

# Single-cell sequencing reveals the landscape of the tumor microenvironment in malignant fibrous histiocytoma

**Liu-Liu Yuan**

Jiangsu Key Laboratory of Bioactive Natural Product Research and State Key Laboratory of Natural Medicines

**Zhong Chen**

Sir Run Run Hospital, Nanjing Medical University

**Jian Qin**

Sir Run Run Hospital, Nanjing Medical University

**Cheng-Jiao Qin**

Jiangsu Key Laboratory of Bioactive Natural Product Research and State Key Laboratory of Natural Medicines

**Jing Bian**

Jiangsu Key Laboratory of Bioactive Natural Product Research and State Key Laboratory of Natural Medicines

**Rui-Fang Dong**

Jiangsu Key Laboratory of Bioactive Natural Product Research and State Key Laboratory of Natural Medicines

**Tang-Bo Yuan**

Sir Run Run Hospital, Nanjing Medical University

**Yi-Ting Xu**

Jiangsu Key Laboratory of Bioactive Natural Product Research and State Key Laboratory of Natural Medicines

**Lingyi Kong**

China Pharmaceutical University <https://orcid.org/0000-0001-9712-2618>

**Yuan-Zheng Xia** (✉ [xiayz@cpu.edu.cn](mailto:xiayz@cpu.edu.cn))

Jiangsu Key Laboratory of Bioactive Natural Product Research and State Key Laboratory of Natural Medicines

---

**Article**

**Keywords:**

**Posted Date:** March 30th, 2022

**DOI:** <https://doi.org/10.21203/rs.3.rs-1427399/v2>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

Malignant fibrous histiocytoma (MFH) is an invasive pleomorphic soft tissue sarcoma with a high degree of malignancy and poor prognoses, and is prone to recurrence and metastasis. The pathophysiology remains elusive and its therapeutic options are limited. The 5-year recurrence rate of patients is 36–61%. Progress in single-cell RNA sequencing (scRNA-seq) provides an opportunity to dissect the pathophysiology of human diseases at unprecedented resolution, particularly in the diseases lacking animal models, such as MFH. We performed scRNA-seq on tumor tissues and adjacent muscle tissues from a patient with MFH. We identified Cell types and the corresponding marker genes by single-cell RNA sequencing. Malignant cells of fibroblasts were evaluated by CopyKAT analysis and differentially expressed genes of sequencing. We identified *PDCD1*, *CTLA4* and *TIGIT* as potential targets. We further showed that C2\_Exhausted CD4 + Treg and C1\_Exhausted CD8 + T cell highly expressed *PDCD1*, *CTLA4* and *TIGIT*. Intervention via PD-1 immune checkpoint inhibitor (trelizumab) enabled disease control and reduced tumor immunosuppression. Thus, scRNA-seq analyses guide a successful therapeutic intervention in the MFH patient, improve our understanding of complex human diseases and provide an alternative approach to personalized medicine.

## Introduction

Malignant fibrous histiocytoma (MFH), also known as undifferentiated pleomorphic sarcoma (UPS), which is initially identified as a histologically distinctive type of sarcoma in 1964<sup>1</sup>. Bone neoplasms are rare solid tumors, accounting for less than 2% of all primary malignancies. The first case of MFH of the bone (MFH-B) was described in 1972, and MFH-B is an extremely rare and aggressive malignancy representing < 1% of all primary malignant bone tumors. MFH-B often occurs in the bone diaphysis or metaphysis that results in invasive bone damage and soft tissue mass. MFH-B is found in the extremities 70–75% of the time with about 50% of the cases in the lower extremities<sup>2</sup>. MFH-B is a matrix-producing malignant tumor with a pleomorphic spindle-cell structure, which is devoid of any specific pattern of histologic differentiation<sup>3</sup>. MFH-B often exhibits an aggressive behavior associated with a high metastatic potential and high frequency of local recurrence. The principles of management of MFH-B are somewhat similar to those of osteosarcoma. A combination approach involving surgical resection and neoadjuvant chemotherapy is considered the mainstay of treatment<sup>4</sup>. MFH was reported as somatic soft tissue sarcoma in many studies, but bony involvement is relatively rare. Here, we report a rare case of MFH-B at the lower femoral end of the left thigh in a 44-year old man. However, current knowledge of MFH-B is limited to case reports and small case series. The clinicopathological features and prognosis, tumor microenvironment, and tumor heterogeneity of these cancers have not been well defined.

Over the past ten years, the rapid development of new technologies has enabled us to quickly obtain a large number of physiological and pathological information of various tumors<sup>5</sup>. Given the recent advances in single-cell RNA sequencing (scRNA-seq), its application in human diseases may enable better understanding of pathological processes. We hypothesized that scRNA-seq might provide a

powerful personalized medicine approach to determine transcriptomic changes that not only deepen our insight into disease mechanisms, but also authorize the identification of overexpressed genes or altered pathways that might be targeted via currently available monoclonal antibodies or small-molecule inhibitors. The advent of scRNA-seq has enabled the profiling of specific cell populations at a single-cell level<sup>6-8</sup>. Notably, scRNA-seq has been widely used to reveal the characteristics of immunity in various fields because it can detect changes in individual cell types. The emergence of single-cell sequencing represents a powerful tool to resolve tumor heterogeneity and delineate the complex communication among the tumor cells with neighboring stromal and immune cells in the tumor microenvironment (TME)<sup>9</sup>. Here, we performed scRNA-seq of MFH adjacent muscle tissues and tumor tissues. These unprecedented data uncovered the transcriptional landscape and phenotypic heterogeneity of tumor and immune cells in malignant fibrous histiocytoma (MFH) and identified their gene expression signature, suggesting specialized functions.

Targeting specific immunological pathways represents a promising approach to fight tumor diseases<sup>10</sup>. We found that C2\_Exhausted CD4 + Treg in the sample tissue of this case positively expressed T cell depletion inhibitory receptors, including *PDCD1* and *CTLA4*. At present, inhibitors corresponding to PDCD1 and CTLA4 are widely used in clinical practice. We suggest that the attending physician of this patient consider using one or both of these inhibitors to treat this patient. There are no guidelines for the employment of immune checkpoint inhibitors (ICIs) in MFH therapy, and the treatment outcome has rarely been reported. We report the first case of MFH with ICIs of programmed death protein-1 (PD-1) treatment, describing the clinical features, imaging, pathological findings, and tumor microenvironment. This case showed better treatment outcome than previous cases of MFH. ICIs of PD-1 may be considered as the first-line therapy for patients with MFH.

## Results

### Imaging examinations and further diagnostic work-up of MFH

Here, we report a rare case of MFH-B in the lower femoral segment of the left thigh in a 44-year-old male patient. The patient complained of pain and discomfort in the distal left thigh and went to a local hospital for symptomatic conservative treatment. The result of treatment was not good in the local hospital. The patient suffered from severe pain and found a hard mass at the distal end of his left thigh for one month. Therefore, he came to our hospital for further diagnosis and treatment. X-ray showed irregular thawing bone destruction at the lower end of the left femur, which invaded the bone marrow cavity. Multiple strip like high-density dead bone shadows were seen inside (Fig. 1A). CT and MRI showed mass of soft tissue at the lower end of the left femur and its surrounding (Supplementary Fig. S1B). HE and Immunohistochemical staining showed SMA (+), CD68 (+), Ki67 (+), CD34 (+), Desmin (-), EMA (-), S-100 (-), Bcl-2(-) and the pathological result was malignant fibrous histiocytoma (Supplementary Fig. S1A)<sup>1</sup>. In

order to reduce the burden of tumor and relieve the symptoms of leg pain, tumor resection was performed to remove the tumor. The patient was initially diagnosed with MFH at first resection of the tumor.

Two months later, the results of the patient's MR imaging showed that patchy abnormal signal was seen in the left side of the L5 spinous process and the posterior edge of the vertebral body, which was slightly more advanced than before, and the enhancement was slightly increased. The abnormal signal at the left posterior border of sacrum was similar to the anterior range and slightly reduced in signal. The possibility of metastasis was considered in the combination with medical history (Fig. 1E). After the patient developed local recurrence and metastasis, he was treated with PD-1 tirelizumab in his second round of chemotherapy.

## Single-cell RNA-seq identified MFH associated cellular constitution in tumor tissues and adjacent muscle tissues

To discover the alteration of gene expression in MFH, we performed scRNA-seq of a patient referred to the SIR RUN RUN HOSPITAL NANJING MEDICAL UNIVERSITY. Thus, we dissociated tumor tissues into a single-cell suspension and performed scRNA-seq analysis. Adjacent muscle tissues served as controls. The harvested cells from different groups were sequenced on a 10 × Genomics platform (Fig. 1F), we obtained a total of 18433 single cell transcriptomes (10532 adjacent muscle tissues; 7901 tumor tissues) from two groups of organizations. We conducted preliminary quality control and evaluation on the sequencing results, removed reads with low sequencing quality, compared reads with the reference genome using cellranger, annotated reads as specific genes, and then corrected unique molecular indexes (UMI) and counted them (Supplementary Fig. S2, S3E).

Unbiased clustering of the cells identified 19 clusters in parallel, based on t-distributed stochastic neighbor embedding (t-SNE) and uniform manifold approximation and projection (UMAP) analyses according to their gene profiles and canonical markers (Supplementary Fig. S3F). Our initial goal was to visualize and ultimately define the various cell subsets in the dataset. Subsequently, we used t-SNE visualization of the cells to reveal 10 major clusters, including CD4 + T cell (2 cell clusters), CD8 + T cell (4 cell clusters), NKT cell (1 cell cluster), Proliferative T cell (1 cell cluster), Macrophage (3 cell clusters), Osteoclast (1 cell cluster), Fibroblast (2 cell clusters), Proliferative Fibroblast (3 cell clusters), B cell (1 cell cluster), Mast cell (1 cell cluster) (Fig. 1G, H). In particular, we identified the marker genes for each clusters as follows <sup>8,11,12</sup>: (1) the CD4 + T cell highly expressing *CD4*, but low expressing *CD8*; (2) the CD8 + T cell highly expressing *CD8*, but low expressing *CD4*; (3) the NKT cell highly expressing NK cell markers and T cell markers *GNLY*, *GZMB*; (4) the Proliferative T cell highly expressing T cell markers and cell proliferating markers *TUBA1B*, and *MKI67*; (5) the Macrophage with high expression of the markers *C1QC*, *C1QA*; (6) the Osteoclast specifically express the markers *CTSK* and *MMP9*; (7) the Fibroblast expressing *COL1A1* and *FN1*; (8) the Proliferative Fibroblast with high expression of proliferating markers and fibroblast markers; (9) the B cell specifically expressing *IGHM*, *CD79A*; and (10) the Mast cell highly expressing *TPSAB1* and *TPSB2* (Fig. 1H). The results were consistent with the similarity analysis of each cell population (Supplementary Fig. S3A). Subsequently, Gene Ontology (GO) enrichment analysis and

Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis with up-expressed genes identified specific processes relevant to them (Supplementary Fig. S4), which were coincident with their cell type.

By comparing adjacent muscle tissues, we obtained the relative abundance of tumor cells and infiltrating immune cells in tumor tissues, which first reveal the landscape of tumors and invasive immune cells in MFH. These results provide deeper insights into the mechanisms of tumor clearance by immune cells in the surrounding microenvironment and contribute to improved targeting and immunotherapy for MFH<sup>13,14</sup>. Next, we attempted to discern the cellular differences between tumor tissues and adjacent muscle tissues. We noticed that almost all types of cell populations were present in the tumor tissues and adjacent muscle tissues except for part of the Proliferative Fibroblast, some of them were almost only recognized in adjacent muscle tissues, and some of them were predominantly recognized in the tumor tissues (Fig. 1G). The proportion of the cellular clusters varied significantly among the tissues, suggesting the intertumoral heterogeneity as well as the consistency among the tissues (Supplementary Fig. S3B).

GO and KEGG analyzed the differentially expressed genes (DEGs) in tumor tissues and adjacent muscle tissues, and we found that DEGs were enriched in different biological processes, indicating that MFH is related to the regulation of immune system process, cell activation, blood vessel development, cell migration and cell motility (Fig. 1B). We found numerous distinctions in the composition ratio and structure of the cell types, as well as in the expression of genes within the clusters, suggesting that the biological features of tumor tissues differed from those of adjacent muscle tissues (Supplementary Fig. S3C). GO and KEGG analyzed the DEGs of non-immune cells in tumor tissues and adjacent muscle tissues, and we revealed that the DEGs were enriched in Pathway in cancer (Fig. 1C). The scRNA-seq data quantified the expression of genes associated with disease developmental pathways, including TGF- $\beta$ , MAPK, NF- $\kappa$ B, and JAK-STAT, as well as signaling pathways associated with epithelial-to-mesenchymal transition (EMT), in various MFH cell populations (Supplementary Fig. S5, Fig. 1D). In our tissue sample, we observed upregulation of ligands and target genes of these pathways in Osteoclast, but few changes were found in immune cells. We also detected downregulated expression of MAPK pathway genes in several Proliferative Fibroblast populations. Some genes related to EMT, NF- $\kappa$ B, and JAK-STAT pathways were upregulated in Osteoclast and Fibroblast within MFH lesions. EMT programs have been indicated to play an important role in cancer invasion, metastasis and acquiring drug resistance. The analysis of gene signatures associated with EMT programming exhibited that the procedure consisted of EMT markers was significantly highly expressed in the Osteoclast and Fibroblast<sup>15</sup>, suggesting that most of the Osteoclast and Fibroblast in this sample were undergoing an active EMT process. Interestingly, immune cells showed few EMT-related genes except Macrophage, which expressed abundant EMT-related genes (Supplementary Fig. S5, Fig. 1D).

In consistent with the high program of EMT, the MFH sample also showed the significant high levels of invasion and metastasis as well as the high angiogenesis genes signature score (Supplementary Fig. S3D), indicating that the MFH in this patient might have an increased capability for high-grade metastasis, which highly correlate with poor prognosis<sup>7</sup>. Indeed, this patient just presented with metastasis based on clinical examination about two months after the surgery. MFH exhibits the highest

number of tumor-infiltrating lymphocytes (TILs), suggesting that MFH could benefit from immune checkpoint blockade (ICB). However, not all MFH patients respond to neoadjuvant ICB. An outstanding question is therefore to identify which underlying mechanisms and associated markers determine therapeutic response. In fact, TILs represent a heterogeneous population of cells with respect to cell type composition, gene expression and functional properties. So far, TIL scores and tumor PD-L1 expression have been proposed to predict clinical outcome, but their ability to act as predictors for MFH remains unclear<sup>16,17</sup>.

## Gene expression heterogeneity of T cell subsets was identified in the MFH

T cells are the key elements of cancer immunotherapy. However, their high heterogeneity regarding their cell-type compositions, gene expression patterns and functional properties significantly influence the outcomes of the T cell-based immunotherapy. Interestingly, we found that T cell clusters were present at high levels in the immune cells (Supplementary Fig. S3E), and the presence of infiltrating T cell in tumor tissues was confirmed using immunohistochemistry (IHC) with CD3E, CD4, CD8 antibodies, which was consistent with the results of the scRNA-seq data (Supplementary Fig. S6A), thus demonstrated that the T cell-based immunotherapy might be efficient in the MFH patients (Fig. 1G). Notably, we observed that the overall number of T cell in the tumor tissues was much lower than that in the adjacent muscle tissues<sup>11</sup>, indicating that T cell infiltration was inefficient (Supplementary Fig. S6D). There are many differences in gene expression within T cell clusters, suggesting that T cell biology in tumor tissues differs from that in adjacent muscle tissues (Supplementary Fig. S6B).

To reveal the intrinsic structure and potential functional subtypes of the overall T cell populations, we performed unsupervised clustering of all T cells via t-SNE and UMAP algorithm, subsequently identified 12 distinct subclusters of T cell. These 12 stable clusters include 8 clusters for CD8<sup>+</sup> and 4 clusters for CD4<sup>+</sup> cells. Each cluster has its unique signature genes (Fig. 2A, 3A and Supplementary Fig. S8F).

Cells of the first CD8<sup>+</sup> cluster, C1\_Exhausted CD8<sup>+</sup> T, expressed high levels of exhaustion markers *LAG3*, *PDCD1*, and *HAVCR2*. Cells of the second CD8<sup>+</sup> cluster, C2\_Naïve CD8<sup>+</sup> T, specifically expressed “naïve” marker genes such as *LEF1*, *TCF7* and *CCR7*. The third cluster, C3\_Proliferative CD8<sup>+</sup> T, characterized by specific expression of *MKI67*, *TOP2A*, and *PCNA*. The fourth cluster, C4\_Effector CD8<sup>+</sup> T, was identified by the high expression of the *CX3CR1*, *FCGR3A*, and *FGFBP2*, which commonly associated with T cell possessing effector functions. The fifth cluster, C5\_Cytotoxic CD8<sup>+</sup> T, was characterized with relatively high expression of the cytotoxic molecules *IL7R*, *TNF* and *CD69*. The sixth cluster, C6\_Proliferative APOE<sup>+</sup> CD8<sup>+</sup> T, shared the proliferative marker with cluster3 and highly expressing APOE. C8\_APOE<sup>+</sup> CD8<sup>+</sup> T, highly expressed *APOE*<sup>8,11,18</sup>. The remaining cells were primarily T helper cells, characterized by the high expression of *GZMA* and *CCL5*, and fallen into the seventh cluster C7\_CD8<sup>+</sup> helper T (Fig. 2E and Supplementary Fig. S7A, B).

We counted the cell number and proportion of each cellular subcluster (Fig. 2C, D and Supplementary Fig. S8B, C). The percentage of C1\_Exhausted CD8 + T cell within CD8 + T cell isolated from tumor tissues and adjacent muscle tissues was much higher than other cell types, revealing potential enrichment of exhausted CD8 + T cell in the TME (Fig. 2C, D). The percentage of C1\_Exhausted CD8 + T cell was increased significantly in tumor tissues, which was consistent with previous findings. GO enrichment analysis showed that C1\_ Exhausted CD8 + T showed a state of loss of function, persisted in the tumor tissues but replied poorly to the tumor cells <sup>16,17</sup>. These exhausted CD8 + T cells expressed high levels of *PDCD1* in TME and can be rescued from unresponsive and depleted state by ICBs (Fig. 2H). At the same time, we observed that the overall number of CD8 + T cells in the tumor tissues was lower than that in the adjacent muscle tissues <sup>11</sup>, indicating that the infiltration efficiency of CD8 + T cells was low and might be associated with poor prognosis (Fig. 2B).

Next, we generated computationally imputed pseudotime trajectories using Partition based graph abstraction (PAGA) <sup>19</sup> and RNA velocity analysis to infer the course of maturation of CD8 + T cells in MFH <sup>20</sup>. Most cells from each cluster were gathered based on the similar gene expression, and variant subsets formed into a relative process in pseudotime. We observed three distinct CD8 + T cell trajectories that began with the C2\_Naïve CD8 + T, which we considered as the root of the trajectories (Fig. 2F and Supplementary Fig. S7C). One of these trajectories formed C3\_Proliferative CD8 + T and ended with C1\_Exhausted CD8 + T. Consequently, those exhausted CD8 + T cells were highly enriched at the late period of the pseudotime, demonstrating that the CD8 + T cell state transformed from activation to exhaustion, Monocle 3 algorithm confirmed these trajectories (Fig. 2G). We analyzed gene expression patterns involved in CD8 + T cell-state transitions. The genes regarding “positive regulation of immune system process” decreased significantly along the pseudo-time axis, while the genes related to “cellular response to cytokine stimulus” increased significantly. Genes referred to “positive regulation of cell adhesion” increased initially and then decreased along the pseudo-time axis (Fig. 2I).

We identified 8 groups of DEGs along the trajectory of C1\_Exhausted CD8 + T cells. Firstly, Naïve T cell markers *CCR7*, *LEF1* and *TCF7* were reduced following the trajectory <sup>20</sup>. The subsequent groups increased at the end of the trajectory and were characterized by effector *TNFSF9*, cytotoxicity *GZMK*, *GZMA*, *NKG7*, and early markers of general exhaustion *PDCD1*, *CTLA4*, and *TIGIT*. In the last two groups, the early-activating genes *TGFB1*, *GZMM* and *TNF* increased midway through the trajectory, but decreased thereafter. In each gene set, we authenticated several genes that are previously unidentified as T cell markers (for example, *COTL1*, *CCDC141* as differentiated exhausted CD8 + T markers) (Fig. 2J).

Cells from the first CD4 + cluster, C1\_Cytotoxic CD4 + T, overexpressed cytotoxic molecules *IL7R*, *TNF* and *CD69*. Cells in the second CD4 + cluster, C2\_Exhausted CD4 + Treg, with high expression of *FOXP3*, *IL2RA*, the canonical exhausted gene signatures such as *PDCD1*, *CTLA4* were also significantly over expressed. The third cluster, C3\_Effector CD4 + helper T, characterized by specific expression of *NKG7*, *GZMK*, representing effector T cells, and also highly expressed *CCL5* and *GZMA*, describing helper T cells. The

fourth cluster, C4\_Naïve CD4 + T<sup>8,11</sup>, specifically expressed “naïve” marker genes such as *LEF1*, *TCF7* and *CCR7* (Fig. 3B and Supplementary Fig. S8A, S8E).

Importantly, we also found that C2\_Exhausted CD4 + Treg and C3\_Effector CD4 + helper T positively expressed the inhibitory receptors and ligands (IRs) including *TIGIT*, *CTLA4*, *PDCD1* and *LAG3*, IRs were associated with the exhaustion program of dysfunctional tumor-infiltrating lymphocytes (TILs), suggesting that these cells became exhausted after the initial activation in the MFH. Treg showed relatively high levels of immune inhibitory molecules *TIGIT*, *CTLA4*, *PDCD1* and *TNFRSF18*, which may contribute to Treg-mediated suppression of anti-tumor immune responses in the MFH. Recently, the anti-TIGIT therapeutics have drawn great attention in treating colorectal cancer, breast cancer, and melanoma through modulating the activities of CD8 + T, Treg, and NK cells. We also noticed that TIGIT was widely expressed in CD8 + T cells corresponding to gradual loss of responsiveness (Supplementary Fig. S6C). In the new era of immunotherapies, ICIs including antibodies against PD-1 (nivolumab) and CTLA4 (ipilimumab) are widely used for cancer treatment. ICIs act by blocking the inhibitory receptors of immune system on T cells (PD-1 and CTLA4), and thereby activate tumor-specific T cells to destroy tumor cells. Recently, Nivolumab in combination with ipilimumab were reported to have survival benefits in patients suffering from hepatic melanoma of unknown primary origin. These messages illustrate that TIGIT, CTLA4 and PDCD1 blockade could be an effective therapy for MFH<sup>16,17</sup>.

According to the tips of the patient's pelvic plain scan on August 20th, 2021, abnormal imaging signal was found around the left iliac vessels and left inguinal region, considering lymph node enlargement. The soft tissue thickening accompanied by abnormal enhancement signal was detected around left L5 lamina. Therefore, the possibility of distant metastasis was considered (Fig. 1E). In mid-August, due to the outbreak in Nanjing and personal economic reasons, the patient underwent a second round of chemotherapy in Nanjing Gaochun people's Hospital by using epirubicin hydrochloride 100mg, ifosfamide 2.0g, trelizumab (PD-1 antibody) 200mg and anlotinib 10mg. On September 18th, 2021 patient reexamined blood routine, the examination of male tumor marker showed that alpha fetoprotein, carcinoembryonic antigen, carbohydrate antigen 19 - 9 and carbohydrate antigen 12 - 5 returned to normal.

Next, we analyzed C2\_Exhausted CD4 + Treg and obtained a list of 54 exhaustion-specific genes by comparing exhausted and non-exhausted CD4 + T cells. Multiple known exhaustion markers, such as *HAVCR2*, *PDCD1*, *ENTPD1*, *CTLA4*, *TIGIT*, *TNFRSF9* and *CD27*, were selected. The 54-gene list also contained several little-described genes<sup>11</sup>, such as *MYO7A*, *TOX* and *CXCL13*, as well as novel exhaustion makers such as *LAYN*, *PHLDA1*, and *SNAP47* (Fig. 3H). Based on the TCGA data, high levels of *LAYN* associated with poor prognosis (Fig. 3I, J). *LAYN*, encoding layilin, was recently reported to be highly expressed in Treg isolated from hepatocellular carcinoma. Besides, *LAYN* is linked to the suppressive function of tumor Treg and exhausted CD8 T cells. Thus, our data not only confirmed previously identified genes associated with exhausted CD4 + T cells and Treg, but also discovered additional markers for these cell types<sup>21</sup>.

The percentage of C1\_Cytotoxic CD4 + T cells within CD4 + T cells isolated from tumor tissues and adjacent muscle tissues was much higher than other cell types, displaying potential enrichment of C1\_Cytotoxic CD4 + T cells in the tumor microenvironment (Supplementary Fig. S8B, C). GO enrichment analysis showed that C2\_Exhausted CD4 + Treg was enriched with functions involving “lymphocyte activation” and “response to cytokine” (Fig. 3G) <sup>22</sup>.

After defining the CD8 + T and CD4 + T cells in our dataset, we recognized the DEGs in tumor tissues and adjacent muscle tissues of MFH. We identified the differential gene set for these CD8 + T and CD4 + T cells that allowed a more in-depth analysis of regulatory pathways (Supplementary Fig. S6B). KEGG enrichment profile analysis demonstrated that the majority of the upregulated genes of the CD8 + T and CD4 + T cells focused on IL-17 signaling pathway <sup>23</sup>, Osteoclast differentiation and TNF signaling pathway (Supplementary Fig. S7D, S8G).

Similarly, we generated computationally imputed pseudotime trajectories using PAGA and RNA velocity analysis <sup>19,20</sup> to infer the course of maturation of CD4 + T cells in MFH. We observed three distinct CD4 + T cell trajectories that began with the C4\_Naïve CD4 + T, which was considered as the root of the trajectories. One of these trajectories developed C1\_Cytotoxic CD4 + T and ended with C2\_Exhausted CD4 + Treg. Accordingly, those C2\_Exhausted CD4 + Treg were highly enriched at the late period of the pseudotime, demonstrating that the CD4 + T cell state transformed from activation to exhaustion (Fig. 3C, D). Monocle 3 algorithm confirmed these trajectories (Fig. 3E), and the profiling of marker genes confirmed their functional annotation (Supplementary Fig. S8D).

We analyzed gene expression patterns involved in CD4 + T cell-state transitions <sup>24</sup>. The genes associated with “cytokine-mediated signaling pathway” and “leukocyte differentiation” decreased significantly along the pseudo-time axis, while the genes related to “protein localization to membrane” and “establishment of protein localization to organelle” increased markedly. Genes referred to “positive regulation of metabolic process” and “positive regulation of biological process” were increased at first and then decreased along the pseudo-time axis (Fig. 3F).

## Two Distinct States of Tumor-Enriched Macrophages

We detected a total of 3972 macrophages that formed 2 clusters (Fig. 4A). Although genes upregulated in C2-Macrophage-THBS1 were enriched for signatures of myeloid-derived suppressor cell (MDSC), those in C1-Macrophage-C1QA simultaneously resembled the signatures of tumor-associated macrophage (TAM) and M1 and M2 macrophage (Fig. 4E). The co-existence of M1 and M2 signatures indicated that TAMs were more complex than the classical M1/M2 model, which was consistent with the previous study. A diffusion map of their global transcriptomes showed that C2-Macrophage-THBS1 (MDSC-like macrophage) and C1-Macrophage-C1QA (TAM-like macrophage) formed a continuum but with distinct expression features (Fig. 4G). Specifically, MDSC-like cells highly expressed the *S100A* family genes *FCN1* and *VCAN*, whereas they possessed low levels of HLA-related genes <sup>18</sup>. In contrast, TAM-like cells expressed a set of genes (*APOE*, *C1QA*, *C1QB* and *TREM2*), which is found previously in TAMs of lung cancer. Furthermore, two additional genes, *SLC40A1*, encoding ferroproteins, and *GPNMB*, encoding type I

transmembrane glycoprotein, showed high levels in TAM-like cells. The transcription factors of these two clusters were diverse, that is TAM-like cells preferentially expressed *MITF*, *RUNX2* and *MAF* and MDSC-like cells harbored high levels of *NR4A1*, *RXRA* and *TCF25* (Fig. 4E).

The number of TAM and MDSC isolated from tumor tissues were higher than that of adjacent muscle tissues, demonstrating that TAM and MDSC were preferentially recruited in tumor microenvironment (Fig. 4B-D) <sup>23</sup>. Indeed, macrophages were increased in tumor tissues compared with the adjacent muscle tissues (Fig. 4J). The slight increase of macrophages in tumor tissue illuminates that macrophage immunotherapy may be effective in patients with MFH <sup>25</sup>. Enrichment analysis of up-regulated gene subsets showed that the function of TAM was mainly “regulation of cell migration” and “regulation of cell motility”. And the function of MDSC was primarily “inflammatory response” and “regulation of immune system process” (Fig. 4F).

After defining the macrophage in our dataset, we identified the DEGs in tumor tissues and adjacent muscle tissues of MFH. We authenticated the differential gene set for these macrophages, allowing for a more in-depth analysis of regulatory pathways (Fig. 4K). We focused on the two-fold upregulated or two-fold downregulated genes from macrophage in tumor tissues compared with the adjacent muscle tissues in MFH. Heatmap for the profile of *CTSK*, *MMP9*, *CKB*, *CCL18* and *COL6A2* demonstrated that these genes were upregulated in macrophage of tumor tissues compared with adjacent muscle tissues (Fig. 4I). GO enrichment analysis demonstrated that the majority of the down-regulated genes in the macrophage subsets were related to “inflammatory response” and “regulation of response to external stimulus” (Fig. 4H).

The t-SNE analysis is instrumental in revealing the heterogeneity of distinct macrophage clusters. However, it is possible that the clusters share common differentiation trajectories. Most macrophages were arranged into a major trajectory with two minor bifurcations by pseudotime ordering. Macrophages from different samples widely distributed in the pseudotime space. The macrophages in adjacent muscle tissues occupied the lower part, while macrophages in the tumor tissues were located in the higher position, indicating that the cells in lower part may be the origin of differentiated macrophages. MDSCs were mainly distributed in the left lower branch, while TAM basically occupied the right upper part, which also suggests that the cells in lower part may be the starting point of differentiation of macrophages (Fig. 4L). We speculate that macrophages have the tendency to transform MDSC phenotype to TAM phenotype <sup>25</sup>.

## Two Distinct States of Tumor-Enriched Osteoclasts

Osteoclast plays a vital role in osteolysis and tumor growth in tumor tissues. Based on t-SNE algorithm (Fig. 5A), two individual subclusters of osteoclast were identified with distinct levels of myeloid markers such as *CD74* and/or mature osteoclastic markers such as *CTSK* and *ACP5* <sup>24</sup>. The subcluster described as C1\_progenitor\_OC with high levels of myeloid markers *CD74* and *CD27* and low levels of OC markers *CTSK* and *ACP5*. C2\_mature\_OC expressed high levels of *CTSK* and *ACP5*, and low levels of *CD74* (Fig. 5G).

Both progenitor osteoclasts and mature osteoclasts isolated from tumor tissues were higher than those from adjacent muscle tissues, suggesting that progenitor osteoclasts and mature osteoclasts may be enriched in tumor microenvironment (Fig. 5B, C). Enrichment analysis showed that the functions of C1\_progenitor\_OC were mainly “type I interferon signaling pathway” and “response to cytokine”. The functions of C2\_mature\_OC were mainly “bone resorption” and “regulation of bone resorption” (Fig. 5E).

Osteoclasts are specialized cells derived from the monocyte/macrophage hematopoietic lineage. They develop and adhere to bone matrix and then secrete acid and lytic enzymes to degrade the bone matrix in a specialized extracellular compartment. Increased bone resorption is the result of osteoclast formation induced by tumor cells and osteoclast formation facilitates bone resorption. Bone is a heterogeneous environment that is benefit for the growth of tumor cells. Among the different cell types present in the bone, osteoclasts are crucial players in the so called “vicious cycle”. This phenomenon is triggered by tumor cells and eventually leads to both tumor proliferation as well as bone deregulation, which promotes the development of bone metastasis <sup>26</sup>.

Subsequently, we performed the trajectory analysis of the osteoclasts to infer the osteoclast maturation process in MFH. Thus, those C2\_mature\_OC were highly enriched at the late period of the pseudotime, demonstrating that the osteoclast state transformed from progenitor to mature (Fig. 5D). Then we analyzed the trajectory of macrophages and osteoclasts. MDSCs mainly occupied the left upper branch, while TAMs were primarily located in the right lower part. Both clusters of osteoclasts were concentrated at the end of the lower branch, suggesting that macrophages had the tendency to differentiate into osteoclasts (Fig. 6B) <sup>27,28</sup>. RNA velocity analysis, Monocle 3 and PAGA algorithm confirmed these trajectories (Fig. 6A, C, D).

We analyzed the gene expression patterns involved in osteoclasts and macrophages. Genes related to “regulation of programmed cell death” decreased observably along the quasi-time axis. Genes associated with “mitotic cell cycle process” increased dramatically along the pseudo-time axis. Genes related to “myeloid leukocyte mediated immunity” increased initially and decreased subsequently along the quasi-time axis. (Fig. 6E).

After defining the osteoclasts in our dataset, we recognized the DEGs in tumor tissues and adjacent muscle tissues. We identified the differential gene set for these osteoclasts that allowed a more in-depth analysis of regulatory pathways (Fig. 5F, H). Gene Set Enrichment Analysis (GSEA) demonstrated that osteoclasts in MFH negatively regulated humoral immune response (Fig. 5I). KEGG enrichment analysis showed that most of the up-regulated genes of osteoclasts in MFH were related to “PD-L1 expression and PD-1 checkpoint pathway in cancer” and “T cell receptor signaling pathway” (Fig. 5J).

## Gene Expression Heterogeneity in Fibroblast Subsets was Identified in the MFH

As the important cell component in the disease lesion, Fibroblasts within adjacent muscle tissues and tumor tissues were compared and various DEGs were identified (Fig. 7I). We found abundant DEGs in

tumor and adjacent normal tissues, so we inferred that cancerous fibroblasts exist in fibroblasts<sup>15</sup>. Therefore, fibroblasts were identified by CopyKAT<sup>12,29</sup> and divided into tumor-associated fibroblasts and normal fibroblasts (Fig. 7G). To determine the intrinsic structure and potential functional subtypes of the entire fibroblast population, we performed unsupervised clustering in these two types of cells to examine their heterogeneity (Fig. 7A and Supplementary Fig. S9D). According to the DEGs, normal fibroblasts were divided into 7 clusters (Supplementary Fig. S9E): C1\_normal\_Fibroblast, C2\_normal\_Fibroblast, C3\_normal\_Fibroblast, C4\_normal\_Fibroblast, C5\_normal\_Fibroblast, C6\_normal\_Fibroblast, C7\_normal\_Fibroblast. Tumor-associated fibroblasts are divided into 8 clusters (Fig. 7E, F): C1\_malignant\_Fibroblast, C2\_malignant\_Fibroblast, C3\_malignant\_Fibroblast, C4\_malignant\_Fibroblast, C5\_malignant\_Fibroblast, C6\_malignant\_Fibroblast, C7\_malignant\_Fibroblast and C8\_malignant\_Fibroblast<sup>12,15,30,31</sup>. Besides, we found that normal fibroblasts are mainly distributed in adjacent muscle tissues, while malignant fibroblasts are primarily distributed in tumor tissues (Fig. 7B-D and Supplementary Fig. S9A-C).

Enrichment analysis of up-regulated gene subsets showed that the functions of C1\_malignant\_Fibroblast were principally “blood vessel development” and “cell migration”. The functions of C6\_malignant\_Fibroblast were largely “immune system process”, “leukocyte activation” and “T cell activation” (Fig. 7H). In addition, GO enrichment analyses revealed that within the malignant cells of MFH, fibroblasts were enriched for genes associated with “cell activation” and “inflammatory response” (Fig. 7J).

## Cell Communication Networks in MFH

We used CellphoneDB to predict receptor-ligand interactions. First, we calculated the interactions in the cell types from tumor tissues and adjacent muscle tissues separately. We observed that cells from tumor tissues had more potential for interaction than those from adjacent normal tissues, especially in malignant fibroblasts, osteoclasts, macrophages and several kinds of T cells<sup>20</sup>. Interestingly, we found a cellular communication network between tumor cells and immune cells through immune checkpoint ligand-receptor interactions (Fig. 8A-C).

The pro-apoptotic interaction between C2\_Naïve CD8 + T cells and mature osteoclasts, osteoclast progenitors or C8\_malignant\_Fibroblast (*TNFSF12\_TNFRSF12A*, *TNFRSF1A\_GRN* and *TNFRSF1B\_GRN*) was increased. Mature osteoclasts, osteoclast progenitors or C8\_malignant\_Fibroblast express *NECTIN2* and *NECTIN3*, thereby transmitting inhibitory signals to *TIGIT* on C2\_Naïve CD8 + T cells, respectively. Some new communications have been observed between mature osteoclasts, osteoclast progenitors or C8\_malignant\_Fibroblast and CD8 + T cells in the malignant tissues indicating that T cells were recruited in tumor tissues (*CCL4L2\_VSIR* and *CCL5\_CCR1*). C2\_Naïve CD8 + T cells showed inhibitory interaction with mature osteoclasts or osteoclast progenitors (*HLA-DPA1\_GAL* and *HLA-FAM3c*), or with C8\_malignant\_Fibroblast (*HLA-DPB1\_NRG1*). The pro-apoptotic interactions of C1\_Exhausted CD8 + T cells with mature osteoclasts, osteoclast progenitors or C8\_malignant\_Fibroblast were increased (*TNFSF12\_TNFRSF12A*, *TNFRSF1B\_GRN* and *TNFRSF1A\_GRN*), but T cell-homing communications

weakened (*CCL4\_CCR5*, *CCL5\_CCR5* and *CXCR3\_CCL20*). Mature osteoclasts, osteoclast progenitors and C6\_malignant\_Fibroblast express *NECTIN2* and *FAM3C*, which transmitted inhibitory signals to *TIGIT* and *PDCD1* on C1\_Exhausted CD8 + T cells, respectively. The co-stimulatory interaction between C2\_Exhausted CD4 + Treg and mature osteoclasts or osteoclast progenitors (*TNFRSF1B\_GRN*, *TNFSF12\_TNFRSF12A* and *MIF\_TNFRSF14*) increased. *NECTIN2*, *CD80* and *CD86*, expressed on mature osteoclasts, osteoclast progenitors C5\_malignant\_Fibroblast or C6\_malignant\_Fibroblast, transferred the suppressive signals to *TIGIT* and *CTLA4* on C2\_Exhausted CD4 + Treg, respectively. It is noteworthy that C2\_Exhausted CD4 + Treg possessed relatively high levels of adhesion molecules, including *CD2* and *ICAM3*. The corresponding receptors, including *CD58* and *aLb2* complex, are widely expressed by mature osteoclasts, osteoclast progenitors, C5\_malignant\_Fibroblast and C6\_malignant\_Fibroblast, which can enhance the adhesion and growth of MFH. Some new interactions have been observed between MDSCs and mature osteoclasts or osteoclast progenitors in tumor tissues, suggesting that MDSCs were recruited in tumor tissues (*CCL3L1\_CCR1*, *CXCL2\_DPP4* and *CCL3L1\_DPP4*). Furthermore, angiogenic signals increased between MDSCs and mature osteoclasts, osteoclast progenitors or C8\_malignant\_Fibroblast (*NRP2\_VEGFA*, *NRP1\_VEGFA* and *NRP1\_VEGFB*). In addition, angiogenic signals (*NRP1\_VEGFB*, *IGF1\_IGF1R*, *NRP1\_VEGFA* and *NRP2\_VEGFA*) and costimulatory effect (*TNFRSF1A\_GRN* and *TNFRSF1B\_GRN*) increased between TAMs and mature osteoclasts, osteoclast progenitors or C8 malignant fibroblasts (Fig. 8D-G) <sup>7,15,32</sup>.

## Discussion

Here, we provide a comprehensive analysis of scRNA-seq data generated from MFH. The deep transcriptome for 18433 individual cells provided an extensive resource for understanding the multi-dimensional characterization of MFH, especially in the tumor immune microenvironment. The higher resolution provided by our dataset was exemplified by the identification of 19 large subsets as well as unique subpopulations such as CD8 + exhausted T cells and CD4 + exhausted Treg <sup>33</sup>. The high quantity and quality of single-cell data allows us to map their developmental trajectory. Besides, the pseudotime analysis and RNA velocity analysis permits us not only to confirm their differentiated relationships in various clusters, but also to deduce their activation status in the cancer microenvironment.

The infiltration of T cells and their characteristics are usually associated with prognostic outcomes. In our study, we determined that *PDCD1*, *CTLA4* and *TIGIT* are involved in the depletion of CD8 + and CD4 + T cells in MFH. Their receptors, such as *FAM3C*, *CD80/CD86* and *NECTIN2*, are widely expressed in malignant fibroblasts, which may be the cause of poor prognosis <sup>17</sup>. In fact, antibody blockade of the PD1 pathway has been shown to reinvigorate exhausted CD8 + T cells with *PDCD1* expression. We discovered that *LAYN* was highly expressed in exhausted CD4 + Treg cells, which was speculated to be associated with CD4 + Treg depletion and poor prognosis. Previous study revealed a regulatory role of *LAYN* in Treg function. Similarly, some studies have proved the induction of *LAYN* after activation of exhausted CD8 + T cells and CD4 + Tregs <sup>21</sup>. Furthermore, the overexpression of *LAYN* on CD8 + T cells in human blood leads to a significant reduction in the production of IFN-g, which is a key cytokine involved

in tumor killing activity and supports *LAYN* as a negative regulator. TCGA data showed that high *LAYN* levels were associated with poor prognosis of various cancers. More studies are needed to further investigate the function of *LAYN* and other genes related to depleted CD4 + Treg cells in MFH.

In addition to CD8 + effector and CD8 + cytotoxic T cells, we also identified CD4 + cytotoxic T cells, which shared similar gene expression characteristics to CD8 + effector and CD8 + cytotoxic T cells, implicating their cytotoxic functions. These cells are more closely linked to CD4 + effector Th cells, which appear to be in a mixed state among known subtypes of T helper cells but are more similar to the effector cells due to the expression of *GZMA*, *GZMK*, *NKG7* and *GNLY*. Therefore, promoting these cells to more closely resemble cytolytic CD4 + T cells might be a new strategy for MFH immunotherapy <sup>34</sup>.

MFH is an aggressive form of soft tissue sarcoma. MFH is extremely rare, with limited information on its pathogenesis, clinical and radiological features, pathological findings, and therapeutic outcomes. Pathophysiology remains elusive and treatment options are limited <sup>2</sup>. Advances in scRNA-seq technology enabled comprehensive analysis of the immune system in an unbiased way at a single cell level. To our knowledge, this study is the first to analyze MFH using scRNA-seq technology <sup>10</sup>. Our single-cell database reveals the detailed characteristics of MFH-infiltrating cells in the microenvironment, including their aggregation, dynamics and developmental trajectory, as well as unique characteristics, in tumor tissues and the corresponding adjacent normal tissues. In Conclusion, we reported the first case of MFH identified by scRNA-seq, described the characteristics of tumor microenvironment, recognized the markers and particular clusters regarding cancer immunotherapy, and determined the treatment strategy. We found that malignant fibroblasts and osteoclasts are significantly enriched in tumor tissues, and most of them are undergoing an active EMT process, which lead to cancer invasion and metastasis. The transition from TAMs to osteoclasts also promotes tumor invasion and metastasis. We detected high levels of CD8 + exhausted T cells accumulate in MFH with abundant aggregation at the late period of pseudotime. In our study, the up-regulation of *PDCD1*, *CTLA4*, *TIGIT* and *LAYN* suggests that inhibitors of these biomarkers may be effective against MFH. At present, there are no guidelines for the use of ICIs in MFH. We confirmed that PD-1 ICIs can be used as first-line treatment for patient with MFH.

## Methods

### Human specimens

The present study has been reviewed and approved by SIR RUN RUN HOSPITAL NANJING MEDICAL UNIVERSITY, informed written consent was obtained from the patient for publication of this report and any accompanying images. The authors have read the CARE Checklist (2016), and the manuscript was prepared and revised according to the CARE Checklist (2016).

### Sample preparation and scRNA-seq

Fresh lesions were stored in tissue preservation solution and processed on ice after the surgery within 30 mins. Single-cell suspensions of the collected tissues were prepared through mechanical dissociation

and enzymatic digestion within 16 h after surgery. Briefly, tissues were cut into pieces that were 2–4 mm in size and transferred to a tube containing the enzyme mix. The tissues were incubated in an enzyme solution (collagenase, DNase I, and Dispase II; prepared in DMEM) at 37 °C for 1 h. The tissue pieces were remixed by gentle pipetting at 20 min intervals during incubation. Each cell suspension was transferred to a new 50 ml (15 ml tube for biopsy samples) tube after being passed through a 70 µm strainer. The volume in the tube was readjusted to 50 ml (or 15 ml) with DMEM medium, and the contents were centrifuged to remove the enzymes. The supernatant was aspirated, the cell pellet was resuspended in 4 ml of DMEM medium, and the dead cells were removed using Ficoll-Paque Plus (GE Healthcare, Chicago, IL, USA) separation. Single cells were encapsulated in droplets using 10 × Genomics GemCode Technology and processed according to the manufacturer's instruction.

## Quality control metrics and data processing

Using CellRanger, reads were compared with the reference genome and annotated as specific genes. After UMI is corrected and counted, the unfiltered feature-barcode matrix is obtained. According to the unfiltered feature-barcode matrix, CellRanger identifies and distinguishes cells and non-cells in the data, and draws a rank-plot to intuitively reflect the effective cell identification results. The cell filtration of Cellranger is automatically identified according to the amount of gene expression, and some abnormal cells will remain, so it is necessary to further filter abnormal cells before subgroup classification. The GEM containing multiple cells in each sample was first detected Doublet Finder is used to calculate the probability of GEM multicellular (pANN value), and then the multicellular rate of each sample is calculated based on the relationship between the effective cell number given by 10X (after Cellranger filtering) and the multicellular rate, and the multicellular filtering threshold of each sample is determined, and the multicellular filtering is carried out in turn. Number of genes identified in a single cell (340.0-6800.0) For the same type of cell, the number of expressed genes is generally maintained within a certain range. If the value is too high, it may be that multiple cell types are wrapped in a GEM, so the barcode is eliminated. Total number of UMI in a single cell (< 44000.0) The total number of mRNA that can exist in a single cell is limited. If the total number of UMI is too high, two or more cells may enter the same GEM, thus eliminating such cells. Mitochondrial gene expression quantity proportion in the single cell apoptosis (< 25.0%) is usually accompanied by a high expression of mitochondrial gene. Mitochondrial genes so high expression means that the poor state of cells, the cells in the process of experiment by bad stimulation, for subsequent analysis by real cells, thus eliminating such cells.

## Histological analysis

For histopathological analysis, hematoxylin eosin (H&E) staining was performed on formaldehyde fixed, paraffin-embedded tissue samples. Microscopic analysis of the staining was evaluated by examining five sections from each tissue. The sections were observed using a laser scanning confocal microscope.

## Immunohistochemistry (IHC)

Tissue sectioning and immunohistochemistry staining of formalin-fixed, paraffin-embedded MFH specimens were performed. All sections were deparaffinized, rehydrated, and washed. Endogenous

peroxidase was blocked using 3% hydrogen peroxide for 10 min. After water-bath heating for antigen retrieval, slides were incubated with primary antibodies followed by horseradish peroxidase (HRP)-linked secondary antibodies and diaminobenzidine staining (G1213-100UL, G1214-100UL). Hospital Pathology Department blinded to clinical data independently assessed staining results for SMA (unnecessary dilution, ZM-0003, ZSGB-BIO), CD68 (unnecessary dilution, ZM-0464, ZSGB-BIO), Ki67 (1:100, ZM-0166, ZSGB-BIO), CD34 (unnecessary dilution, ZM-0046, ZSGB-BIO), Desmin (unnecessary dilution, ZA-0686, ZSGB-BIO), EMA (unnecessary dilution, ZM-0095, ZSGB-BIO), S-100 (unnecessary dilution, ZM-0224, ZSGB-BIO), BCL-2 (unnecessary dilution, ZA-0536, ZSGB-BIO). Servicebio blinded to clinical data independently assessed staining results for CD3 (1:250, GB11014, Servicebio), CD4 (1:400, GB11064-1, Servicebio), CD8 (1:200, GB11068-1, Servicebio). Quantification was performed by counting positive cells in 6 to 10 high-powered fields (magnification,  $\times 40$ ) in a blinded fashion.

## **Analysis of differentially expressed genes (DEGs) in tumor tissues and adjacent muscle tissues**

We analyzed the tumor tissues and adjacent muscle tissues of a patient with MFH to find DEGs for a group in one cluster. We identified the DEGs based on the following criteria:  $|\log FC| \geq 0.25$ ,  $p\_value\_adj \leq 0.05$  and the percentage of cells where the gene is detected in specific cluster is more than 25%. We used the gene ontology (GO) enrichment analysis to discover certain biological functions in each cell type. Initially, all peak related genes were mapped to GO terms in the Gene Ontology database; gene numbers were calculated for every term and significantly enriched GO terms in the peak related genes versus to the genome background were defined by hypergeometric test.

## **Pseudotime Analysis**

Single cell trajectory was analyzed using matrix of cells and gene expressions by PAGA<sup>19</sup> and Monocle. The PAGA graph was made using the preprocessed Seurat object. PAGA achieves consistent and topology-preserving embeddings by initializing an embedding of a fine-grained graph using the coordinates of a coarse-grained graph. After assigning the starting cell, the software automatically calculated the pseudo-time value of each cell by referring to the DPT algorithm. The Rfast2 software package was used to perform correlation statistics and significance calculations for each gene. And we identified significantly differential expression genes with  $FDR < 1e-7$  on the pseudo-time axis. At the same time, we grouped genes which had the similar trends in expression. Monocle reduced the space down to one with two dimensions and ordered the cells ( $\sigma = 0.001$ ,  $\lambda = \text{NULL}$ ,  $\text{param.gamma} = 10$ ,  $\text{tol} = 0.001$ ). We identified the key genes related to the development and differentiation process with  $FDR < 1e-5$  and grouped the genes with similar trends in expression reasoning that such groups may share common biological functions and regulators. Monocle developed BEAM to test for branch-dependent gene expression by formulating the problem as a contrast between two negative binomial GLMs. Differential gene testing for the pseudo-time analysis was based on the previously identified cell clusters.

## **RNA velocity-based cell fate tracing**

To perform the RNA velocity analysis, the spliced reads and unspliced reads were recounted by the velocity python package based on previous aligned bam files of scRNA-seq data. The calculation of RNA velocity values for each gene in each cell and embedding RNA velocity vector to low-dimension space were done by following the scvelo python pipeline. We calculated the velocity-based cell transition matrix by `transition_matrix()` function from scvelo, of which the element was the Pearson correlation coefficient between the velocity vector and cell state difference vectors of the column cell as previously described. We estimated the destination of a cell by identifying the highest correlation value. Then Fisher's exact test was performed on 2x2 cluster-by-cluster or cluster-by-tissue contingency tables to test the fate destinations of interested cell clusters. To infer the migration directions of T cell, osteoclast and macrophage, we first constructed partition-based graph abstraction for T cell population, osteoclast population and macrophage population respectively, and then oriented edges among cell populations using RNA velocity information as previously described.

## Copy number karyotyping of aneuploid tumors (CopyKAT)

To distinguish malignant cells from normal cells in fibroblasts, we chose CopyKAT (copynumber karyotyping of tumors) software to calculate the CNV level of each cell<sup>29</sup>. CopyKAT is a software that combines integrative Bayesian method and hierarchical clustering to classify cells according to copy number. We use scRNAseq technology to get the gene expression data of fibroblasts in adjacent muscle tissues and tumor tissues. The gene expression matrix of unique molecular identifier (UMI) of two fibroblasts from different sources is the input of CopyKAT. The software cluster the UMI data after processing, and select the diploid cells with high confidence first. Then, using hierarchical clustering, tumor cells with significant difference from normal cells are obtained. For non-significant genomes, Gaussian mixture model (GMM) is used to identify them one by one. Finally, get the gene expression profiles of malignant cells and normal cells.

## Cell-to-cell communication of scRNA-seq data

The CellPhoneDB algorithm ([https://github.com/ Teichlab /cellphonedb; v2.1.1](https://github.com/Teichlab/cellphonedb)) was used to infer cell-to-cell interactions. Briefly, the algorithm allows the detection of ligand–receptor interactions between cell types in scRNA-seq data using the statistical framework described in refs. We took the union of the significant interactions found in tumor tissues and adjacent muscle tissues to explore specific interactions. Next, we assessed the amount of interactions that are shared and specific for tumor tissues and adjacent muscle tissues, and explored specific interactions indicated as curated (that is, annotated by the CellPhoneDB developers).

## Statistical analysis

Statistical analyses were performed using GraphPad Prism. The data are expressed as the mean  $\pm$  SEM unless indicated otherwise. Unpaired Student's t-test was used to determine statistically significant differences. A value of  $P < 0.05$  was considered significant at the 95% confidence level. Date analysis was performed using the OmicShare tools, a free online platform for data analysis.

### Data availability

The 10 × Genomics datasets generated during this study is publicly available via Genome Sequence Archive (GSA), under the BioProject number: [PRJCA008549](https://www.genome.gov/PRJCA008549).

### Code availability

The codes generated during this study are available at the OmicShare tools, a free online platform (<https://www.omicshare.com/>).

## Declarations

### Competing interests

The authors declare that they have no competing interests.

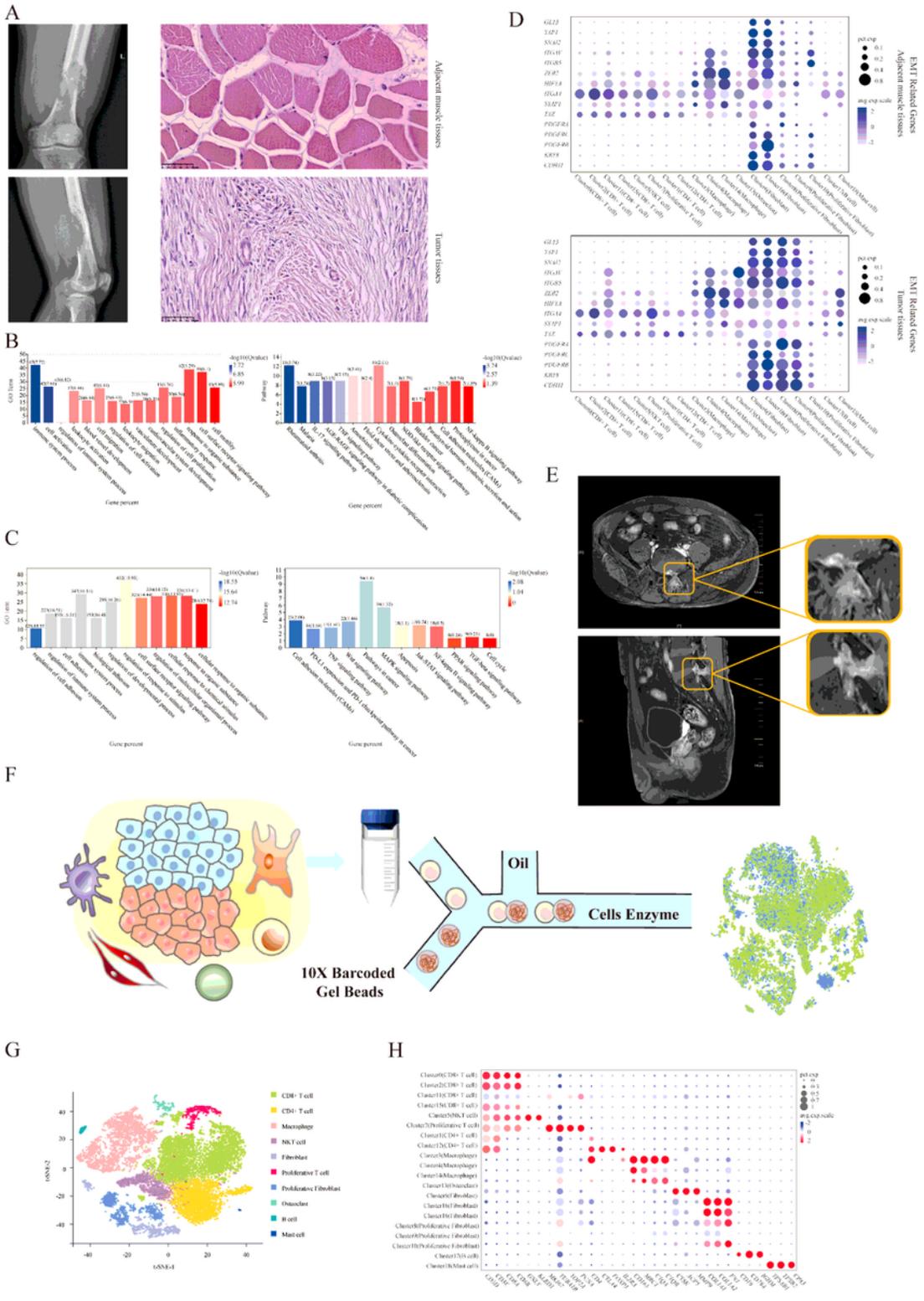
## References

1. Weiss, S.W. & Enzinger, F.M. Malignant fibrous histiocytoma: an analysis of 200 cases. *Cancer* **41**, 2250–2266 (1978).
2. Malik, A.T., *et al.* Malignant fibrous histiocytoma of bone: A survival analysis from the National Cancer Database. *J Surg Oncol* **121**, 1097–1103 (2020).
3. Peiper, M., Zurakowski, D., Knoefel, W.T. & Izbicki, J.R. Malignant fibrous histiocytoma of the extremities and trunk: An institutional review. *Surgery* **135**, 59–66 (2004).
4. Liu, B., *et al.* Clinicopathological characteristics and survival of malignant fibrous histiocytoma of the bone: A population-based study using the SEER database. *PLoS One* **15**, e0232466 (2020).
5. Navin, N., *et al.* Tumour evolution inferred by single-cell sequencing. *Nature* **472**, 90–94 (2011).
6. Lambrechts, D., *et al.* Phenotype molding of stromal cells in the lung tumor microenvironment. *Nat Med* **24**, 1277–1289 (2018).
7. Kim, N., *et al.* Single-cell RNA sequencing demonstrates the molecular and cellular reprogramming of metastatic lung adenocarcinoma. *Nat Commun* **11**, 2285 (2020).
8. Guo, X., *et al.* Global characterization of T cells in non-small-cell lung cancer by single-cell sequencing. *Nat Med* **24**, 978–985 (2018).
9. Lee, H.W., *et al.* Single-cell RNA sequencing reveals the tumor microenvironment and facilitates strategic choices to circumvent treatment failure in a chemorefractory bladder cancer patient. *Genome Med* **12**, 47 (2020).
10. Kim, D., *et al.* Targeted therapy guided by single-cell transcriptomic analysis in drug-induced hypersensitivity syndrome: a case report. *Nat Med* **26**, 236–243 (2020).
11. Zheng, C., *et al.* Landscape of Infiltrating T Cells in Liver Cancer Revealed by Single-Cell Sequencing. *Cell* **169**, 1342–1356 e1316 (2017).
12. Zhang, M., *et al.* Dissecting transcriptional heterogeneity in primary gastric adenocarcinoma by single cell RNA sequencing. *Gut* **70**, 464–475 (2021).

13. Tirosh, I., *et al.* Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189–196 (2016).
14. Savas, P., *et al.* Single-cell profiling of breast cancer T cells reveals a tissue-resident memory subset associated with improved prognosis. *Nat Med* **24**, 986–993 (2018).
15. Ma, J., *et al.* Single-cell transcriptomic analysis of endometriosis provides insights into fibroblast fates and immune cell heterogeneity. *Cell Biosci* **11**, 125 (2021).
16. Cheng, A.C., Lin, Y.J., Chiu, S.H. & Shih, Y.L. Combined immune checkpoint inhibitors of CTLA4 and PD-1 for hepatic melanoma of unknown primary origin: A case report. *World J Clin Cases* **9**, 2641–2648 (2021).
17. Andrews, L.P., Yano, H. & Vignali, D.A.A. Inhibitory receptors and ligands beyond PD-1, PD-L1 and CTLA-4: breakthroughs or backups. *Nat Immunol* **20**, 1425–1434 (2019).
18. Zhang, Q., *et al.* Landscape and Dynamics of Single Immune Cells in Hepatocellular Carcinoma. *Cell* **179**, 829–845 e820 (2019).
19. Wolf, F.A., *et al.* PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome Biol* **20**, 59 (2019).
20. Bassez, A., *et al.* A single-cell map of intratumoral changes during anti-PD1 treatment of patients with breast cancer. *Nat Med* **27**, 820–832 (2021).
21. De Simone, M., *et al.* Transcriptional Landscape of Human Tissue Lymphocytes Unveils Uniqueness of Tumor-Infiltrating T Regulatory Cells. *Immunity* **45**, 1135–1147 (2016).
22. Ma, X., *et al.* Identification of a distinct luminal subgroup diagnosing and stratifying early stage prostate cancer by tissue-based single-cell RNA sequencing. *Mol Cancer* **19**, 147 (2020).
23. Qie, C., *et al.* Single-cell RNA-Seq reveals the transcriptional landscape and heterogeneity of skin macrophages in *Vsir(-/-)* murine psoriasis. *Theranostics* **10**, 10483–10497 (2020).
24. Zhou, Y., *et al.* Single-cell RNA landscape of intratumoral heterogeneity and immunosuppressive microenvironment in advanced osteosarcoma. *Nat Commun* **11**, 6322 (2020).
25. Trac, N.T. & Chung, E.J. Peptide-based targeting of immunosuppressive cells in cancer. *Bioact Mater* **5**, 92–101 (2020).
26. Kamran, N., Chandran, M., Lowenstein, P.R. & Castro, M.G. Immature myeloid cells in the tumor microenvironment: Implications for immunotherapy. *Clin Immunol* **189**, 34–42 (2018).
27. Takahashi, N., Ejiri, S., Yanagisawa, S. & Ozawa, H. Regulation of osteoclast polarization. *Odontology* **95**, 1–9 (2007).
28. Mutsaers, A.J. & Walkley, C.R. Cells of origin in osteosarcoma: mesenchymal stem cells or osteoblast committed cells? *Bone* **62**, 56–63 (2014).
29. Gao, R., *et al.* Delineating copy number and clonal substructure in human tumors from single-cell transcriptomes. *Nat Biotechnol* **39**, 599–608 (2021).
30. Zhao, J., *et al.* Single cell RNA-seq reveals the landscape of tumor and infiltrating immune cells in nasopharyngeal carcinoma. *Cancer Lett* **477**, 131–143 (2020).

31. Sebastian, A., *et al.* Single-Cell Transcriptomic Analysis of Tumor-Derived Fibroblasts and Normal Tissue-Resident Fibroblasts Reveals Fibroblast Heterogeneity in Breast Cancer. *Cancers (Basel)* **12**(2020).
32. Zhang, J., *et al.* Two waves of pro-inflammatory factors are released during the influenza A virus (IAV)-driven pulmonary immunopathogenesis. *PLoS Pathog* **16**, e1008334 (2020).
33. Pauken, K.E. & Wherry, E.J. Overcoming T cell exhaustion in infection and cancer. *Trends Immunol* **36**, 265–276 (2015).
34. Saeidi, A., *et al.* T-Cell Exhaustion in Chronic Infections: Reversing the State of Exhaustion and Reinvigorating Optimal Protective Immune Responses. *Front Immunol* **9**, 2569 (2018).

## Figures

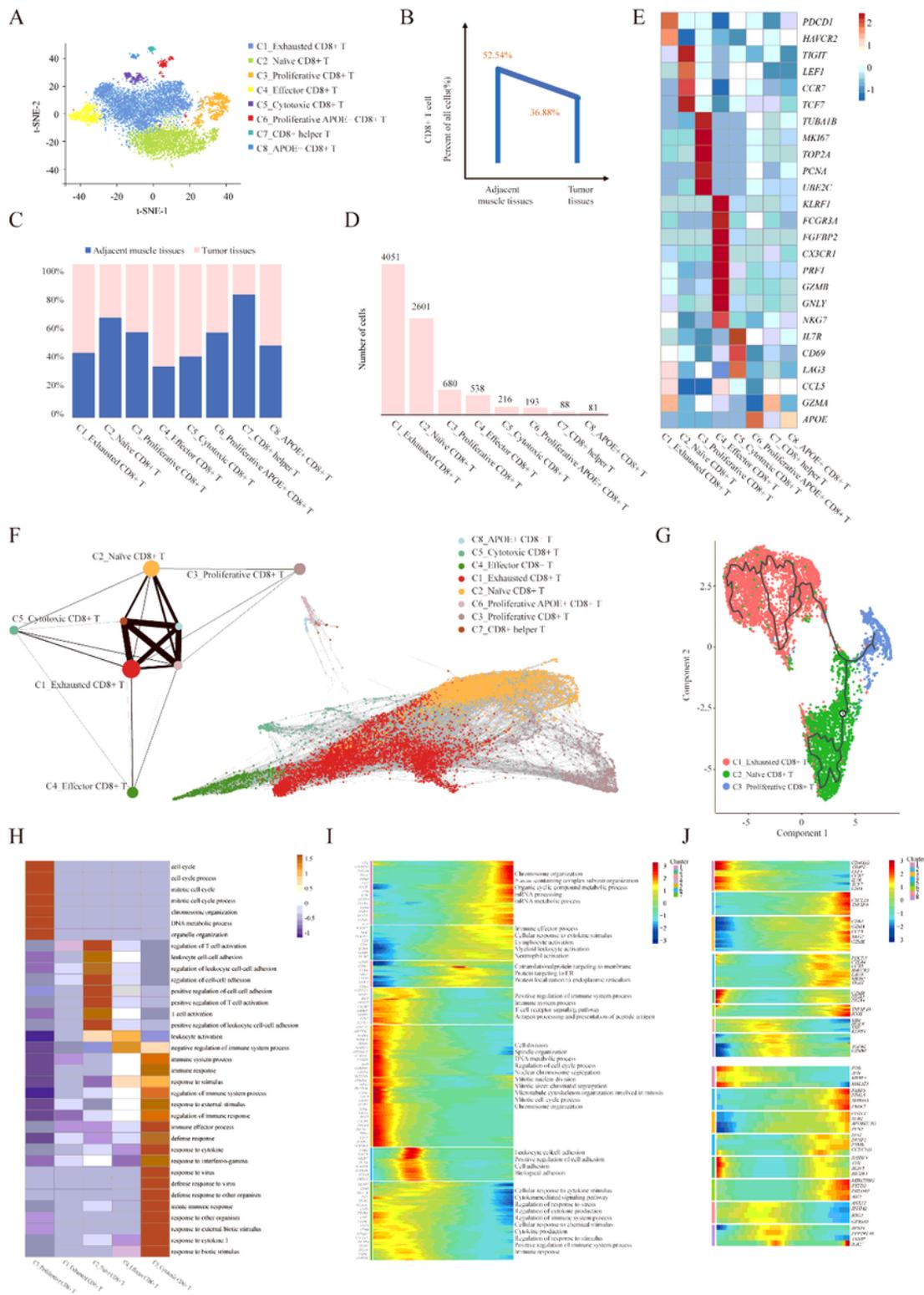


**Figure 1**

**Single-cell transcriptomic analysis of MFH.**

**A.** Bone destruction of the left lower femur with a soft tissue mass (upper left: Anteroposterior X-ray; lower left: Lateral X-ray film). H&E staining of the tumor tissues and adjacent muscle tissues (Scale bar 50µm).  
**B.** GO and KEGG enrichment analysis was performed for DEGs in MFH between tumor tissues and

adjacent muscle tissues. **C.** GO and KEGG enrichment analysis were performed for DEGs in non-immune cells between tumor tissues and adjacent muscle tissues. **D.** Expression of EMT related genes were shown for each cluster of adjacent muscle tissue and tumor tissue. Dot size corresponded to the percentage of cells in the cluster expressing a gene, and dot color corresponded to the average expression level for the gene in the cluster. **E.** MR medical impact (upper: pelvic plain scan; left: femoral plain scan). **F.** Schematic diagram of the experimental design. **G.** t-SNE plot of cells from tumor tissues and adjacent muscle tissues in MFH, with each cell color-coded according to its associated cell type. **H.** Dot plot showing the 30 signature gene expressions across the 10 cellular clusters. The size of dots represents the proportion of cells expressing the particular marker, and the spectrum of color indicates the mean expression levels of the markers.

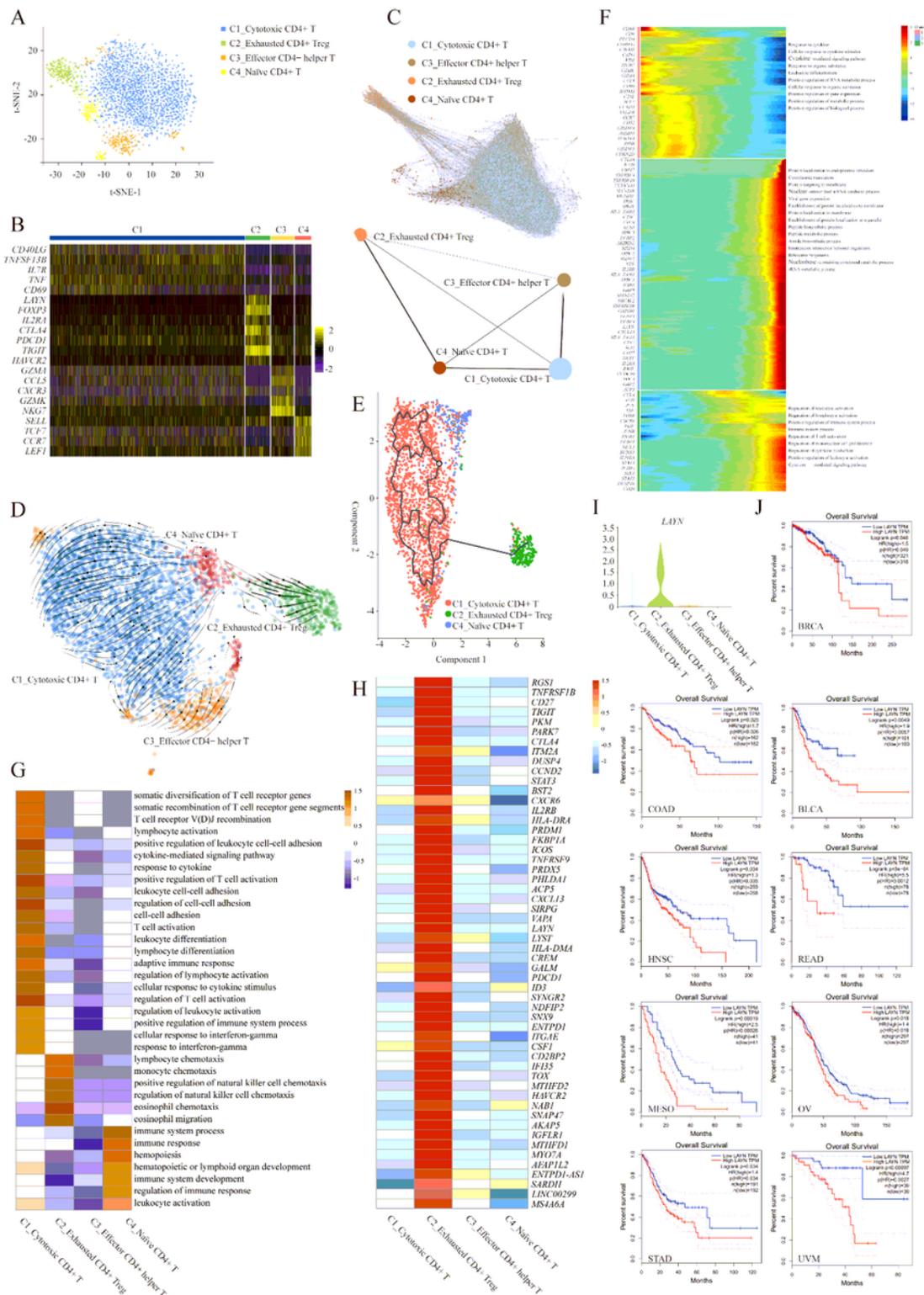


**Figure 2**

**Gene expression heterogeneity of CD8+ T cell subsets was identified in the MFH.**

**A.** t-SNE plots of CD8+ T cells from tumor tissues and adjacent muscle tissues in MFH. **B.** The percentages of CD8+ T cells in adjacent muscle tissues and tumor tissues. **C-D.** The cell number and proportion of each CD8+ T cell cluster. **E.** Heatmap plot of specific marker genes in each CD8+ T cell

cluster. **F.** PAGA analysis on CD8+ T cell phenotypes independently confirming the maturation trajectories. **G.** Trajectories of C2\_Naïve CD8+ T, C3\_Proliferative CD8+ T and C1\_Exhausted CD8+ T using Monocle 3 algorithms. **H.** Functional enrichment analysis of upregulated genes in each CD8 + T cell cluster was performed with GO analysis. Representative significantly enriched function processes are shown. **I.** The DEGs (in rows,  $q\text{-value} < 10^{-10}$ ) in CD8+ T cells along the pseudotime were hierarchically clustered into different subclusters. The top annotated GO terms in each cluster were provided. **J.** Gene expression dynamics along the C1\_Exhausted CD8+ T trajectory. Genes cluster into different gene sets, each characterized by specific expression profiles (upper). For each gene cluster (indicated by different colors), expression of some novel genes along the trajectory is shown (lower).

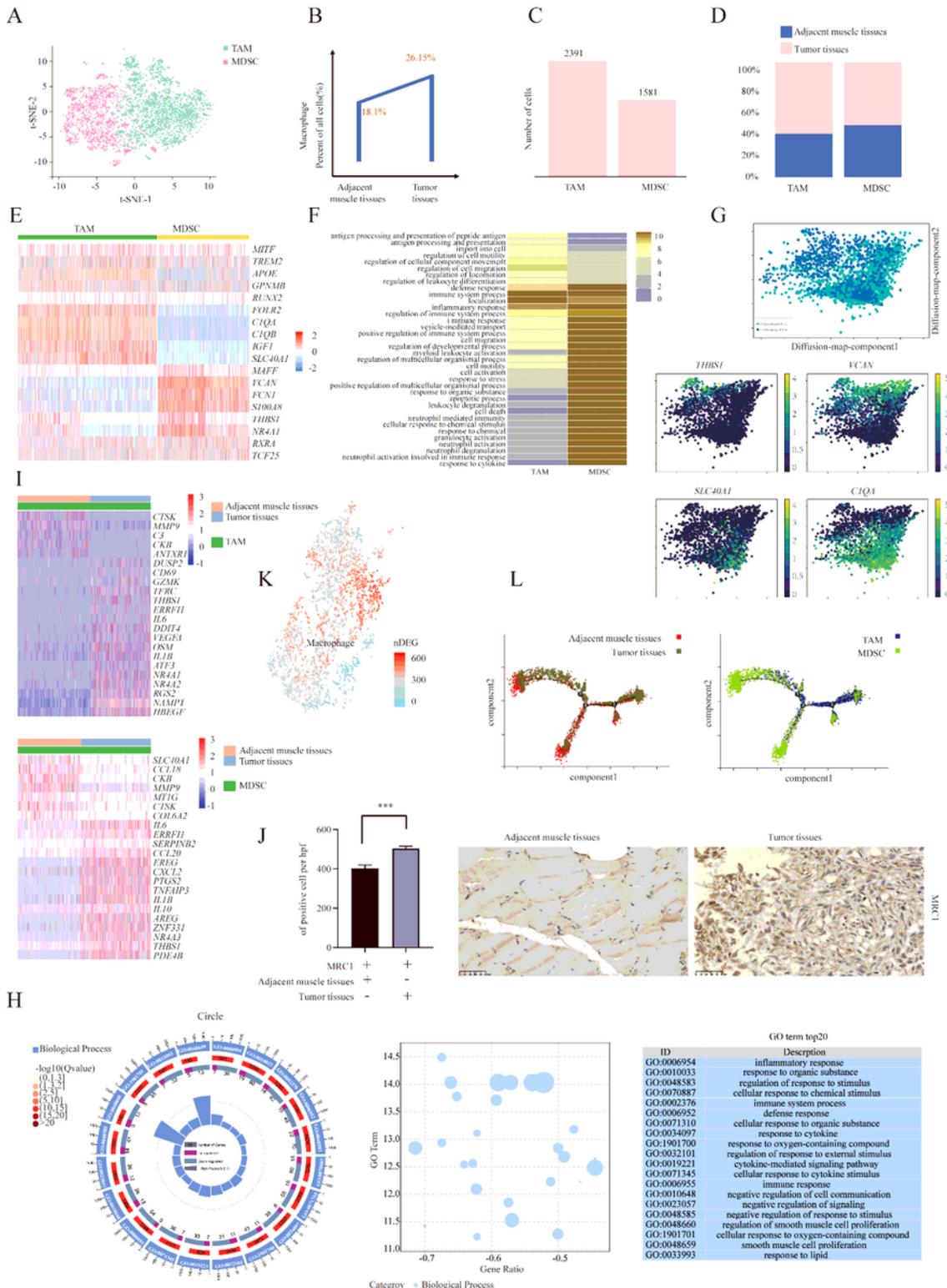


**Figure 3**

Gene expression heterogeneity of CD4+ T cell subsets was identified in the MFH.

**A.** t-SNE plots of CD4+ T cells from tumor tissues and adjacent muscle tissues in MFH. **B.** Heatmap plot of specific marker genes in each CD4+ T cell cluster. **C-D.** PAGA and RNA velocity analysis on CD4+ T cell phenotypes independently confirming the maturation trajectories. **E.** Trajectories of C1\_Cytotoxic CD4+ T,

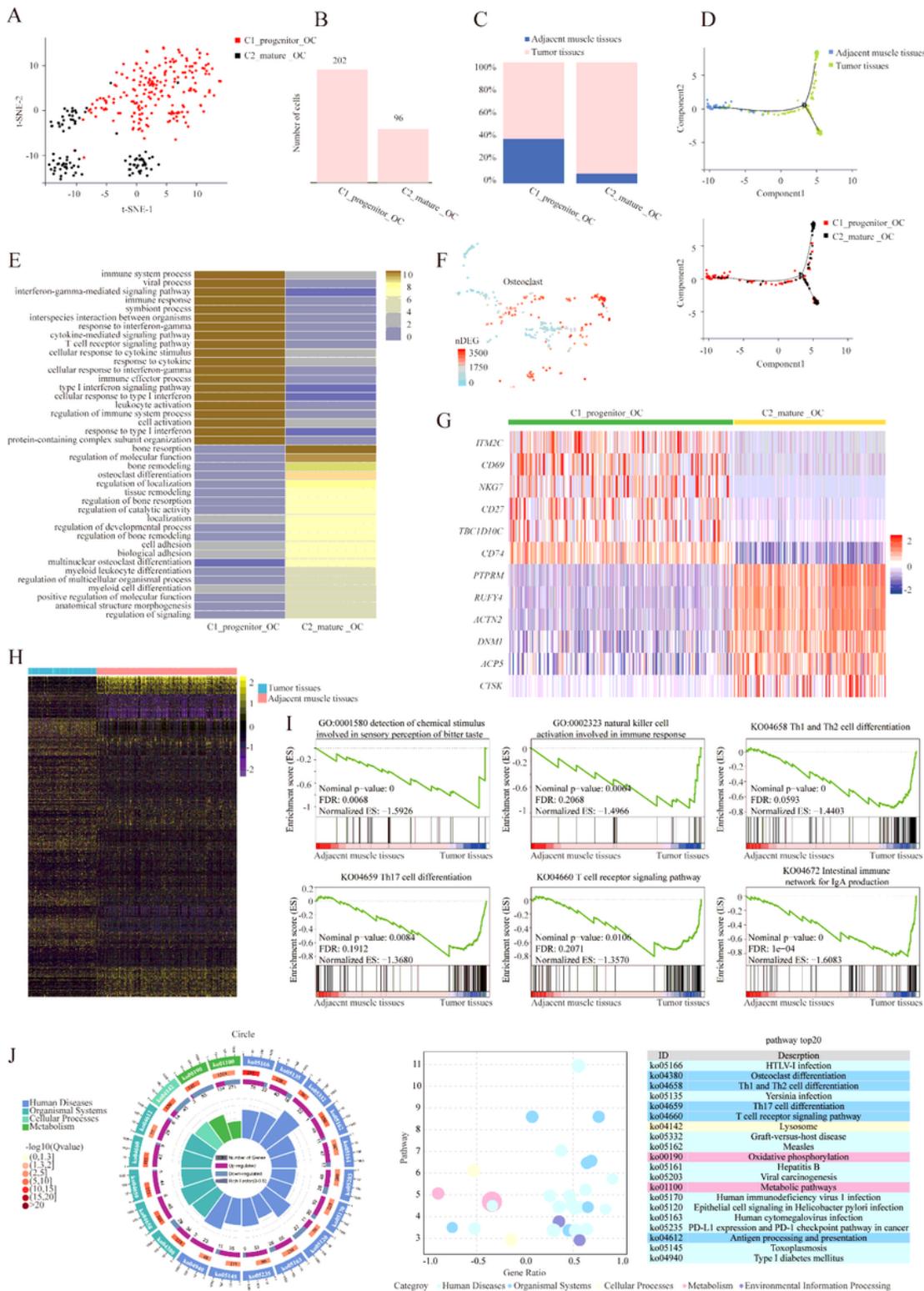
C2\_Exhausted CD4+ Treg and C4\_Naive CD4+ T using Monocle 3 algorithms. **F.** The DEGs in CD4+ T cells along the pseudotime were hierarchically clustered into different subclusters. The top annotated GO terms in each cluster were provided. **G.** Functional enrichment analysis of upregulated genes in each CD4+ T cell cluster was performed with GO analysis. **H.** Heatmap plots comparing the expression of exhaustion-related genes in CD4+ T cell cluster. **I.** Violin plot comparing the expression of LAYN in CD4+ T cell cluster. **J.** The disease-free survival curve based on TCGA data showing patients with higher expression of APOE had poor prognosis.



## Figure 4

### Two Distinct States of Tumor-Enriched Macrophages.

**A.** t-SNE plots of Macrophages from tumor tissues and adjacent muscle tissues in MFH. **B.** The percentages of Macrophages in adjacent muscle tissues and tumor tissues. **C-D.** The cell number and proportion of each macrophage cluster. **E.** Heatmap plot of specific marker genes in each Macrophage cluster. **F.** Functional enrichment analysis of upregulated genes in each Macrophage cluster was performed with GO analysis. **G.** Diffusion map showing the continuous connection of the two macrophage states (upper) and signature gene expression based on 10x (lower). **H.** GO term analysis of DEGs in Macrophage of tumor tissues versus adjacent muscle tissues was performed. **I.** DEGs of TAMs and MDSCs in tumor tissues vs adjacent muscle tissues in MFH was analyzed. **J.** Immunohistochemistry of MFH show the expression of MRC1 in adjacent muscle tissues and tumor tissues. **K.** Number of DEGs between adjacent normal tissues and tumor tissues within Macrophages projected onto the t-SNE map. **L.** Monocle 2 analysis on Macrophage phenotype independently confirming the maturation trajectories.



**Figure 5**

## Two Distinct States of Tumor-Enriched Osteoclasts.

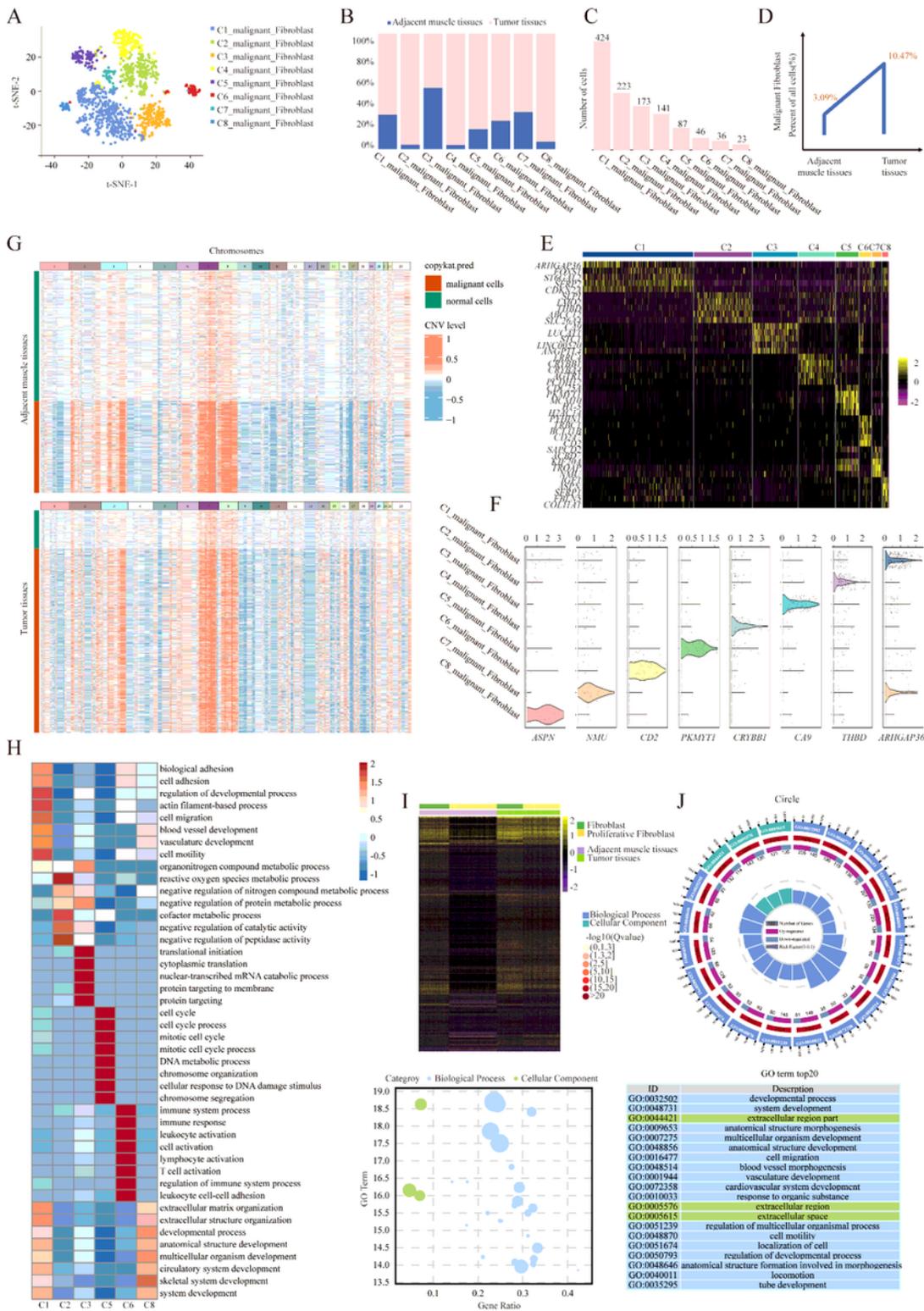
**A.** t-SNE plots of Osteoclasts from tumor tissues and adjacent muscle tissues in MFH. **B-C.** The cell number and proportion of each osteoclast cluster. **D.** Monocle 2 analysis on Osteoclast phenotype independently confirming the maturation trajectories. **E.** Functional enrichment analysis of upregulated



## Figure 6

### The trajectory analysis of osteoclasts and macrophages.

**A.** RNA velocity analysis, **B.** Monocle 2 analysis, **C.** PAGA analysis and **D.** Monocle 3 analysis on Osteoclast and macrophage phenotypes independently confirming the maturation trajectories. **E.** The DEGs in Osteoclast and macrophage along the pseudotime were hierarchically clustered into different subclusters. The top annotated GO terms in each cluster were provided.



**Figure 7**

Gene expression heterogeneity of malignant fibroblast subsets was identified in the MFH.

**A.** t-SNE plots of malignant fibroblasts from tumor tissues and adjacent muscle tissues in MFH. **B-C.** The cell number and proportion of each malignant fibroblast cluster. **D.** The percentages of malignant fibroblasts in adjacent muscle tissues and tumor tissues. **E-F.** Heatmap plots and Violin plots of specific



## Figure 8

### Cell Communication Networks in MFH.

**A-B.** Heatmap of the number of ligand-receptor pairs each cell has (each row/column represents a cell, and the number of ligand-receptor pairs each cell has is colored, the bluer the cell has fewer ligand receptor pairs, and the redder the cell has more ligand receptor pairs). **C.** Results of the ratio of the number of ligand receptor pairs in tumor tissue compared with adjacent normal muscle tissue. **D-H.** Overview of selected ligand–receptor interactions of macrophages, Osteoclast, malignant fibroblast cells and T cells. P values are indicated by circle size, Mean values are indicated by circle color.

## Supplementary Files

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

- [AdditionalfileSupplementaryFigures.docx](#)
- [rs.pdf](#)
- [nreditorialpolicychecklist.pdf](#)
- [SummaryandSoftwareList.zip](#)