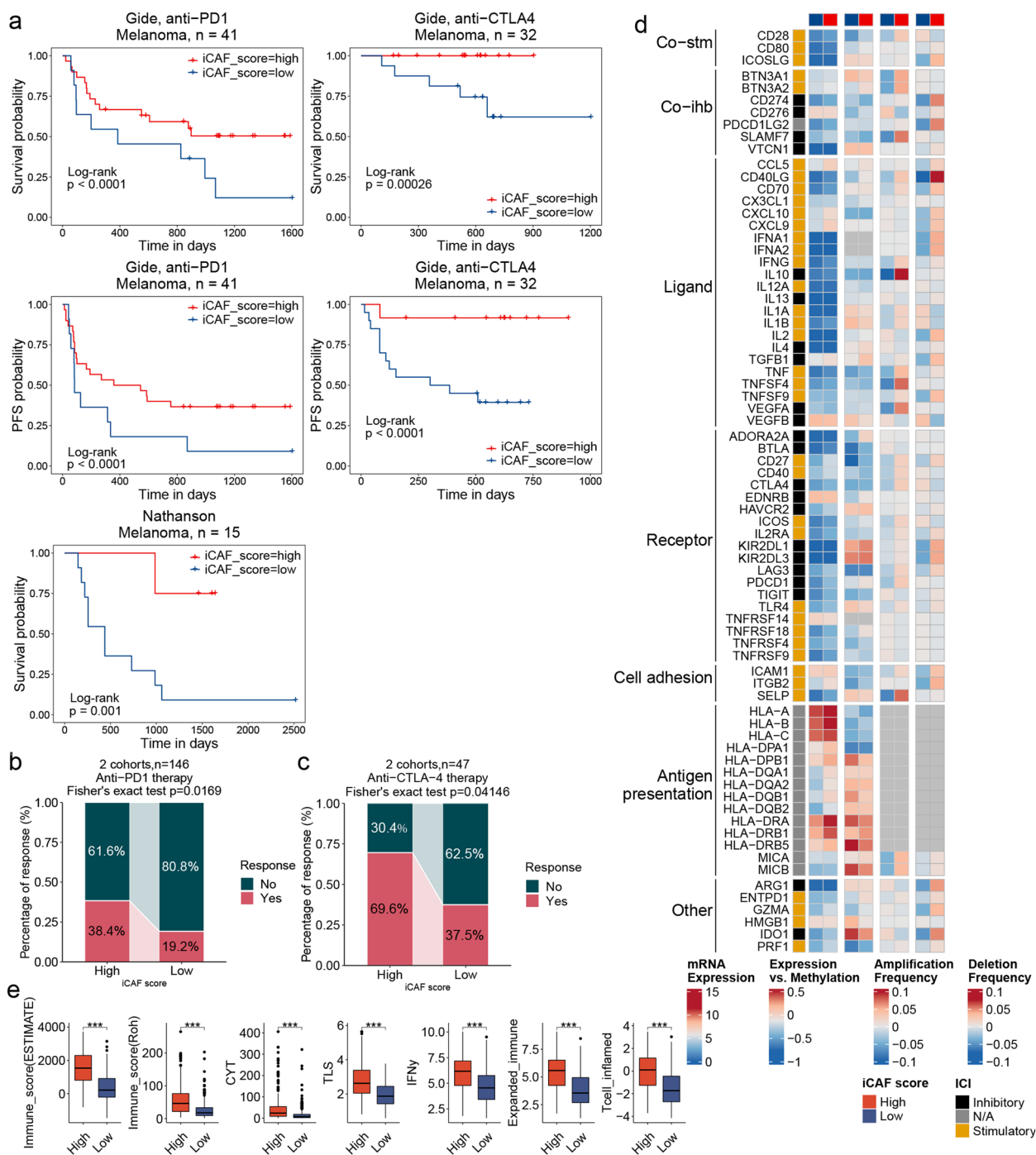


Fig. 6 iCAF score correlate with immunotherapy response. **a** Kaplan–Meier plots showing the prognostic value of iCAF score in the melanoma immunotherapy cohorts. *P*-values were calculated by log-rank test. **b** Percentage of anti – PD1 therapy response among melanoma patients with high and low iCAF scores. Statistical analysis was performed using Fisher's exact test. **c** Percentage of anti – CTLA – 4 therapy response among melanoma patients with high and low iCAF scores. Statistical analysis was performed using Fisher's exact test. **d** Heatmap showing immune modulators in melanoma patients with high and low iCAF scores. From left to right: mRNA expression (median-normalized expression levels of immune modulators); expression versus methylation (Spearman correlation between expression of immune modulators and DNA methylation beta-values); amplification frequency (difference in the proportion of immune modulators amplifications between patients with high or low iCAF scores and the proportion of immune modulators amplifications in all patients.); and the deletion frequency (as amplifications). **e** Boxplot showing the comparison of immune related scores in melanoma patients with high and low iCAF scores. Statistical analysis was performed using Wilcoxon rank-sum tests; **P* < 0.05, ***P* < 0.01, ****P* < 0.001

pathways (JAK–STAT, NF κ B, and *TNFA*) were enriched in high iCAF score patients (Figure S20f and g). These data suggest that the benefit of high iCAF score patients in tumor immunotherapy may rely on increased immune cell infiltration and the intricate interplay of immune modulators.

Discussion

We have collected scRNA-seq data and ST data from patients with six prevalent cancer types to conduct a comprehensive study on the biological characteristics of CAFs in TME. While the proportions of distinct CAF subtypes varied among different cancer types, both iCAFs and mCAFs consistently emerged as the primary subtypes across all common cancer types. Interestingly, our observations extended to two rare cancer types, EMC and MEC, where the presence of iCAFs and mCAFs was also found.

It's worth noting that we also made an intriguing discovery regarding pCAFs, which exhibited heightened activity in IFN-I production. The role of IFN-I (Type I interferon) in cancer presents a dual-edged sword effect [59–63]. Acute exposure to high concentrations of IFN-I can lead to the growth arrest and apoptosis of cancer cells, whereas prolonged exposure to low concentrations of IFN-I may promote the survival of cancer cells [60]. Additionally, IFN-I plays a critical role in facilitating cDC1 cross-priming and CD8+T cell reactivation [61, 63]. Therefore, a promising treatment strategy could involve combination immunotherapy targeting pCAF.

As widely recognized, metabolic reprogramming serves as a crucial hallmark of cancer cells, facilitating the establishment of a tumor-promoting microenvironment [64]. Recent studies have shed light on the impact of CAFs on cancer cell metabolism through their intrinsic metabolic reprogramming [24]. However, these studies often overlook the heterogeneity of CAFs, merely revealing the average metabolic characteristics across all subtypes. Employing a comprehensive pan-cancer single-cell analysis, we found diverse metabolic reprogramming mechanisms within distinct CAF subpopulations. Specifically, mCAFs exhibited enrichment in fatty acid biosynthesis, pCAFs displayed enrichment in the TCA cycle, while meCAFs demonstrated metabolic enrichment in glycolysis, alanine, aspartate, and glutamate metabolism. Consequently, the development of therapeutic strategies targeting the metabolic reprogramming of CAFs should consider the distinct characteristics exhibited by various subtypes of CAFs.

Pericytes are vital mural cells that can undergo pericyte–fibroblast transition (PFT) under the influence of

changes in matrix stiffness and tumor-secreted factors [65, 66]. This phenotypic transition plays a crucial role in promoting tumor growth and metastasis [26]. In this study, we have uncovered a pericyte-iCAF transition pathway, suggesting that the initial fibroblasts derived from pericytes may be iCAFs. Along the transition pathway from pericytes to iCAFs, expression of genes associated with inflammatory response and ECM was significantly upregulated, indicating a potential involvement of PFT in facilitating the formation of an immunosuppressive microenvironment and ECM remodeling. Further exploration is warranted to unravel the functional roles and underlying mechanisms of PFT in this context.

In the spatial analysis, the four subpopulations of fibroblasts demonstrated a relatively closer spatial proximity compared to other cell types. We found an enrichment of endothelial cells in close proximity to mCAFs, and their intercellular communication was found to promote angiogenesis within TME. While CD8+T cells were not found to be enriched in close proximity to iCAFs, we observed numerous instances of iCAFs-CD8+T cells co-localization in situ across various tumor types, which was further confirmed by immunofluorescence. The complex interplay between iCAFs and immune cells, particularly macrophages and CD8+T cells, facilitates the establishment of an immunosuppressive microenvironment. Notably, analyzing scRNA-seq data from BRCA patients receiving anti-PD-1 immunotherapy, we have identified that anti-PD-1 immunotherapy enhances the capacity of iCAFs to promote the establishment of an immunosuppressive microenvironment. Additionally, there was a significant correlation observed between the iCAF score constructed based on iCAF marker genes and the immune therapy response in melanoma patients. Therefore, the combination of targeted interventions against iCAFs with anti-PD-1 treatment holds promising potential as a valuable therapeutic approach.

Conclusion

In conclusion, our comprehensive analysis of pan-cancer spatial and single-cell data has unraveled the heterogeneity of CAFs, shedding light on their spatial distribution patterns and intricate cell communication with TME. To facilitate further exploration of CAF heterogeneity, we have developed an interactive website (<https://chenxisd.shinyapps.io/pancaf/>) using the ShinyCell R package [67]. Our pan-cancer study not only enhances our understanding of CAF biological characteristics but also provides important insights for the development of targeted therapeutic approaches aimed at CAFs.

Methods

Sample acquisition and processing

The study was approved by the Ethics Committee of Qilu Hospital of Shandong University (KYL-2017-256) and conducted in accordance with the Declaration of Helsinki. All subjects gave written informed consent before participating in the study. One sample of MEC, one sample of EMC, one sample of HNSCC and two non-malignant samples (control group, one adjacent normal tissue from MEC and one adjacent normal tissue from HNSCC), were obtained from the Qilu Hospital of Shandong University, Jinan, China. The clinical information for these samples is provided in Table S6. All samples were processed immediately after being obtained from oral cancer surgery according to the standard procedures. According to the manufacturer's instructions, the Human Tumor Dissociation Kit (Miltenyi Biotec; Order no: 130-095-929) was used to obtain single cells from the tissues.

Single-cell RNA sequencing

According to the manufacturer's protocol, Chromium Single cell 3' Reagent v3 kits were used to prepare barcoded scRNA-seq libraries. The cell concentration was adjusted to 700–1200 cells/ μ L. The gel beads, carrying barcode information, were combined with a mixture of cells and enzymes, and subsequently enveloped by oil droplets, forming gel beads in emulsions (GEMs). The gel beads within GEMs underwent dissolution, releasing mRNA upon cell lysis. Reverse transcription was then performed to generate barcoded cDNA for sequencing. After disrupting the liquid oil layer, cDNA amplification was carried out, followed by purification and quality inspection. Subsequently, the cDNA was digested into fragments of approximately 200–300 bp, and then subjected to the traditional second-generation sequencing library construction process, which included the addition of sequencing adapter P5 and sequencing primer R1, followed by PCR amplification to obtain the DNA library. Finally, the constructed library was subjected to high-throughput sequencing using the Illumina NovaSeq 6000 platform in a paired-end sequencing mode.

scRNA-seq data and ST data processing

The newly generated raw scRNA-seq data were processed by Cell Ranger (v 3.1.0) to generate a UMI count matrix. The human genome (hg38) was used as a reference. Raw gene expression matrices were constructed into a Seurat object and imported into R software by Seurat R package [68]. Low-quality cells ($>40,000$ UMI/cell, <500 genes/cell, $>5,000$ genes/cell and $>20\%$ mitochondrial genes) were excluded. Doublets were identified and removed by DoubletFinder R package [69]. The harmony R package

[70] was utilized for batch effect correction. We utilized the local inverse Simpson's Index (LISI) to evaluate batch effects [70]. We performed principal component analysis (PCA) to reduce the dimensionality of scRNA-seq data. Top 30 principal components (PCs) were selected for UMAP. The FindClusters function was used to identify cell clusters.

We applied the same processing pipeline to publicly available scRNA-seq datasets from the 10 \times Genomics platform. For the public scRNA-seq datasets sourced from the inDrop platform, we performed quality control by filtering out cells with UMI counts greater than 40,000, cells with gene counts less than 200, cells with gene counts exceeding 5000, and cells with mitochondrial gene count surpassing 30%.

We imported the publicly available ST dataset into Seurat using the Load10X_Spatial function. Subsequently, we filtered out low-quality spots with gene counts below 500 and mitochondrial gene count exceeding 30%.

Recognition of malignant and non-malignant epithelial cells

The copykat R package [71] was used to identify malignant and non-malignant epithelial cells with default parameters. The cells from TME were used as a normal reference.

CellTrek analysis

To acquire the spatial coordinates of the cells, we conducted a combined analysis of the scRNA-seq data and ST data using the CellTrek R package [14] with its default parameters.

We utilized the run_kdist function from the CellTrek package to calculate the spatial k-distance between different cell types. The analysis followed the parameters: ref_type = "all", keep_nn = F, k = 10.

Comparison dendrograms

To conduct a phylogenetic analysis of the different cell subpopulations within the pan-cancer scRNA-seq dataset, we utilized the BuildClusterTree function from the Seurat R package. To visualize the results, the ggtree R package [72] was applied.

Differential expression analysis and functional enrichment analysis

To identify DEGs in the scRNA-seq data, we utilized the "FindAllMarkers" or "FindMarkers" functions in Seurat. The thresholds were set as $|\log_2FC| > 1$ and $\text{adj.p.val} < 0.05$. Subsequently, we conducted functional enrichment analysis of the DEGs using the WebGestaltR R package [73]. For this analysis, we selected the "genome protein-coding" as the reference set.

Cancer preferences analysis

To assess the cancer preferences of CAF subtypes, odds ratios (OR) were calculated using the computational method described by Zhang et al. [74]. This involved constructing a 2 by 2 contingency table for each combination of CAF subtypes *i* and cancer types *j*. The table included the number of cells from CAF subtypes *i* in cancer types *j*, the number of cells from CAF subtypes *i* in other cancer types, the number of cells from non-*i* CAF subtypes in cancer types *j*, and the number of cells from non-*i* CAF subtypes in other cancer types. Fisher's exact test was then performed on the contingency table.

SCENIC analysis

To calculate the regulon activity scores (RAS) of CAFs, we used the pySCENIC Python package [75] for SCENIC analysis. First, GRNBoost2 was used to infer the co-expression modules between TFs and candidate target genes. Then, RcisTarget was used to analyze the genes in each co-expression module to identify the enriched motifs (a TF and its potential direct target genes were defined as a regulon). Finally, AUCell was used to evaluate the activity of each regulon in each cell.

We measured the cell-type specificity of a regulon by calculating the regulon specificity score (RSS) using the computational method described by Suo et al. [76]. First, we defined a probability distribution of RAS $P^R = (P_1^R, \dots, P_n^R)$ and normalized it so that $\sum_{i=1}^n P_i^R = 1$. Second, we defined a probability distribution of cell types $P^C = (P_1^C, \dots, P_n^C)$ to indicate whether a cell belongs to a specific cell-type ($P_i^C = 1$) or not ($P_i^C = 0$) and normalized it so that $\sum_{i=1}^n P_i^C = 1$. Then, we calculated the Jensen-Shannon Divergence (JSD) $JSD(P^R, P^C) = H\left(\frac{P^R + P^C}{2}\right) - \frac{H(P^R) + H(P^C)}{2}$ to measure the difference between the two probability distributions. Finally, RSS was calculated as: $RSS(R, C) = 1 - \sqrt{JSD(P^R, P^C)}$.

Gene set scoring

To score gene sets in the scRNA-seq data, we utilized the "AUCell" method from the irGSEA package. The gene set files for GO Biological Processes (GOBP), HALLMARK, and REACTOME were obtained from The Molecular Signatures Database (MSigDB) (<https://www.gsea-msigdb.org/gsea/msigdb>) using the msigdbR package (Table S5). The signature genes of M2 macrophage polarization were derived from the supplementary materials of a previously published study by Azizi et al. [77] (Table S5).

Single-cell metabolic activity analysis

To evaluate the metabolic pathway activity of meCAFs, we utilized the SCPA R package [78] to analyze meCAFs in the pan-cancer single-cell dataset, using the metabolic

pathway gene sets obtained from the supplementary materials of Bibby et al.'s study [78].

Furthermore, we employed the scMetabolism R package [79] with default parameters to quantify the metabolic activity of four distinct subtypes of CAFs.

Trajectory analysis

The slingshot R package [80] was used for inferring cell lineages and pseudotime. It utilized a clustering-based minimum spanning tree (MST) to identify the lineage structure and applies simultaneous principal curves to fit branch curves to these lineages. The getCurves function was employed to obtain smoothed trajectory curves.

Based on the inferred pseudotime, we utilized the GeneSwitches R package [81] to identify gene expression events within specific trajectory. The binarize_exp function was employed to convert the single-cell gene expression matrix into a binary state, using the following parameters: binarize_cutoff=0.05 and fix_cutoff=TRUE. Subsequently, a logistic regression model was fitted and the switching time was estimated using the find_switch_logistic_fastglm function. Genes that satisfied the criteria (zero_pct=0.9, r2cutoff=0.02) were selected as switch genes. To determine the switch pathways, genes were initially filtered based on the following parameters: zero_pct=0.9 and r2cutoff=0.1. Finally, the find_switch_pathway function (sig_FDR=0.05, pathways=msigdb_h_c2_c5) was employed, utilizing a hypergeometric test, to extract the switch pathways.

Transcriptional homogeneity analysis

In order to estimate the heterogeneity of different CAFs subpopulations, we performed transcriptional homogeneity analysis on CAFs in the pan-cancer scRNA-seq dataset, adopting the computational approach described by Marjanovic et al. [82]. Specifically, using the top 100 marker genes of each cluster found by the FindAllMarkers function of the Seurat R package, we discretized expression per gene into 10 bins. Then, we subsampled 100 cells for each tumor sample 100 times and calculated the median value of the pairwise normalized mutual information (NMI). NMI between each pair of cells *x* and *y* was calculated according to the following 3 steps: (1) mutual information (MI) $I(X; Y) = \sum_x \sum_y p(x, y) \log \frac{p(x, y)}{p(x)p(y)}$; (2) entropy of each cell $H(X) = \sum_x p(x) \log(p(x))$; (3) $NMI(X, Y) = \frac{I(X; Y)}{\sqrt{H(X)H(Y)}}$

Spatial trajectory analysis

To investigate the dynamic biological processes occurring between high-density regions of iCAFs and mCAFs within the spatial context, we utilized the SPATA2 R package. Firstly, we transformed the Seurat object into

a Spata object using the transformSeuratToSpata function. Then, employing the createTrajectories function, we generated a spatial trajectory starting from the high-density region of iCAFs and ending at the high-density region of mCAFs. The plotTrajectoryFeaturesDiscrete function was used to visualize the changes in cell proportions along the trajectory. Additionally, the plotTrajectoryGeneSets function was utilized to depict the variations in gene sets along the trajectory.

Cell–cell interaction analysis

To infer cell–cell interaction within the spatial context, we utilized the SpaTalk R package [83]. Firstly, we created a SpaTalk object using the createSpaTalk function. Subsequently, the dec_cci function was applied with default parameters to identify ligand–receptor pairs involved in the interaction between CAFs and TME. The ligand–receptor pairs for each tissue slice were ranked based on their occurrence frequency, and the results from all tissue slices were integrated using the RRA algorithm [27]. To visualize the inferred intracellular signaling pathways, we employed the plot_path2gene function.

We conducted comparative analysis of cell communication between iCAFs and immune cells in a cohort of BRCA patients receiving anti-PD-1 immunotherapy using the CellChat R package [84]. The netAnalysis_signalingChanges_scatter function and netVisual_bubble function were utilized to visualize the changes in signaling pathways and ligand–receptor pairs from pre-treatment to on-treatment in BRCA patients.

Melanoma immunotherapy dataset collection

We collected the expression matrix and clinical information of the GSE91061 dataset [85] (referred to as the Riaz cohort) from the Gene Expression Omnibus (GEO) database (<http://www.ncbi.nlm.nih.gov/geo/>). FPKM normalized gene expression data was converted into log₂ (TPM + 1) data. Furthermore, we obtained the expression matrix and clinical information of the Gide cohort [86] and Nathanson cohort [87] from the Tumor Immune Dysfunction and Exclusion (TIDE) database (<http://tide.dfci.harvard.edu/>) [88]. We performed batch effect correction using the "ComBat" function in the SVA R package.

TCGA RNA-seq data processing

RNA-seq data and clinical profiles of BRCA, CRC, LIHC, OVCA, PRAD, UCEC, and skin cutaneous melanoma (SKCM) from the TCGA database were downloaded from GDC API. Count data was converted into log₂ (TPM + 1) data.

Survival analysis

Kaplan–Meier survival analyses were performed with survival R package and survminer R package. The cut-off value of continuous variables in the survival data was determined by the surv_cutpoint function of survminer R package. *P*-values were calculated by log-rank test.

Immune score

The estimate R package [89] was utilized to calculate the Immune_score (estimate). The Immune_score (Roh) was determined as the geometric mean of gene expression levels of cytolytic markers, HLA molecules, IFN- γ pathway genes, chemokines, and adhesion molecules [55]. Cytolytic activity (CYT) was calculated as the geometric mean of GZMA and PRF1 [56]. Tertiary lymphoid structures (TLS) were determined based on the mean expression of TLS-signature genes [57]. IFN γ and expanded immune scores were obtained by averaging the gene expression levels of the included genes for the IFN- γ (6-gene) and expanded immune (18-gene) signatures, respectively [58]. Lastly, the Tcell inflamed score was calculated as the weighted sum of Tcell inflamed signature genes after housekeeping normalization [58]. The detailed information is provided in Table S11.

Mutation, CNV, and DNA methylation analysis

We utilized the TCGAbiolinks R package [90] to download somatic mutation data and CNV data from the TCGA database for melanoma patients. The maftools R package [91] was employed for analyzing and visualizing the somatic mutation data. The TMB was calculated using the tmb function. Fisher's exact test was conducted to identify mutation genes with differential frequencies between groups with high and low iCAF scores. For the CNV data, the GISTIC2.0 [92] analysis module available on the GenePattern (<https://cloud.genepattern.org>) [93] was used to detect significantly amplified or deleted genomic regions. The burden of copy number alterations was quantified by counting the total number of genes exhibiting copy number gains or losses at both the focal and arm levels. The DNA methylation data and CNV data for melanoma patients, obtained from the UCSC Xena database (<https://xenabrowser.net/datapages/>), were employed for the analysis of immune modulators.

Estimation of immune cell infiltration levels

We obtained gene signatures of 28 tumor-infiltrating lymphocytes (TILs) from the TISIDB database (<http://cis.hku.hk/TISIDB>) [94]. Subsequently, we employed the ssGSEA algorithm from the GSVA R package [95]

to estimate the immune cell enrichment scores for each tumor sample.

PROGENy analysis

The progeny R package [96] was utilized to infer the activity of 14 cancer-related pathways using default parameters.

Immunofluorescence staining

The tissue microarrays for BRCA and LIHC were purchased from Shanghai Qutdo Biotech Company (Shanghai, China). For the immunofluorescence staining aimed at validating the existence of CAF subtypes, we followed the protocol outlined below. Immunofluorescence staining was performed using the Quadruple-Fluorescence immunohistochemical mouse/rabbit kit (Immunoway) to detect the expression of specific markers. The microarray was placed on a slide warmer and baked at 60 °C for 60 min to ensure adhesion. A two-in-one dewaxing and antigen retrieval reagent was added to a retrieval box and heated to boiling. Subsequently, the microarray was immersed in the boiling dewaxing and antigen retrieval reagent, ensuring complete submersion of the tissue. They were heated at medium flame for 30 min. The retrieval box was then removed from the heat source and allowed to naturally cool to room temperature. Following this, the microarray was transferred to a beaker containing distilled water and rinsed 5–6 times. Excess moisture around the tissue was blotted using filter paper. The microarray was then incubated with peroxidase-blocking buffer at room temperature for 15 min, followed by washing with PBS three times for 2 min each. Next, primary antibodies, including CENPF (Rabbit, 1:200, Immunoway), HILPDA (Rabbit, 1:200, Bioss), MMP-11 (Rabbit, 1:200, Immunoway), and CFD (Rabbit, 1:200, Immunoway), were diluted and applied to the microarray, ensuring complete coverage. Incubation was carried out at 37 °C for 1–2 h (or overnight at 4 °C in a humid chamber), followed by three washes with PBS for 2 min each. After blotting the excess moisture, the microarray was incubated with an HRP-conjugated anti-rabbit/mouse IgG secondary antibody at room temperature for 30 min. The sections were washed again with PBS three times for 2 min each. For fluorescence labeling, Tyramide working solution was added and incubated for 10 min. Subsequently, the sections were washed with PBS three times for 2 min each. The microarray was placed in a retrieval box, and an antibody stripping solution was added. Microwave heating was performed at high power for 3 min and at medium–low power for 15 min. After natural cooling, the sections were washed with PBS three times for 2 min each. Finally, DAPI staining solution was added and mounting medium was applied to

cover the microarray, ensuring contact without trapping air bubbles. Subsequently, the sections were scanned and imaged using a digital slide scanner microscope.

For the immunofluorescence staining aimed at validating the spatial proximity of CFD-positive cells and CD8-positive cells, we followed the protocol outlined below. The microarray was placed on a slide warmer at 60 °C for 30 min. Subsequently, it was sequentially immersed in xylene (first and second), followed by various concentrations of ethanol and water, each for 5 min. Antigen retrieval was performed using trypsin at 37 °C for 20 min. The microarray was then washed three times with PBS buffer for 5 min each. Endogenous peroxidase activity was blocked by incubating the microarray with a peroxidase blocking agent at room temperature for 10 min, followed by three washes with PBS buffer. To block non-specific binding, goat serum was applied to the microarray at 37 °C for 15 min, after which the excess serum was removed. Primary antibodies (CD8 Ms 1:100, CFD Rb1:100) were added and left overnight at 4 °C. The microarray was then washed three times with PBS buffer, and secondary antibodies (Alexa Fluor 488@Ms, Alexa Fluor 594@Rb) were applied. The microarray was incubated at 37 °C for 30 min, and all subsequent steps were performed in a light-protected environment. After three washes with PBS buffer, DAPI staining solution was added, followed by three additional washes with PBS buffer. Finally, the microarray was mounted using a tissue mounting medium containing an anti-fading agent, ensuring the absence of bubbles. The imaging of the microarray slices was performed using a digital slide scanner microscope, capturing the desired observations.

Abbreviations

CAF	Cancer-associated fibroblast
TME	Tumor microenvironment
scRNA-seq	Single-cell RNA sequencing
EMC	Epithelial-myoepithelial carcinoma
MEC	Mucoepidermoid carcinoma
mCAF	Matrix cancer-associated fibroblast
iCAF	Inflammatory cancer-associated fibroblast
BRCA	Breast cancer
EMT	Epithelial-mesenchymal transition
ECM	Extracellular matrix
BC	Bladder carcinoma
HNSCC	Head and neck squamous cell carcinoma
PTC	Papillary thyroid carcinoma
LC	Lung cancer
ST	Spatial transcriptomics
CRC	Colorectal cancer
LIHC	Liver hepatocellular carcinoma
OVCA	Ovarian cancer
PRAD	Prostate adenocarcinoma
UCEC	Uterine corpus endometrial carcinoma
UMI	Unique molecular identifier
CNV	Copy number variation
SMC	Smooth muscle cell
meCAF	Metabolic cancer-associated fibroblast
PDAC	Pancreatic ductal adenocarcinoma

pCAF	Proliferative cancer-associated fibroblast
NSCLC	Non-small cell lung cancer
RRA	Robust rank aggregation
LRI	Ligand-receptor interaction
TCGA	The Cancer Genome Atlas
TF	Transcription factor
DEG	Differentially expressed gene
ICB	Immune checkpoint blockade
OS	Overall survival
PFS	Progression-free survival
TMB	Tumor mutational burden
CYT	Cytolytic activity
TLS	Tertiary lymphoid structures signature
PFT	Pericyte–fibroblast transition
GEM	Gel beads in emulsion
OR	Odds ratio
RAS	Regulon activity score
RSS	Regulon specificity score
MSigDB	Molecular Signatures Database
MST	Minimum spanning tree
GEO	Gene Expression Omnibus
SKCM	Skin cutaneous melanoma
LISI	Local inverse Simpson's Index

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12943-023-01876-x>.

Additional file 1: Table S1. Sample information of public single cell RNA sequencing data.

Additional file 2: Table S2. Sample information of public spatial transcriptomic data.

Additional file 3: Table S3. Marker genes for each cell type.

Additional file 4: Table S4. The top differentially expressed genes in each CAF subtype.

Additional file 5: Table S5. Gene signatures for various CAF functions.

Additional file 6: Table S6. Sample information of newly generated single cell RNA sequencing data.

Additional file 7: Table S7. Regulon specificity scores for each CAF subtype.

Additional file 8: Table S8. Metabolic pathway enrichment analysis results of mCAFs using SCPA R Package.

Additional file 9: Table S9. Integrated ranking of LRIs from each CAF subtype to TME using RRA algorithm.

Additional file 10: Table S10. Ligands promoting M2 macrophage polarization.

Additional file 11: Table S11. Immune scores.

Additional file 12: Figure S1. Quality control of scRNA-seq and ST data a Numeric table of quality-controlled scRNA-seq and ST data b Boxplot showing number of UMIs per cell from different samples in scRNA-seq data c Boxplot showing number of genes per cell from different samples in scRNA-seq data d Boxplot showing number of UMIs per cell from different sections in ST data e Boxplot showing number of genes per cell from different sections in ST data.

Additional file 13: Figure S2. UMAP plots showing the major cell types in each scRNA-seq dataset of this pan-cancer analysis.

Additional file 14: Figure S3. UMAP plots showing the CopyKAT classification results in each scRNA-seq dataset of this pan-cancer analysis.

Additional file 15: Figure S4. UMAP plots showing the subsets of myeloid cells in each scRNA-seq dataset of this pan-cancer analysis.

Additional file 16: Figure S5. UMAP plots showing the subsets of T&NK in each scRNA-seq dataset of this pan-cancer analysis.

Additional file 17: Figure S6. UMAP plots showing the co-embedding results of scRNA-seq and ST data using CellTrek.

Additional file 18: Figure S7. Transcriptional heterogeneity among CAF subtypes. a Phylogenetic tree of major cell types across six cancer types. b Boxplot showing transcriptional homogeneity of different CAF subtypes quantified by NMI. Statistical analysis was performed using Wilcoxon rank-sum tests; * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

Additional file 19: Figure S8. Boxplot showing the local inverse Simpson's Index (LISI) of fibroblasts before and after batch correction. Statistical analysis was performed using Wilcoxon rank-sum tests; * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

Additional file 20: Figure S9. Representative immunofluorescence images of CFD (deep yellow, iCAF), MMP11 (light yellow, mCAF), HILPDA (red, meCAF) and CENPF (green, pCAF) in tissues from patients with BRCA. Scale bar represents 20 μm .

Additional file 21: Figure S10. CAF heterogeneity in other cancer types. a UMAP plots showing the subsets of CAFs in NSCLC. b UMAP plots showing the subsets of CAFs in Melanoma. c UMAP plots showing the subsets of CAFs in HNSCC. d UMAP plots showing the subsets of CAFs in EMC. e UMAP plots showing the subsets of CAFs in MEC. f Volcano plot showing genes upregulated in fibroblasts derived from tumor tissues of HNSCC patients, compared to fibroblasts derived from adjacent non-tumor tissues. g Enriched GO functions of upregulated genes in fibroblasts derived from tumor tissues of HNSCC patients. h Volcano plot showing genes upregulated in fibroblasts derived from tumor tissues of MEC patients, compared to fibroblasts derived from adjacent non-tumor tissues. i Enriched GO functions of upregulated genes in fibroblasts derived from tumor tissues of MEC patients.

Additional file 22: Figure S11. Bubble heatmap showing metabolic pathway activities scored by scMetabolism in each CAF subtype.

Additional file 23: Figure S12. Spatial trajectory analysis of CAFs. a Spatial trajectory from high-density areas of iCAFs to high-density areas of mCAFs in OVCA1. B Changes in cell proportions of CAF subtypes along the trajectory direction in OVCA1. C Changes in pathway activity of CAF subtypes along the trajectory direction in OVCA1. D Spatial trajectory from high-density areas of iCAFs to high-density areas of mCAFs in CRC1. E Changes in cell proportions of CAF subtypes along the trajectory direction in CRC1. F Changes in pathway activity of CAF subtypes along the trajectory direction in CRC1.

Additional file 24: Figure S13. Integrated ranking of various functional LRIs based on number of LRIs from mCAFs to endothelial cells using RRA algorithm across 22 tissue slices.

Additional file 25: Figure S14. Integrated ranking of various functional LRIs based on number of LRIs from iCAFs to macrophages using RRA algorithm across 22 tissue slices.

Additional file 26: Figure S15. Integrated ranking of various functional LRIs based on number of LRIs from iCAFs to CD8+ T cells using RRA algorithm across 22 tissue slices.

Additional file 27: Figure S16. Effect of iCAFs on immune cells through paracrine signaling. A GO enrichment of ligands from iCAFs to B cells. b GO enrichment of ligands from iCAFs to dendritic cells. c GO enrichment of ligands from iCAFs to mast cells. d GO enrichment of ligands from iCAFs to neutrophils. E GO enrichment of ligands from iCAFs to NK cells. f GO enrichment of ligands from iCAFs to Tregs. G Integrated ranking of LRIs based on number of LRIs from iCAFs to B cells using RRA algorithm across 22 tissue slices. h Integrated ranking of LRIs based on number of LRIs from iCAFs to dendritic cells using RRA algorithm across 22 tissue slices. i Integrated ranking of LRIs based on number of LRIs from iCAFs to mast cells using RRA algorithm across 22 tissue slices. j Integrated ranking of LRIs based on number of LRIs from iCAFs to neutrophils using RRA algorithm across 22 tissue slices. k Integrated ranking of LRIs based on number of LRIs from iCAFs to NK cells using RRA algorithm across 22 tissue slices. l Integrated ranking of LRIs based on number of LRIs from iCAFs to Tregs using RRA algorithm across 22 tissue slices.

Additional file 28: Figure S17. Spearman correlation analysis of LGALS1 and NFATC2 with PDCD1. A Spearman correlation analysis of LGALS1 with PDCD1. B Spearman correlation analysis of NFATC2 with PDCD1.

Additional file 29: Figure S18. Downstream analysis of the interaction between iCAFs and CD8+ T cells. a Intracellular signaling network triggered by LGALS1- PTPRC interaction. B Slingshot trajectory analysis of the 13 clusters of CD8+ T cells. c Pseudotime of the 6 lineages of CD8+ T cells calculated by Slingshot. D Bubble heatmap showing the expression of marker genes for CD8+ T cell subtypes in pan-cancer. E GeneSwitches analysis of switching genes in Lineage 1.

Additional file 30: Figure S19. Anti-PD1 treatment influences transcriptional characteristics of iCAFs. A Boxplot showing the differences in cell proportions between patients with clonal expansion before and during anti-PD-1 treatment. Statistical analysis was performed using unpaired t-tests; * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. b Boxplot showing the differences in cell proportions between patients without clonal expansion before and during anti-PD-1 treatment. Statistical analysis was performed using unpaired t-tests; * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. c Volcano plot showing DEGs of iCAFs between pre- and on-anti-PD-1 treatment. d GO enrichment of DEGs of iCAFs between pre- and on-anti-PD-1 treatment. e Density heatmaps showing the comparison of pathway activities of iCAFs scored by AUCell between pre- and on-anti-PD-1 treatment. Statistical analysis was performed using Wilcoxon rank-sum tests; * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. f UMAP plot showing the myeloid cells subpopulations in BRCA immunotherapy cohort. g Bubble heatmap showing the expression of marker genes for myeloid cells subpopulations in BRCA immunotherapy cohort. h UMAP plot showing the T cells subpopulations in BRCA immunotherapy cohort. i Bubble heatmap showing the expression of marker genes for T cells subpopulations in BRCA immunotherapy cohort.

Additional file 31: Figure S20. The immune and genomic landscape of melanoma patients based on high and low iCAF scores. a Boxplot showing the comparison of iCAF scores in anti-PD1 therapy responders and non-responders, as well as anti-CTLA-4 therapy responders and non-responders, among melanoma patients. Statistical analysis was performed using Wilcoxon rank-sum tests; * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. b Mutation landscapes of melanoma patients with high and low iCAF scores. Statistical analysis was performed using Fisher's exact test; * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. c Boxplot showing the comparison of TMB in melanoma patients with high and low iCAF score. Statistical analysis was performed using Wilcoxon rank-sum tests; * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. d Boxplot showing the comparison of copy number gain burden or copy number loss burden at arm-level or focal-level in melanoma patients with high and low iCAF score. Statistical analysis was performed using Wilcoxon rank-sum tests; * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. e Boxplot showing the comparison of immune modulators expression levels in melanoma patients with high and low iCAF score. Statistical analysis was performed using Wilcoxon rank-sum tests; * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. f Boxplot showing the comparison of pathway activities scored by PROGENy in melanoma patients with high and low iCAF score. Statistical analysis was performed using Wilcoxon rank-sum tests; * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. g Heatmap showing immune cell infiltration levels in melanoma patients based on high and low iCAF scores.

Acknowledgements

Not applicable.

Authors' contributions

Q.F., C.X.M., and C.Z.Y. designed experiments. C.X.M. and Y.T.S. contributed to bioinformatics analysis. C.Z.Y., A.P., X.L.J., J.M., and L.W. were engaged in the clinical sample collection and immunofluorescence experiments. C.X.M. constructed the website. C.X.M., Q.F., and S.S. wrote and revised the manuscript.

Funding

We sincerely thank all colleagues in our laboratory for all of the kindly advices and support. We sincerely thank the foundation support of the National Natural Science Foundation of China (No. 82071122, 82270980), the National Young Scientist Support Foundation (2019), Excellent Young Scientist Foundation of Shandong Province (No. ZR202102230369), Taishan Young Scientist

Project of Shandong Province (2019), Periodontitis innovation team of Jinan City (2021GXRC021), Major Innovation Projects in Shandong Province (No. 2021SFGC0502), Oral Microbiome Innovation Team of Shandong Province (No. 2020KJK001), Shandong Province Key Research and Development Program (No. 2021ZDSYS18), Shandong Province Major Scientific and Technical Innovation Project (No. 2021SFGC0502).

Availability of data and materials

The raw sequence data reported in this paper have been deposited in the Genome Sequence Archive in National Genomics Data Center, China National Center for Bioinformation / Beijing Institute of Genomics, Chinese Academy of Sciences (GSA-Human: HRA004998) that are publicly accessible at <https://ngdc.cncb.ac.cn/gsa-human>. The integrated ST data from 22 tissue slices and their corresponding scRNA-seq data have been deposited in Synapse under the accession code syn51758773 (<https://www.synapse.org/#Synapse:syn51758773/>). The analysis and visualization of CAFs in pan-cancer can be performed at <https://chenxisd.shinyapps.io/pancaf/>. Reanalyzed publicly available scRNA-seq data can be accessed from the GEO database under accession codes: GSE176078 [97], GSE166555 [98], GSE149614 [99], GSE184880 [100], GSE203612 [101], GSE207422 [18], GSE215120 [19], GSE181919 [7]. The scRNA-seq data of PRAD from Chen et al. were downloaded from <http://www.pradcellatlas.com/>. The scRNA-seq data of BRCA patients receiving pembrolizumab from Bassez et al. were downloaded from <https://lambrechtslab.sites.vib.be/en/single-cell> [41]. Reanalyzed publicly available ST data can be accessed from the GEO database under accession codes: GSE176078 [97], GSE203612 [101]. The ST data of CRC from Wu et al. were downloaded from <http://www.cancerdiversity.asia/scCRLM/> [79]. The ST data for LIHC1, LIHC2, LIHC3, and LIHC4 from Wu et al. were downloaded from <http://lifeome.net/supp/livercancer-st/> [102]. The ST data for OVCA_10x were downloaded from <https://www.10xgenomics.com/resources/datasets/human-ovarian-cancer-1-standard>. The ST data for PRAD1 were downloaded from <https://www.10xgenomics.com/resources/datasets/human-prostate-cancer-acinar-cell-carcinoma-ffpe-1-standard>. The ST data for PRAD2 were downloaded from <https://www.10xgenomics.com/resources/datasets/human-prostate-cancer-adenocarcinoma-with-invasive-carcinoma-ffpe-1-standard-1-3-0>. Reanalyzed publicly available RNA-seq data of melanoma patients undergoing immunotherapy (Riaz cohort) can be accessed from the GEO database under accession code GSE91061 [85]. Reanalyzed publicly available RNA-seq data of melanoma patients undergoing immunotherapy (Gide cohort [86] and Nathanson cohort [87]) were downloaded from TIDE database (<http://tide.dfci.harvard.edu/>) [88].

Declarations

Ethics approval and consent to participate

The study was approved by the Ethics Committee of Qilu Hospital of Shandong University (KYL-2017-256) and conducted in accordance with the Declaration of Helsinki. All subjects gave written informed consent before participating in the study.

Competing interests

The authors declare no competing interests.

Received: 11 July 2023 Accepted: 2 October 2023

Published online: 13 October 2023

References

- Li Y, Jin J, Bai F. Cancer biology deciphered by single-cell transcriptomic sequencing. *Protein Cell*. 2022;13:167–79. <https://doi.org/10.1007/s13238-021-00868-1>.
- Xu M, Zhang T, Xia R, Wei Y, Wei X. Targeting the tumor stroma for cancer therapy. *Mol Cancer*. 2022;21:208. <https://doi.org/10.1186/s12943-022-01670-1>.
- 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>.
- Pei L, et al. Roles of cancer-associated fibroblasts (CAFs) in anti-PD-1/PD-L1 immunotherapy for solid cancers. *Mol Cancer*. 2023;22:29. <https://doi.org/10.1186/s12943-023-01731-z>.

5. Tsoumakidou M. The advent of immune stimulating CAFs in cancer. *Nat Rev Cancer*. 2023;23:258–69. <https://doi.org/10.1038/s41568-023-00549-7>.
6. Chen Z, et al. Single-cell RNA sequencing highlights the role of inflammatory cancer-associated fibroblasts in bladder urothelial carcinoma. *Nat Commun*. 2020;11:5077. <https://doi.org/10.1038/s41467-020-18916-5>.
7. Choi JH, et al. Single-cell transcriptome profiling of the stepwise progression of head and neck cancer. *Nat Commun*. 2023;14:1055. <https://doi.org/10.1038/s41467-023-36691-x>.
8. Pu W, et al. Single-cell transcriptomic analysis of the tumor ecosystems underlying initiation and progression of papillary thyroid carcinoma. *Nat Commun*. 2021;12:6058. <https://doi.org/10.1038/s41467-021-26343-3>.
9. Hanley CJ, et al. Single-cell analysis reveals prognostic fibroblast subpopulations linked to molecular and immunological subtypes of lung cancer. *Nat Commun*. 2023;14:387. <https://doi.org/10.1038/s41467-023-35832-6>.
10. Galbo PM Jr, Zang X, Zheng D. Molecular Features of Cancer-associated Fibroblast Subtypes and their Implication on Cancer Pathogenesis, Prognosis, and Immunotherapy Resistance. *Clin Cancer Res*. 2021;27:2636–47. <https://doi.org/10.1158/1078-0432.CCR-20-4226>.
11. Luo H, et al. Pan-cancer single-cell analysis reveals the heterogeneity and plasticity of cancer-associated fibroblasts in the tumor micro-environment. *Nat Commun*. 2022;13:6619. <https://doi.org/10.1038/s41467-022-34395-2>.
12. Tian L, Chen F, Macosko EZ. The expanding vistas of spatial transcriptomics. *Nat Biotechnol*. 2022. <https://doi.org/10.1038/s41587-022-01448-2>.
13. Xue R, et al. Liver tumour immune microenvironment subtypes and neutrophil heterogeneity. *Nature*. 2022;612:141–7. <https://doi.org/10.1038/s41586-022-05400-x>.
14. Wei R, et al. Spatial charting of single-cell transcriptomes in tissues. *Nat Biotechnol*. 2022;40:1190–9. <https://doi.org/10.1038/s41587-022-01233-1>.
15. Liu C, et al. Single-cell dissection of cellular and molecular features underlying human cervical squamous cell carcinoma initiation and progression. *Sci Adv*. 2023;9:eadd8977. <https://doi.org/10.1126/sciadv.add8977>.
16. Wang Y, et al. Single-cell analysis of pancreatic ductal adenocarcinoma identifies a novel fibroblast subtype associated with poor prognosis but better immunotherapy response. *Cell Discov*. 2021;7:36. <https://doi.org/10.1038/s41421-021-00271-4>.
17. Lavie D, Ben-Shmuel A, Erez N, Scherz-Shouval R. Cancer-associated fibroblasts in the single-cell era. *Nat Cancer*. 2022;3:793–807. <https://doi.org/10.1038/s43018-022-00411-z>.
18. Hu J, et al. Tumor microenvironment remodeling after neoadjuvant immunotherapy in non-small cell lung cancer revealed by single-cell RNA sequencing. *Genome Med*. 2023;15:14. <https://doi.org/10.1186/s13073-023-01164-9>.
19. Zhang C, et al. A single-cell analysis reveals tumor heterogeneity and immune environment of acral melanoma. *Nat Commun*. 2022;13. <https://doi.org/10.1038/s41467-022-34877-3>.
20. Saito M, et al. CDX2 is involved in microRNA-associated inflammatory carcinogenesis in gastric cancer. *Oncol Lett*. 2017;14:6184–90. <https://doi.org/10.3892/ol.2017.6956>.
21. Tang X, et al. Stromal miR-200s contribute to breast cancer cell invasion through CAF activation and ECM remodeling. *Cell Death Differ*. 2016;23:132–45. <https://doi.org/10.1038/cdd.2015.78>.
22. Zhang J, et al. KLF16 Affects the MYC Signature and Tumor Growth in Prostate Cancer. *Onco Targets Ther*. 2020;13:1303–10. <https://doi.org/10.2147/OTT.S233495>.
23. Liang L, et al. 'Reverse Warburg effect' of cancer-associated fibroblasts (Review). *Int J Oncol*. 2022;60. <https://doi.org/10.3892/ijo.2022.5357>.
24. Li Z, Sun C, Qin Z. Metabolic reprogramming of cancer-associated fibroblasts and its effect on cancer cell reprogramming. *Theranostics*. 2021;11:8322–36. <https://doi.org/10.7150/thno.62378>.
25. Madar S, Goldstein I, Rotter V. 'Cancer associated fibroblasts'—more than meets the eye. *Trends Mol Med*. 2013;19:447–53. <https://doi.org/10.1016/j.molmed.2013.05.004>.
26. Hosaka K, et al. Pericyte-fibroblast transition promotes tumor growth and metastasis. *Proc Natl Acad Sci U S A*. 2016;113:E5618–5627. <https://doi.org/10.1073/pnas.1608384113>.
27. Kolde R, Laur S, Adler P, Vilo J. Robust rank aggregation for gene list integration and meta-analysis. *Bioinformatics*. 2012;28:573–80. <https://doi.org/10.1093/bioinformatics/btr709>.
28. Yuan Z, et al. Extracellular matrix remodeling in tumor progression and immune escape: from mechanisms to treatments. *Mol Cancer*. 2023;22:48. <https://doi.org/10.1186/s12943-023-01744-8>.
29. Cao Y, Langer R, Ferrara N. Targeting angiogenesis in oncology, ophthalmology and beyond. *Nat Rev Drug Discov*. 2023. <https://doi.org/10.1038/s41573-023-00671-z>.
30. Huinen ZR, Huijbers EJM, van Beijnum JR, Nowak-Sliwinska P, Griffioen AW. Anti-angiogenic agents - overcoming tumour endothelial cell energy and improving immunotherapy outcomes. *Nat Rev Clin Oncol*. 2021;18:527–40. <https://doi.org/10.1038/s41571-021-00496-y>.
31. Liu Y, et al. Recent progress on vascular endothelial growth factor receptor inhibitors with dual targeting capabilities for tumor therapy. *J Hematol Oncol*. 2022;15:89. <https://doi.org/10.1186/s13045-022-01310-7>.
32. Kerneux C, Cano CE, Olive D. Major pathways involved in macrophage polarization in cancer. *Front Immunol*. 2022;13:1026954. <https://doi.org/10.3389/fimmu.2022.1026954>.
33. Liu S, Ren J, Ten Dijke P. Targeting TGFbeta signal transduction for cancer therapy. *Signal Transduct Target Ther*. 2021;6:8. <https://doi.org/10.1038/s41392-020-00436-9>.
34. Videla-Richardson GA, et al. Galectins as Emerging Glyco-Checkpoints and Therapeutic Targets in Glioblastoma. *Int J Mol Sci*. 2021;23. <https://doi.org/10.3390/ijms23010316>.
35. Yu X, et al. Galectin-1: A Traditionally Immunosuppressive Protein Displays Context-Dependent Capacities. *Int J Mol Sci*. 2023;24. <https://doi.org/10.3390/ijms24076501>.
36. Chen J, et al. MIF inhibition alleviates vitiligo progression by suppressing CD8+ T cell activation and proliferation. *J Pathol*. 2023;260:84–96. <https://doi.org/10.1002/path.6073>.
37. Chen W. TGF-β Regulation of T Cells. *Annu Rev Immunol*. 2023. <https://doi.org/10.1146/annurev-immunol-101921-045939>.
38. Lau LS, Mohammed NBB, Dimitroff CJ. Decoding Strategies to Evade Immunoregulators Galectin-1, -3, and -9 and Their Ligands as Novel Therapeutics in Cancer Immunotherapy. *Int J Mol Sci*. 2022;23. <https://doi.org/10.3390/ijms232415554>.
39. Martinez GJ, et al. The transcription factor NFAT promotes exhaustion of activated CD8(+) T cells. *Immunity*. 2015;42:265–78. <https://doi.org/10.1016/j.immuni.2015.01.006>.
40. Oestreich KJ, Yoon H, Ahmed R, Boss JM. NFATc1 regulates PD-1 expression upon T cell activation. *J Immunol*. 2008;181:4832–4839. <https://doi.org/10.4049/jimmunol.181.7.4832>.
41. Bassez A, et al. A single-cell map of intratumoral changes during anti-PD1 treatment of patients with breast cancer. *Nat Med*. 2021;27:820–32. <https://doi.org/10.1038/s41591-021-01323-8>.
42. Xu N, et al. Chitinase-3-Like-1 Promotes M2 Macrophage Differentiation and Induces Choroidal Neovascularization in Neovascular Age-Related Macular Degeneration. *Invest Ophthalmol Vis Sci*. 2019;60:4596–605. <https://doi.org/10.1167/iovs.19-27493>.
43. de Azevedo RA, et al. MIF inhibition as a strategy for overcoming resistance to immune checkpoint blockade therapy in melanoma. *Oncimmunology*. 2020;9:1846915. <https://doi.org/10.1080/2162402X.2020.1846915>.
44. Jeong H, Lee SY, Seo H, Kim BJ. Recombinant Mycobacterium smegmatis delivering a fusion protein of human macrophage migration inhibitory factor (MIF) and IL-7 exerts an anticancer effect by inducing an immune response against MIF in a tumor-bearing mouse model. *J Immunother Cancer*. 2021;9. <https://doi.org/10.1136/jitc-2021-003180>.
45. Liu X, Qiao Y, Chen J, Ge G. Basement membrane promotes tumor development by attenuating T cell activation. *J Mol Cell Biol*. 2022;14. <https://doi.org/10.1093/jmcb/mjac006>.
46. Pixley FJ, Stanley ER. CSF-1 regulation of the wandering macrophage: complexity in action. *Trends Cell Biol*. 2004;14:628–38. <https://doi.org/10.1016/j.tcb.2004.09.016>.

47. Buechler MB, Fu W, Turley SJ. Fibroblast-macrophage reciprocal interactions in health, fibrosis, and cancer. *Immunity*. 2021;54:903–15. <https://doi.org/10.1016/j.immuni.2021.04.021>.
48. Hume DA, MacDonald KP. Therapeutic applications of macrophage colony-stimulating factor-1 (CSF-1) and antagonists of CSF-1 receptor (CSF-1R) signaling. *Blood*. 2012;119:1810–20. <https://doi.org/10.1182/blood-2011-09-379214>.
49. Xiang X, et al. Cancer-associated fibroblasts: Vital suppressors of the immune response in the tumor microenvironment. *Cytokine Growth Factor Rev*. 2022;67:35–48. <https://doi.org/10.1016/j.cytogfr.2022.07.006>.
50. Anagnostou V, Bardelli A, Chan TA, Turajlic S. The status of tumor mutational burden and immunotherapy. *Nat Cancer*. 2022;3:652–6. <https://doi.org/10.1038/s43018-022-00382-1>.
51. Hellmann MD, et al. Tumor Mutational Burden and Efficacy of Nivolumab Monotherapy and in Combination with Ipilimumab in Small-Cell Lung Cancer. *Cancer Cell*. 2019;35:329. <https://doi.org/10.1016/j.ccell.2019.01.011>.
52. Bagaev A, et al. Conserved pan-cancer microenvironment subtypes predict response to immunotherapy. *Cancer Cell*. 2021;39:845–865 e847. <https://doi.org/10.1016/j.ccell.2021.04.014>.
53. Sia D, et al. Identification of an Immune-specific Class of Hepatocellular Carcinoma Based on Molecular Features. *Gastroenterology*. 2017;153:812–26. <https://doi.org/10.1053/j.gastro.2017.06.007>.
54. Shen R, et al. Identification of Distinct Immune Subtypes in Colorectal Cancer Based on the Stromal Compartment. *Front Oncol*. 2019;9:1497. <https://doi.org/10.3389/fonc.2019.01497>.
55. Roh W, et al. Integrated molecular analysis of tumor biopsies on sequential CTLA-4 and PD-1 blockade reveals markers of response and resistance. *Sci Transl Med*. 2017;9. <https://doi.org/10.1126/scitranslmed.aah3560>.
56. Rooney MS, Shukla SA, Wu CJ, Getz G, Hacohen N. Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell*. 2015;160:48–61. <https://doi.org/10.1016/j.cell.2014.12.033>.
57. Cabrita R, et al. Tertiary lymphoid structures improve immunotherapy and survival in melanoma. *Nature*. 2020;577:561–5. <https://doi.org/10.1038/s41586-019-1914-8>.
58. Ayers M, et al. IFN-gamma-related mRNA profile predicts clinical response to PD-1 blockade. *J Clin Invest*. 2017;127:2930–40. <https://doi.org/10.1172/JCI91190>.
59. Arico E, Castiello L, Capone I, Gabriele L, Belardelli F. Type I Interferons and Cancer: An Evolving Story Demanding Novel Clinical Applications. *Cancers (Basel)*. 2019;11. <https://doi.org/10.3390/cancers11121943>.
60. Cheon H, Wang Y, Wightman SM, Jackson MW, Stark GR. How cancer cells make and respond to interferon- γ . *Trends in cancer*. 2023;9:83–92. <https://doi.org/10.1016/j.trecan.2022.09.003>.
61. Liang Y, Hannan R, Fu YX. Type I IFN Activating Type I Dendritic Cells for Antitumor Immunity. *Clin Cancer Res*. 2021;27:3818–24. <https://doi.org/10.1158/1078-0432.Ccr-20-2564>.
62. Lim J, et al. Harnessing type I interferon-mediated immunity to target malignant brain tumors. *Front Immunol*. 2023;14:1203929. <https://doi.org/10.3389/fimmu.2023.1203929>.
63. Snell LM, McGaha TL, Brooks DG. Type I Interferon in Chronic Virus Infection and Cancer. *Trends Immunol*. 2017;38:542–57. <https://doi.org/10.1016/j.it.2017.05.005>.
64. You M, et al. Signaling pathways in cancer metabolism: mechanisms and therapeutic targets. *Signal Transduct Target Ther*. 2023;8:196. <https://doi.org/10.1038/s41392-023-01442-3>.
65. Murgai M, et al. KLF4-dependent perivascular cell plasticity mediates pre-metastatic niche formation and metastasis. *Nat Med*. 2017;23:1176–90. <https://doi.org/10.1038/nm.4400>.
66. Feng F, Feng X, Zhang D, Li Q, Yao L. Matrix Stiffness Induces Pericyte-Fibroblast Transition Through YAP Activation. *Front Pharmacol*. 2021;12:698275. <https://doi.org/10.3389/fphar.2021.698275>.
67. Ouyang JF, Kamaraj US, Cao EY, Rackham OJL. ShinyCell: simple and sharable visualization of single-cell gene expression data. *Bioinformatics*. 2021;37:3374–6. <https://doi.org/10.1093/bioinformatics/btab209>.
68. 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>.
69. McGinnis CS, Murrow LM, Gartner ZJ. DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. *Cell Syst*. 2019;8:329–337 e324. <https://doi.org/10.1016/j.cels.2019.03.003>.
70. Korsunsky I, et al. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat Methods*. 2019;16:1289–96. <https://doi.org/10.1038/s41592-019-0619-0>.
71. Gao R, et al. Delineating copy number and clonal substructure in human tumors from single-cell transcriptomes. *Nat Biotechnol*. 2021;39:599–608. <https://doi.org/10.1038/s41587-020-00795-2>.
72. Yu G. Using ggtree to Visualize Data on Tree-Like Structures. *Curr Protoc Bioinformatics*. 2020;69: e96. <https://doi.org/10.1002/cpbi.96>.
73. Liao Y, Wang J, Jaehnig EJ, Shi Z, Zhang B. WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Res*. 2019;47:W199–205. <https://doi.org/10.1093/nar/gkz401>.
74. Zheng L, et al. Pan-cancer single-cell landscape of tumor-infiltrating T cells. *Science*. 2021;374:abe6474. <https://doi.org/10.1126/science.abe6474>.
75. Van de Sande B, et al. A scalable SCENIC workflow for single-cell gene regulatory network analysis. *Nat Protoc*. 2020;15:2247–76. <https://doi.org/10.1038/s41596-020-0336-2>.
76. Suo S, et al. Revealing the Critical Regulators of Cell Identity in the Mouse Cell Atlas. *Cell Rep*. 2018;25:1436–1445 e1433. <https://doi.org/10.1016/j.celrep.2018.10.045>.
77. Azizi E, et al. Single-Cell Map of Diverse Immune Phenotypes in the Breast Tumor Microenvironment. *Cell*. 2018;174:1293–1308 e1236. <https://doi.org/10.1016/j.cell.2018.05.060>.
78. Bibby JA, et al. Systematic single-cell pathway analysis to characterize early T cell activation. *Cell Rep*. 2022;41: 111697. <https://doi.org/10.1016/j.celrep.2022.111697>.
79. Wu Y, et al. Spatiotemporal Immune Landscape of Colorectal Cancer Liver Metastasis at Single-Cell Level. *Cancer Discov*. 2022;12:134–53. <https://doi.org/10.1158/2159-8290.CD-21-0316>.
80. Street K, et al. Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics*. 2018;19:477. <https://doi.org/10.1186/s12864-018-4772-0>.
81. Cao EY, Ouyang JF, Rackham OJL. GeneSwitches: ordering gene expression and functional events in single-cell experiments. *Bioinformatics*. 2020;36:3273–5. <https://doi.org/10.1093/bioinformatics/btaa099>.
82. Marjanovic ND, et al. Emergence of a High-Plasticity Cell State during Lung Cancer Evolution. *Cancer Cell*. 2020;38:229–246 e213. <https://doi.org/10.1016/j.ccell.2020.06.012>.
83. Shao X, et al. Knowledge-graph-based cell-cell communication inference for spatially resolved transcriptomic data with SpaTalk. *Nat Commun*. 2022;13:4429. <https://doi.org/10.1038/s41467-022-32111-8>.
84. Jin S, et al. Inference and analysis of cell-cell communication using Cell Chat. *Nat Commun*. 2021;12:1088. <https://doi.org/10.1038/s41467-021-21246-9>.
85. Riaz N, et al. Tumor and Microenvironment Evolution during Immunotherapy with Nivolumab. *Cell*. 2017;171:934–949 e916. <https://doi.org/10.1016/j.cell.2017.09.028>.
86. Gide TN, et al. Distinct Immune Cell Populations Define Response to Anti-PD-1 Monotherapy and Anti-PD-1/Anti-CTLA-4 Combined Therapy. *Cancer Cell*. 2019;35:238–255 e236. <https://doi.org/10.1016/j.ccell.2019.01.003>.
87. Nathanson T, et al. Somatic Mutations and Neoepitope Homology in Melanomas Treated with CTLA-4 Blockade. *Cancer Immunol Res*. 2017;5:84–91. <https://doi.org/10.1158/2326-6066.CIR-16-0019>.
88. Jiang P, et al. Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nat Med*. 2018;24:1550–8. <https://doi.org/10.1038/s41591-018-0136-1>.
89. Yoshihara K, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun*. 2013;4:2612. <https://doi.org/10.1038/ncomms3612>.
90. Colaprico A, et al. TCGAAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res*. 2016;44: e71. <https://doi.org/10.1093/nar/gkv1507>.
91. Mayakonda A, Lin DC, Assenov Y, Plass C, Koeffler HP. Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res*. 2018;28:1747–56. <https://doi.org/10.1101/gr.239244.118>.
92. Mermel CH, et al. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol*. 2011;12:R41. <https://doi.org/10.1186/gb-2011-12-4-r41>.

93. Reich M, et al. GenePattern 2.0. *Nat Genet.* 2006;38:500–501. <https://doi.org/10.1038/ng0506-500>.
94. Ru B, et al. TISIDB: an integrated repository portal for tumor-immune system interactions. *Bioinformatics.* 2019;35:4200–2. <https://doi.org/10.1093/bioinformatics/btz210>.
95. 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>.
96. Schubert M, et al. Perturbation-response genes reveal signaling footprints in cancer gene expression. *Nat Commun.* 2018;9:20. <https://doi.org/10.1038/s41467-017-02391-6>.
97. Wu SZ, et al. A single-cell and spatially resolved atlas of human breast cancers. *Nat Genet.* 2021;53:1334–47. <https://doi.org/10.1038/s41588-021-00911-1>.
98. Uhlitz F, et al. Mitogen-activated protein kinase activity drives cell trajectories in colorectal cancer. *EMBO Mol Med.* 2021;13: e14123. <https://doi.org/10.15252/emmm.202114123>.
99. Lu Y, 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>.
100. Xu J, et al. Single-Cell RNA Sequencing Reveals the Tissue Architecture in Human High-Grade Serous Ovarian Cancer. *Clin Cancer Res.* 2022;28:3590–602. <https://doi.org/10.1158/1078-0432.CCR-22-0296>.
101. Barkley D, et al. Cancer cell states recur across tumor types and form specific interactions with the tumor microenvironment. *Nat Genet.* 2022;54:1192–201. <https://doi.org/10.1038/s41588-022-01141-9>.
102. Wu R, et al. Comprehensive analysis of spatial architecture in primary liver cancer. *Sci Adv.* 2021;7:eabg3750. <https://doi.org/10.1126/sciadv.abg3750>.

Publisher's Note

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

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

