SPP1+ macrophages: A malignant macrophage subset in the colorectal cancer microenvironment revealed by single-cell sequencing

Jianyong Zheng (zhjy68@163.com)
Fourth Military Medical University

Zhenyu Xie
Fourth Military Medical University

Liaoran Niu
Fourth Military Medical University

Gaozan Zheng
Fourth Military Medical University

Ruikai Li
Fourth Military Medical University

Hanjun Dan
Fourth Military Medical University

Lili Duan
Fourth Military Medical University

Hongze Wu
Fourth Military Medical University

Guangming Ren
Xi’an Medical University

Xinyu Dou
Xi’an Medical University

Fan Feng
Fourth Military Medical University

Jian Zhang
School of Basic Medicine, Fourth Military Medical University  https://orcid.org/0000-0001-9129-9092

Article

Keywords:

Posted Date: September 7th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1984416/v1
Abstract

Accumulating single-cell studies suggest that SPP1 + macrophages are key players in the tumor microenvironment. However, a systematic investigation of SPP1 + macrophages in colorectal cancer (CRC) has not been conducted. A total of eight single-cell RNA-seq datasets and 16 bulk RNA-seq datasets were included in this study. On the basis of existing research, we propose the SPP1 + macrophage model paradigm, which can explain the clinical features and functional changes of macrophages in CRC better than the M1/M2 polarization theory. We identified four macrophage subsets from CRC myeloid cells: FCN1 + macrophages, C1QC + macrophages, SPP1 + macrophages, and MKI67 + macrophages. Inflammation, phagocytosis, malignancy, and proliferation were identified as the most prominent features related to each of the four macrophage subsets. Our results show that SPP1 + macrophages can serve as markers of CRC occurrence, progression, metastasis and a poor prognosis and exhibit enhanced transcription of genes associated with angiogenesis, epithelial-mesenchymal transition, glycolysis, hypoxia, and immunosuppressive signatures. CellPhoneDB analysis further indicated that SPP1 may mediate crosstalk between SPP1 + macrophages and other cells via the SPP1-CD44, SPP1-PTGER4 and SPP1-a4b1 complex axes. Additionally, our study suggests that SPP1 + macrophages are associated with the benefit of immune checkpoint blockade (ICB) therapy and that immunotherapy targeting SPP1 + macrophages is expected to improve the prognosis of CRC patients. Furthermore, anti-CSF1R treatment is more likely to preferentially deplete CSF1R-enriched C1QC + macrophages versus SPP1 + macrophages, which may account for the minimal effect of this monotherapy.

Introduction

Colorectal cancer (CRC) is the third most common malignancy and the second leading cause of cancer-related death. Following the great success of immune checkpoint blockade (ICB) in various advanced solid tumors, interest in immunotherapy for CRC is increasing. However, immunotherapy targeting PD-1 is only effective in patients with mismatch repair deficiency (dMMR)/high levels of microsatellite instability (MSI-H), which is present in 5% of metastatic CRC cases. Therefore, it is necessary to understand the complex mechanisms of cellular and molecular interactions in the CRC tumor microenvironment (TME) and to search for potential targets to develop new immunotherapies.

Macrophages are key mediators in the TME and are involved in multiple aspects of tumor immunity. M1 (inflammatory; antitumoral) and M2 (anti-inflammatory; protumoral) macrophage polarization systems have been used to characterize the activation state of macrophages in vitro. However, macrophages exhibit a more complex phenotype in vivo, which contradicts this simple in vitro classification. Macrophages in the TME are often referred to as tumor-associated macrophages (TAMs). TAM is a heterogeneous cell type that contributes to angiogenesis, extracellular matrix (ECM) remodeling, promotes epithelial-mesenchymal transition (EMT) of tumor cells, and activates immunosuppression, promoting the progression and metastasis of CRC. Immunotherapies targeting TAMs, such as those
reducing the number of TAMs in the TME by blocking the CSF1-CSF1R axis\textsuperscript{14–16}, have been applied in the clinic with minimal efficacy as monotherapy in malignancies\textsuperscript{16,17}.

Accumulating single-cell studies have reported that SPP1 + macrophages are important players in the TME\textsuperscript{9,18–20}, are characterized as malignancy driven and are associated with a poor prognosis in CRC patients\textsuperscript{9}. However, the function of SPP1 + macrophages remains unclear. Our single-cell analysis indicated that SPP1 is not only a specific biomarker for SPP1 + macrophages in CRC but also a key functional molecule. Secreted phosphoprotein 1 (SPP1), also called osteopontin (OPN), is overexpressed in most types of cancer\textsuperscript{21–23} and is associated with a poor prognosis in breast, cervical, bladder, colorectal, head and neck, liver, lung, and esophageal cancers\textsuperscript{21,23}. SPP1 can promote CRC progression and metastasis through EMT\textsuperscript{24} and hypoxia pathways\textsuperscript{25}. It has been reported that SPP1 functions by binding to its receptors CD44 and integrin to regulate several signaling pathways (ERK, JNK1, and PI3K/Akt)\textsuperscript{26–31}. However, research on SPP1 + macrophages is still in its infancy, and systematic research is lacking.

The recent rapid development of single-cell RNA-sequencing (scRNA-seq) has revealed the complexity of myeloid cells, especially the predominant macrophages, in a variety of tumors\textsuperscript{9,19,32,33}. However, unlike the major cell types, which have distinctive features and classification markers, there is no generally accepted classification criterion for myeloid cells. Based on the CRC single-cell public dataset, we constructed the largest CRC myeloid cell atlas to date and performed a systematic study of SPP1 + macrophages. Accumulating single-cell evidence suggests that SPP1 + macrophages may be the key to uncovering the heterogeneity and functional changes of macrophages during CRC liver metastasis (CRLM)\textsuperscript{9,18}. On the basis of existing data, we further proposed the SPP1 + macrophage model, which can explain the clinical characteristics and functional changes of macrophages in CRC better than the M1/M2 polarization theory; this model is also expected to help aid in the development of macrophage-targeted immunotherapeutic strategies.

**Result**

**ScRNA-seq revealed the myeloid cell landscape of CRC**

The assessed tissue types of the CRLM dataset included colorectal cancer (CRC), normal colorectum (NC), liver metastases (LM), normal liver (NL), and peripheral blood mononuclear cells (PBMCs) (Fig. 1a). NC and NL samples included both adjacent and healthy tissue. The CRC single-cell data used in this study were obtained from seven public datasets (GSE132465, GSE144735, GSE146771, GSE164522, GSE178318, GSE178341, and Wu et al.) (Fig. 1b). The GSE164522 and Wu et al. datasets were also further defined as CRLM datasets as they included CRC, NC, LM, NL and PBMC samples.

After data processing and annotation, we identified nine main cell types in the seven CRC datasets: T cells, NK cells, B cells, plasma cells, myeloid cells, mast cells, neutrophils, epithelial cells and stromal cells
The proportion of myeloid cells among all immune cells was significantly higher (6/6) in CRC samples than in NC samples (Fig. 1c and Table S1), suggesting that myeloid cells may play an important role in the CRC tumor immune microenvironment.

Myeloid cells were then screened, reclusted and annotated (Fig. S1b). We defined a total of 14 well-characterized myeloid cell subsets, including one plasmacytoid DC subset (pDC-LILRA4), four cDC subsets (cDC1-CLEC9A, cDC2-CD1C, cDC3-LAMP3, and cDC4-LTB), four macrophage subsets (Macro-C1QC, Macro-FCN1, Macro-MKI67, and Macro-SPP1), three doublet subsets (Doublets-B, Doublets-Epi and Doublets-T), and monocyte and Kupffer cell subsets (Fig. S1b and Table S1).

We then analyzed the tissue distribution of these myeloid cell subsets in each of the seven cohorts using odds ratios (ORs) (Fig. 1d). The data showed that cDC1-CLEC9A and cDC2-CD1C were relatively enriched in NC samples, while cDC3-LAMP3 and Macro-SPP1 were relatively enriched in CRC samples; the remaining myeloid subsets were not found to have stable and significant enrichment in NC or CRC samples. In addition, monocytes and Kupffer cells were significantly enriched in PBMC and NL samples, respectively.

To construct a more robust and accurate model of the CRC myeloid cell lineage, CRC myeloid cells in the GSE132465, GSE144735, GSE146771, GSE178318, and GSE178341 datasets were integrated into the large cohort CRC-Mix (n = 48,815) (Fig. S1c). The CRC-Mix data were then reclusted and annotated after harmony correction for batch effects. Finally, we identified 12 myeloid cell subsets in CRC tissue, including pDC-LILRA4, cDC1-CLEC9A, cDC2-CD1C, cDC3-LAMP3, cDC4-LTB, Macro-C1QC, Macro-FCN1, Macro-MKI67, Macro-SPP1, Doublets-B, Doublets-Epi and Doublets-T subsets (Fig. 1e and S1d). In addition, a summary table including representative genes was made to facilitate the identification of these 12 myeloid cell subsets by other researchers, and details of differentially expressed genes (DEGs) between subsets are in the Supplementary Material (Fig. 1f and Table S2).

**Changes in monocytes/macrophages during CRLM**

Single-cell sequencing provides the opportunity to study genetic and functional differences at the cellular level, allowing us to explore changes in the transcriptional profile and functional pathways of monocytes/macrophages in the TME during CRLM.

To observe changes in the macrophage transcriptome during CRC occurrence, we identified DEGs of monocytes/macrophages between NC and CRC samples (GSE164522) and specifically annotated the marker genes of the four major macrophage subtypes in the heatmap (Fig. 2a and Table S3). The results showed that the expression of SPP1 and MKI67 was significantly elevated in CRC samples, while C1QC was highly expressed in NC samples. The same method was used to identify DEGs of monocytes/macrophages in six CRC single-cell datasets (NC vs. CRC) (Fig. 2b and Table S3). The common DEGs (n = 216) in these six datasets were then analyzed by Metascape to further explore the functional and pathway changes in monocytes/macrophages during CRC occurrence (Fig. 2c and Fig. S2a).
Similarly, to observe changes in the macrophage transcriptome during liver metastasis of CRC, we identified DEGs of monocytes/macrophages between CRC and LM samples (GSE164522). The results showed that the expression of \textit{SPP1} and \textit{FCN1} was significantly elevated in LM samples, while the expression of \textit{C1QC} was higher in CRC samples (Fig. 2d). The same method was used to identify DEGs of monocytes/macrophages in three single-cell datasets (CRC vs. LM) (Fig. 2e and Table S4). Common DEGs (n = 1,383) in the three datasets were then analyzed by Metascape to further explore the functional and pathway changes in monocytes/macrophages during liver metastasis of CRC (Figs. 2f and S2b).

To characterize the origin and differentiation fate of monocytes/macrophages during CRC occurrence, development, and liver metastasis, we observed trajectories of monocyte/macrophage subsets, marker genes and tissue types based on pseudo-temporal analysis of Monocle2 (Wu \textit{et al.}) (Fig. 2g-i). The results show that monocytes from PBMCs first differentiate into FCN1 + macrophages after entering colon or liver tissue, and then FCN1 + macrophages polarize toward C1QC + macrophages and SPP1 + macrophages, while Kupffer cells in NL samples are relatively isolated (Fig. 2g and 2h). The pseudo-chronological results of monocyte/macrophage marker genes also showed that the expression of \textit{CD14} and \textit{FCN1} decreased significantly when monocytes/macrophages moved from the PBMC-enriched area to the solid tissue-enriched area (Fig. 2i), suggesting the rapid differentiation of monocytes into macrophages after entering tissues. The expression of \textit{SPP1} was specifically elevated in CRC and LM-concentrated regions, suggesting that SPP1 + macrophages play an important role in the occurrence and metastasis of CRC. In addition, the expression of \textit{MARCO} was concentrated in the NL-rich region.

To explore the functional heterogeneity of monocytes/macrophages during CRLM, gene set variation analysis (GSVA) was used to analyze hallmark pathways of different monocyte/macrophage subsets (Wu \textit{et al.}) (Fig. 2j). The results showed a significant enrichment of angiogenesis-, epithelial mesenchymal transition-, glycolysis-, and hypoxia-related genes in SPP1 + macrophages, and the results were also reliably validated in multiple datasets (Fig. S2c-i). Notably, these hallmark pathways were also significantly altered in monocytes/macrophages during CRC occurrence and LM (Fig. S2a and S2b), suggesting that SPP1 + macrophages play a dominant role in monocytes/macrophages changes during CRLM.

**Tissue distribution and potential mechanisms of SPP1\(^+\) macrophages**

Next, our study focused on the tissue distribution characteristics and biological functions of SPP1 + macrophages. Five scRNA-seq datasets (GSE132465, GSE144735, GSE146771, GSE164522, and GSE178341) showed a significant increase (5/6) in the ratio of SPP1 + macrophages to all macrophages in CRC (NC vs. CRC) (Fig. 3a and Table S1), and data from three scRNA-seq datasets (GSE164522, GSE178318, and Wu \textit{et al.}) further showed a significant increase (3/3) in the SPP1 + macrophage proportion in LM samples (CRC vs. LM) (Fig. 3b and Table S1). The increasing proportion of SPP1 + macrophages during CRC development and liver metastases suggests its critical role in CRLM. To explore whether the increased proportion of SPP1 + macrophages is cancer specific, we integrated CRLM
macrophage (GSE164522) and hepatocellular carcinoma (HCC) macrophage data (GSE156625) and performed reclustering. The results showed that the proportion of SPP1 + macrophages in HCC was not significantly different from that in adjacent tissues but was significantly lower than that in CRLM samples (Fig. 3c and Table S1). This suggests that the elevation of SPP1 + macrophages is only observed in specific cancer tissues (e.g., primary foci and liver metastases of CRC) but is not significant in HCC.

Then, we explored the biological function of SPP1 + macrophages in CRC based on the large CRC-Mix dataset. Macrophages in CRC-Mix included C1QC + macrophages (35.5%, n = 13,640), FCN1 + macrophages (30.6%, n = 11,784), SPP1 + macrophages (27.7%, n = 10,647), and MKI67 + macrophages (6.2%, n = 2,403) (Fig. 3d). MKI67 + macrophages have transcriptional profile highly similar to C1QC + macrophages but specifically express proliferation-related genes such as *MKI67* (Fig. 3h). GSVA and signature scores also showed enrichment of the proliferation signature in MKI67 + macrophages (Figs. 2j and 3a). Considering the low proportion and well-defined proliferation function of MKI67 + macrophages, our follow-up study focused on the three major macrophages subsets (C1QC + macrophages, FCN1 + macrophages and SPP1 + macrophages) in CRC.

To validate the function of each macrophage subset in the literature and our GSVA results, we assessed angiogenesis, EMT, glycolysis, hypoxia, phagocytosis, and MHC-II signatures in different macrophage subsets in CRC-Mix (Figs. 3d-I and S3b-d). The results showed that C1QC + macrophages had the highest phagocytosis, MHC-II score and classical MHC class I genes (*HLA-A, HLA-B, HLA-C*) (Fig. S3c and S3d), which indicated that C1QC + macrophages were the dominant macrophages for antigen presentation and phagocytosis of pathogens. These features also support a protective role of C1QC + macrophages in the CRC TME. A recent CRC single-cell study reported that SPP1-TAMs showed the highest angiogenesis score and the lowest phagocytosis score compared with C1QC-TAMs and MKI67-TAMs. Our study included FCN1 + macrophages and further showed that FCN1 + macrophages had higher angiogenesis scores and lower phagocytosis scores than SPP1 + macrophages (Fig. S3b and S3c). This suggests that although SPP1 + macrophages are characterized by high angiogenesis and low phagocytosis, these are not unique features that constitute the specificity of SPP1 + macrophages. In addition, our results showed specific elevation of EMT, glycolysis, and hypoxia scores in SPP1 + macrophages (Fig. 3e-g), which is consistent with the GSVA results.

The classical M1/M2 polarization theory is based on inflammatory signatures to describe the heterogeneity of macrophages, and we also assessed the inflammatory, anti-inflammatory, M1, and M2 signatures of different macrophage subsets in Mix-CRC (Fig. 3h and 3i). FCN1 + macrophages showed the highest inflammatory score and the lowest anti-inflammatory score (Fig. 3h); similarly, FCN1 + macrophages also showed higher M1 scores and lower M2 scores than C1QC + macrophages and SPP1 + macrophages (Fig. 3i). In addition, we performed enrichment analysis of elevated DEGs in different macrophage subsets separately using Metascape and observed enrichment of the inflammatory response pathway in FCN1 + macrophages (Fig. 3h). The above evidence suggests an inflammatory profile of FCN1 + macrophages. In contrast to FCN1 + macrophages, C1QC + macrophages and SPP1 +
macrophages exhibited more anti-inflammatory features. Enrichment analysis also showed enrichment of the negative regulation of immune system process pathway in C1QC+ macrophages. Studies in glioma\textsuperscript{35} and HCC\textsuperscript{36} have reported the correlation of SPP1 expression with the M2 phenotype, and our data further showed that SPP1+ macrophages have a higher M2 signature than FCN1+ macrophages and C1QC+ macrophages (Fig. 3i). In addition, SPP1+ macrophages exhibited specific expression of CD274 (PD-L1) and HLA-G (a major immune checkpoint)\textsuperscript{37} (Fig. S3d), suggesting an immunosuppressive characteristic of SPP1+ macrophages. Notably, consistent with the M1/M2 signature found in multiple tumor scRNA-seq studies\textsuperscript{38}, we also observed a significant positive correlation between the M1 signature and the M2 signature in CRC macrophages (Fig. 3i).

To explore the specific molecular mechanism by which SPP1+ macrophages exert their biological functions, we used CellPhoneDB to identify the expression of potential crosstalk molecules based on ligand–receptor interactions between different macrophage subsets and major cell types (GSE178341) (Fig. 3j and k). The most abundant receptor–ligand pair interactions were concentrated between different macrophage subsets, and SPP1+ macrophages also showed relatively abundant interactions with stromal cells, epithelial cells, and T cells (Fig. 3j). Interestingly, we found three pairs of receptor–ligand interactions specific for SPP1+ macrophages: SPP1-CD44, SPP1-PTGER4, and the SPP1-a4b1 complex (Fig. 3k). SPP1+ macrophages express high levels of SPP1, and the receptor CD44 is widely present in different cells, while PTGER4 expression is enriched in epithelial cells and T cells, and integrin a4b1 expression is enriched in T cells and plasma cells. These findings reveal the potential molecular mechanisms of SPP1+ macrophages and indicate that SPP1 is not only a key marker gene of SPP1+ macrophages but also a core molecule for the biological functions of SPP1+ macrophages.

In summary, SPP1+ macrophages are malignant cells that are specifically enriched in CRC, with characteristics of angiogenesis, EMT, glycolysis, hypoxia, and immunosuppression.

**SPP1, a novel macrophage marker in CRC**

Single-cell data have revealed the important role of SPP1+ macrophages in CRC. The existing bulk RNA-seq dataset with a large number of samples and rich clinical information is also undoubtedly valuable for studying SPP1+ macrophages, and SPP1 can be the key biomarker to unlocking this treasure trove. Next, we first need to answer two fundamental and important questions: whether SPP1 is a specific marker of SPP1+ macrophages and how SPP1 differs from classical macrophage markers.

To describe changes in the proportion of macrophages among immune cells during CRC occurrence, we compared the proportions of 22 immune cell types in 10 CRC bulk RNA-seq datasets based on the CIBERSORT algorithm (NC vs. CRC) (Fig. 4a). The results, which showed the same significant trend in the comparison of half or more datasets (n ≥ 5), are considered to have high credibility. The data showed that the proportions of M0 macrophages and M1 macrophages were significantly increased in CRC, while the proportion of M2 macrophages was significantly decreased. Next, to describe changes in the numbers of macrophages during CRC occurrence, we compared the expression levels of five classical macrophage markers in 10 CRC bulk datasets (NC vs. CRC) (Fig. 4a). The results did not show significant changes in
the pan-macrophage marker CD68 or the M1 macrophage markers CD68 and iNOS (NOS2), while the M2 macrophage markers CD163 and CD206 (MRC1) were significantly downregulated in CRC.

The above results of changes in the proportion and number of macrophages during CRC occurrence seem to contradict our single-cell results that the proportion of myeloid cells is significantly elevated in CRC. We speculate that this is related to the overall decline in immune cells in the tumor immune microenvironment during CRC occurrence. Therefore, we used the immune score based on the ESTIMATE algorithm and the immune cell marker CD45 (PTPRC) to evaluate the difference in the level of immune infiltration between NC and CRC samples (Fig. 4a), and the results showed a significant decrease in immune infiltration in CRC tissues. We further assessed the ratio of macrophages to immune cells using CD68/CD45, and the results showed that CD68/CD45 was significantly elevated in CRC. These findings indicated that the number of macrophages in tissues did not change significantly during the occurrence of CRC, while the decrease in the number of overall immune cells resulted in a significant increase in the proportion of macrophages.

Unlike classical macrophage markers whose levels were unchanged or decreased, SPP1 (9/9) and SPP1/CD68 (6/6) were significantly upregulated in CRC (Fig. 4a). Correlation analysis between SPP1 expression and immune cell proportions (CIBERSORT) showed that SPP1 expression in TCGA-CRC was positively correlated with four myeloid cell types, including neutrophils, M0 macrophages, M2 macrophages, and mast cells (Fig. 4b). The correlation analysis between SPP1 and six types of immune cells in TCGA-CRC based on TIMER showed that SPP1 had the highest correlation with macrophages (Fig. S4a). In addition, correlation analysis showed that SPP1 was highly positively correlated with CD45, CD68, CD86, CD163, and CD206 but negatively correlated with iNOS (Fig. S4b). This evidence indicates that SPP1 expression is highly correlated with the infiltration of macrophages, especially M2 macrophages, in CRC. Single-cell sequencing data further revealed that SPP1 expression was only enriched in specific myeloid cells in CRC, and the enrichment pattern was almost consistent with our definition of SPP1 + macrophage subsets (Figs. 4c, d, S4c and d). In other myeloid cells, other immune cells, and nonimmune cells, SPP1 is expressed at very low levels. These results demonstrate that SPP1 is indeed a specific marker for SPP1 + macrophages in CRC.

The IHC staining from The Human Protein Atlas (THPA, https://www.proteinatlas.org/) database also verified that SPP1 protein was specifically expressed in CRC samples, while no staining was detected in NC, NL and HCC samples (Fig. 4e). In addition, to assess the function of the SPP1 gene in CRC, we used the Gene Set Enrichment Analysis (GSEA) function of LinkedOmics to analyze the coexpressed genes of SPP1 in TCGA-CRC (Fig. 4f). The top five significantly enriched GO BP terms were extracellular structure organization, cell adhesion mediated by integrin, cellular response to lipoprotein particle stimulus, cellular response to VEGF stimulus and macrophage activation.

In summary, SPP1 is a specific biomarker of SPP1 + macrophages and is significantly elevated in CRC. Clinical value of SPP1 + macrophages in CRC (bulk RNA-seq)
Five CRC bulk datasets (TCGA, GSE14333, GSE17536, GSE41258, and GSE161158) were used to compare $SPP1$ expression in different stages (stage I/II vs. stage III/IV) (Fig. 5a). The results showed elevated $SPP1$ expression in patients with stage III/IV (advanced) CRC (5/5). We also used three CRC bulk datasets (TCGA, GSE39582, and GSE41258) to compare the expression of $SPP1$ in samples with different TNM classifications (T1/2 vs. T3/4, N0 vs. N1-4, and M0 vs. M1) (Fig. 5b-d). The results showed higher $SPP1$ expression in T3/4 (3/3) and N1-3 (2/3) samples than in T1/2 and N0 samples. The close relationship between $SPP1$ expression level and the degree of tumor infiltration and lymph node metastasis suggests an important role of $SPP1 +$ macrophages in CRC progression.

Furthermore, we compared the expression of $SPP1$ between CRC and LM samples in six bulk datasets (GSE6988, GSE41258, GSE41568, GSE68468, GSE131418 Consortium, and GSE131418 MCC) (Fig. 5e). $SPP1$ expression was significantly higher in LM samples than in CRC samples (4/6), suggesting an important role for $SPP1 +$ macrophages in CRC metastasis. We also observed a further decrease in immune infiltration in LM samples compared to CRC samples (5/6).

Macrophage enrichment indicates a poor prognosis in most tumors, but macrophage enrichment in CRC tends to be associated with a favorable prognosis$^{13,40,41}$. Univariate Cox analysis of four datasets (TCGA, GSE17536, GSE39582, and GSE41258) indicated that $SPP1$ expression was significantly associated with poor overall survival (OS) (3/4) (Fig. 5f). We performed the same analysis for traditional macrophage markers: $iNOS$ showed a protective effect, while $CD68$, $CD86$, $CD163$ and $CD206$ were not significantly related to CRC progression, metastasis or prognosis.

Considering the colinearity between $SPP1 +$ macrophages ($SPP1$) and macrophages ($CD68$), we used $SPP1/CD68$ coexpression status to study the relationship between $SPP1 +$ macrophages and the clinical features of CRC (Table 1). The TCGA-CRC samples were divided into high (H-$SPP1/CD68$) and low (L-$SPP1/CD68$) groups according to the median value of $SPP1/CD68$. Patients in the H-$SPP1/CD68$ group had a higher mean age; a higher proportion of right-side tumors, advanced T classification, lymph node metastasis, MSI-H status and high TMB; and shorter survival time (OS and PFI). In addition, GSEA verified the enrichment of hallmark pathways such as EMT, angiogenesis and hypoxia in the H-$SPP1/CD68$ group (Fig. 5g).
### Table 1
Comparison of clinical parameters between the L-\textit{SPP1/CD68} and H-\textit{SPP1/CD68} groups in CRC.

<table>
<thead>
<tr>
<th>Clinical parameters</th>
<th>L-\textit{SPP1/CD68} (N= 309)</th>
<th>H-\textit{SPP1/CD68} (N= 308)</th>
<th>(p) value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age (years)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean ± SD</td>
<td>64.7 ± 12.7</td>
<td>67.8 ± 12.5</td>
<td>0.002\textsuperscript{a}</td>
</tr>
<tr>
<td>NA</td>
<td>3</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>140 (45.8%)</td>
<td>146 (47.4%)</td>
<td>0.682\textsuperscript{b}</td>
</tr>
<tr>
<td>Male</td>
<td>166 (54.2%)</td>
<td>162 (52.6%)</td>
<td></td>
</tr>
<tr>
<td>NA</td>
<td>3</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td><strong>Location</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left</td>
<td>180 (61.2%)</td>
<td>160 (53.2%)</td>
<td>0.047\textsuperscript{b}</td>
</tr>
<tr>
<td>Right</td>
<td>114 (38.8%)</td>
<td>141 (46.8%)</td>
<td></td>
</tr>
<tr>
<td>NA</td>
<td>15</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td><strong>Pathologic stage</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>62 (21.3%)</td>
<td>41 (13.5%)</td>
<td>0.078\textsuperscript{b}</td>
</tr>
<tr>
<td>II</td>
<td>109 (37.5%)</td>
<td>117 (38.6%)</td>
<td></td>
</tr>
<tr>
<td>III</td>
<td>82 (28.2%)</td>
<td>96 (31.7%)</td>
<td></td>
</tr>
<tr>
<td>IV</td>
<td>38 (13.1%)</td>
<td>49 (16.2%)</td>
<td></td>
</tr>
<tr>
<td>NA</td>
<td>18</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td><strong>T classification</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>14 (4.6%)</td>
<td>5 (1.6%)</td>
<td>0.005\textsuperscript{b}</td>
</tr>
<tr>
<td>T2</td>
<td>63 (20.7%)</td>
<td>41 (13.4%)</td>
<td></td>
</tr>
<tr>
<td>T3</td>
<td>199 (65.5%)</td>
<td>219 (71.3%)</td>
<td></td>
</tr>
<tr>
<td>T4</td>
<td>28 (9.2%)</td>
<td>42 (13.7%)</td>
<td></td>
</tr>
<tr>
<td>NA</td>
<td>5</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

\textsuperscript{a} Mann-Whitney test, \textsuperscript{b} Chi-square test, \textsuperscript{c} Log-rank test. TMB, tumor mutation burden; high TMB (> 10mut/MB); CI, confidence interval; OS, overall survival; PFI, progression-free interval; Location left includes the cecum to the transverse colon; Location left includes the splenic flexure to the rectum.
<table>
<thead>
<tr>
<th>Clinical parameters</th>
<th>L-SPP1/CD68 (N=309)</th>
<th>H-SPP1/CD68 (N=308)</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>N classification</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N0</td>
<td>182 (60.1%)</td>
<td>165 (53.7%)</td>
<td>0.025\textsuperscript{b}</td>
</tr>
<tr>
<td>N1</td>
<td>77 (25.4%)</td>
<td>71 (23.1%)</td>
<td></td>
</tr>
<tr>
<td>N2</td>
<td>44 (14.5%)</td>
<td>71 (23.1%)</td>
<td></td>
</tr>
<tr>
<td>NA</td>
<td>6</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td><strong>Metastasis</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M0</td>
<td>216 (85.4%)</td>
<td>236 (82.8%)</td>
<td>0.417\textsuperscript{b}</td>
</tr>
<tr>
<td>M1</td>
<td>37 (14.6%)</td>
<td>49 (17.2%)</td>
<td></td>
</tr>
<tr>
<td>NA</td>
<td>56</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td><strong>MSI-H</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>32 (10.5%)</td>
<td>50 (16.8%)</td>
<td>0.024\textsuperscript{b}</td>
</tr>
<tr>
<td>No</td>
<td>272 (89.5%)</td>
<td>247 (83.2%)</td>
<td></td>
</tr>
<tr>
<td>NA</td>
<td>5</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td><strong>TMB</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>32 (11.7%)</td>
<td>50 (20.5%)</td>
<td>0.006\textsuperscript{b}</td>
</tr>
<tr>
<td>Low</td>
<td>242 (88.3%)</td>
<td>194 (79.5%)</td>
<td></td>
</tr>
<tr>
<td>NA</td>
<td>35</td>
<td>64</td>
<td></td>
</tr>
<tr>
<td><strong>OS (years)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (95%CI)</td>
<td>7.97 (6.94–9.01)</td>
<td>6.49 (5.43–7.54)</td>
<td>0.046\textsuperscript{c}</td>
</tr>
<tr>
<td>Outcome event</td>
<td>55/299</td>
<td>68/294</td>
<td></td>
</tr>
<tr>
<td><strong>PFI (years)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (95%CI)</td>
<td>8.09 (7.08–9.09)</td>
<td>6.03 (4.99–7.06)</td>
<td>0.030\textsuperscript{c}</td>
</tr>
<tr>
<td>Outcome event</td>
<td>67/298</td>
<td>86/293</td>
<td></td>
</tr>
</tbody>
</table>

\textsuperscript{a} Mann-Whitney test, \textsuperscript{b} Chi-square test, \textsuperscript{c} Log-rank test. TMB, tumor mutation burden; high TMB (> 10mut/MB); CI, confidence interval; OS, overall survival; PFI, progression-free interval; Location left includes the cecum to the transverse colon; Location left includes the splenic flexure to the rectum.
In summary, SPP1 + macrophages can be used as a malignant biomarker to evaluate CRC progression, metastasis and prognosis.

**Significance of SPP1 + macrophages for CRC treatment**

Kaplan–Meier survival analysis showed that the H-*SPP1/CD68* group had shorter OS than the L-*SPP1/CD68* group (log-rank, P < 0.001) (Fig. 6a). CRC single-cell data revealed the relationship between preoperative chemotherapy and SPP1 expression in macrophages (Fig. 6b and c). Patients with partial response (PR) or stable disease (SD) were defined as responders, while patients with progressive disease (PD) were defined as nonresponders. The results showed that SPP1 expression in macrophages was significantly decreased in patients treated with preoperative chemotherapy, and SPP1 expression in macrophages was further decreased in responders (those with PR/SD) compared with nonresponders (those with PD). This suggests that preoperative chemotherapy may exert a therapeutic effect by specifically reducing the number of SPP1 + macrophages in the tumor immune microenvironment. We also speculate that immunotherapy targeting SPP1 + macrophages may improve the prognosis of CRC patients with a poor prognosis.

Interestingly, both scRNA-seq data and bulk RNA-seq data showed a strong association between SPP1 + macrophages and genome instability and mutation (dMMR, MSI-H and high TMB) of CRC (Fig. 6d-I and Table 1), suggesting that CRC patients with high SPP1 + macrophages had the characteristics of immunotherapy benefit. We also used SubMap algorithms to predict immunotherapy responsiveness in the L-*SPP1/CD68* and H-*SPP1/CD68* groups (Fig. 6j). The results indicated that the H-*SPP1/CD68* group was more likely to respond to anti-PD-1 therapy (Bonferroni corrected $P = 0.021$) and anti-CTLA-4 therapy (Bonferroni corrected $P = 0.023$). In addition, the results from the GSE39582 dataset also validate that patients with high *SPP1* expression are more likely to respond to anti-PD-1 therapy and anti-CTLA-4 therapy (Fig. 6k).

The development and maintenance of macrophages requires a continuous supply of the nutritional factor *CSF1*, and *CSF1R* is also essential for the development, survival, recruitment, and proliferation of most macrophages. CRC-Mix data showed that *CSF1R* expression was significantly enriched in C1QC + macrophages but not SPP1 + macrophages (Fig. 6l), which was also validated in two other scRNA-seq cohorts (GSE164522 and Wu *et al*.) (Fig. S5a and b). This suggests that immunotherapy that blocks *CSF1R* preferentially depletes C1QC + macrophages over SPP1 + macrophages.

**Discussion**

We investigated SPP1 + macrophages in CRC from various aspects; we assessed their cell origin and tissue distribution, clinical value as a marker of CRC malignancy, potential functional pathways and cell–cell interactions, and implications for treatment (Fig. 7a-e). Based on the literature and the results of this study, we propose the SPP1 + macrophage model theory. Compared with the M1/M2 polarization theory, this theory can better explain the changes in macrophages in the whole process of CRLM and guide clinical diagnosis, condition assessment and treatment. Importantly, the SPP1 + macrophage model
provides a unique single-cell perspective to help us understand the heterogeneity of macrophages in the CRC tumor immune microenvironment.

We identified four macrophage subsets from CRC tissue, including FCN1 + macrophages, C1QC + macrophages, SPP1 + macrophages and MKI67 + macrophages (Fig. 7a). Enrichment analysis and phenotypic signatures indicated that inflammation, phagocytosis, malignancy and proliferation were the most prominent features related to these four macrophage subsets.

SPP1 + macrophages were specifically enriched in primary foci and liver metastases of CRC but were almost absent in normal colorectal and liver tissues and HCC tissues. Evidence from trajectory analysis and tissue distribution analysis of macrophage subsets supports the following hypothesis regarding the origin and differentiation of SPP1 + macrophages: monocytes from peripheral blood are recruited into colonic tissues and rapidly differentiate into macrophages, which differentiate into monocyte-like FCN1 + macrophages that polarize toward C1QC + macrophages and SPP1 + macrophages. Subsequently, SPP1 + macrophages that are enriched in CRC primary foci are transferred to the liver along with tumor tissue and further enriched (Fig. 7b).

The expansion of myeloid cells in CRC tissue suggests their active role in shaping the TME. Our data also confirmed the significant increase in the proportion of myeloid cells and macrophages in CRC. However, our further study found that the number of macrophages in tissues did not change significantly during CRC tumorigenesis, but the decrease in the number of overall immune cells resulted in a significant increase in the proportion of macrophages (Fig. 7c). The decrease in immune infiltration during the occurrence and metastasis of CRC suggests that antitumor immunity weakens as the disease progresses, while the importance of macrophages in the TME continues to increase.

Accumulating single-cell studies have reported the association of SPP1 + TAMs with malignancies, especially CRC. A recent study also found that SPP1 + TAMs in CRC were associated with LM. Numerous studies have reported that SPP1 is a diagnostic and prognostic marker for CRC. Our results integrate the content of the above studies. We found that SPP1 is a specific marker of SPP1 + macrophages, and SPP1 + macrophages can be used as an indicator to assess the malignancy of CRC. The number and proportion of SPP1 + macrophages constantly increase during the occurrence, progression and metastasis of CRC and indicate a poor prognosis (Fig. 7c). These results support that SPP1 + macrophages in the TME may play an important driving role in the occurrence, progression and metastasis of CRC. SPP1 + macrophages are also expected to be a comprehensive marker for CRC diagnosis, disease severity assessment and prognosis evaluation.

The “classically activated” M1 and “alternately activated” M2 macrophage polarization systems have been widely used to describe the activation state of macrophages in vitro. However, macrophages in vivo exhibit a more complex phenotype, which contradicts this simple in vitro classification. Our studies in CRC also demonstrate the limitations of this in vitro polarization model. First, the M1/M2 polarization theory considers M1 and M2 activation states as mutually exclusive discrete states, while multiple recent
single-cell studies have shown that M1 and M2 gene signatures are coexpressed in macrophage subsets in nearly all cancer types\textsuperscript{19,32,38,47}. Our data also found that M1 and M2 signatures coexisted in CRC macrophages and were significantly positively correlated. Second, according to the M1/M2 polarization theory, proinflammatory M1 macrophages are thought to prevent tumor progression, while M2 macrophages promote tumor growth and metastasis\textsuperscript{34,47}. This theory can well explain the changes in macrophages in the TME in some tumors\textsuperscript{34,48}. However, our results based on CIBERSORT showed that the proportion of M1 macrophages was significantly increased in CRC, while the proportion of M2 macrophages was significantly decreased, which contradicted the M1/M2 polarization theory. Furthermore, in our study of CRC progression, metastasis and prognosis, no significant differences were observed in M1/M2 markers except for the M1 marker \textit{iNOS}, which presented a protective effect. The above evidence indicates that the M1/M2 polarization theory is not applicable in the CRC TME, which also suggests that the malignant features of SPP1 + macrophages are not derived from their M2 signatures.

The biological function of SPP1 + macrophages is still unclear. A recent study suggested that SPP1 + TAMs promote CRC through enhancing angiogenesis and attenuating phagocytosis\textsuperscript{9}. However, our study found that FCN1 + macrophages exhibited higher angiogenesis signatures and lower phagocytosis signatures than SPP1 + macrophages, suggesting that SPP1 + macrophages also promote tumors through other important pathways. Various analyses in this study confirmed the enrichment of EMT, hypoxia, glycolysis, and immunosuppressive signatures in SPP1 + macrophages, suggesting that SPP1 + macrophages may promote CRC through related pathways (Fig. 7d). Furthermore, our study found that SPP1 is also a key functional molecule of SPP1 + macrophages. CellPhoneDB predicted that SPP1 + macrophages interact with other cells through the \textit{SPP1-CD44}, \textit{SPP1-PTGER4}, and \textit{SPP1-a4b1 complex} axes. The \textit{CD44} receptor plays an important role in tumor EMT, invasion and metastasis\textsuperscript{49, PTGER4\textsuperscript{50}}, and integrin a4b1\textsuperscript{51} have also been reported to be involved in cancer progression and metastasis. This suggests that SPP1 + macrophages may promote CRC progression and metastasis through the interaction of secreted \textit{SPP1} with receptors \textit{CD44}, \textit{PTGER4}, and integrin a4b1 (Fig. 7d).

The understanding of SPP1 + macrophages is expected to provide new options for the treatment of CRC. Chemotherapy-induced macrophage phenotype switching has been reported in breast cancer\textsuperscript{52}. We also found that preoperative chemotherapy significantly reduced \textit{SPP1} expression in CRC macrophages, particularly in responders. This suggests that effective chemotherapy may specifically inhibit SPP1 + macrophages. Moreover, it suggests the possibility of immunotherapy targeting SPP1 + macrophages. Interestingly, we found that although SPP1 + macrophages predict a poor prognosis, patients with high SPP1 + macrophage levels show higher levels of genome instability and mutation characteristics (MSI-H, dMMR, and high TMB) (Fig. 7d). dMMR/MSI-H status is a recognized immunotherapy biomarker in colorectal cancer and other solid tumors\textsuperscript{53}. In addition, a high TMB also predicts better therapeutic efficacy of ICBs in several solid tumors, including CRC\textsuperscript{54–56}. This suggests that CRC patients with high SPP1 + macrophages are more likely to benefit from ICB immunotherapy. This notion is further supported by the prediction results of SubMap algorithm and the enrichment of \textit{PD-L1} and \textit{HLA-G} in SPP1 +
macrophages. In addition, a large number of studies have reported that targeted blockade of SPP1 can inhibit tumor progression and metastasis\textsuperscript{21,25,57−59}. Substantial evidence from this study supports the development of drugs targeting SPP1 + macrophages as a viable option for CRC immunotherapy.

Notably, our SPP1 + macrophage theory also reveals a possible problem in current macrophage-targeted immunotherapy research. Inhibition of the \textit{CSF1-CSF1R} axis to reduce TAM levels is a major research direction of macrophage-targeted immunotherapy\textsuperscript{6,14−16}. However, we found that \textit{CSF1R} was significantly enriched in C1QC + macrophages compared to other macrophages, suggesting that \textit{CSF1R} blockade may preferentially deplete the C1QC + macrophage subset while sparing SPP1 + macrophages. Anti-CSF1R therapy may not be sufficient to deplete all tumor-promoting SPP1 + macrophages, which may contribute to the poor efficacy of anti-CSF1R monotherapy in the Renca mouse tumor model and in human cancer patients\textsuperscript{16,17,19}.

However, this study has some limitations. First, the lack of a consensus classification system for myeloid cell subsets identified by single-cell sequencing has become a key obstacle limiting further research on subsets. Therefore, we mainly referred to the study of Zhang et al.\textsuperscript{19} for the classification of macrophage subsets and only identified four macrophage subsets that are ubiquitous and well characterized, while some rare macrophage subsets (such as \textit{CXCL10}- and ISG-enriched macrophage subsets\textsuperscript{18,32,60}) were generally incorporated into the macrophage subsets with similar transcriptomes. Second, we observed specific enrichment of SPP1 + macrophages in LM samples. These SPP1 + macrophages may be derived from peripheral blood monocytes acclimated by the TME of the LM or from the abnormal activation and proliferation of transferred SPP1 + macrophages. However, our current techniques remain unable to determine the origin of SPP1 + macrophages in metastases. Third, extensive data in this study suggest that CRC patients with high SPP1 + macrophage levels are more likely to benefit from immunotherapy, but this evidence is all based on correlational inferences. We have not yet found a large-scale CRC immunotherapy dataset that can be used to validate the hypothesis. Conversely, a recently published study suggested that SPP1 + macrophages reduced the efficacy of PD-L1 blockade treatment\textsuperscript{24}. Therefore, additional in-house clinical data are needed to validate the relationship between SPP1 + macrophages and ICB immunotherapy.

In conclusion, based on public scRNA-seq and bulk RNA-seq datasets, we propose the SPP1 + macrophage model paradigm, which can explain the characteristics and changes of macrophages in CRC better than the M1/M2 polarization theory. We found that SPP1 + macrophages can serve as a malignant marker of CRC occurrence, progression, metastasis and poor prognosis. We also explored the underlying mechanisms of SPP1 + macrophages and suggest that immunotherapy targeting SPP1 + macrophages (in terms of abundance and function) is a promising CRC therapeutic strategy.

**Materials And Methods**

**Material**
The public datasets used in this study are from the GEO and TCGA and the website that stores the scRNA-seq data of Wu et al. These datasets include eight scRNA-seq datasets and 16 bulk RNA-seq datasets (Table 2). To ensure the stability and comparability of single-cell data, only scRNA-seq datasets based on the 10x Genomics platform were included in this study; seven CRC datasets (GSE132465, GSE144735, GSE146771, GSE164522, GSE178318, GSE178341, and Wu et al.) and one HCC dataset (GSE156625) met this criterion. GSE164522 and Wu et al. were further defined as CRLM datasets due to their tissue sample types, including CRC, NC, LM, NC, and PBMC. The bulk RNA-seq dataset includes high-throughput sequencing data from TCGA-CRC and 15 GEO microarray datasets (GSE6988, GSE14333, GSE17536, GSE20842, GSE20916, GSE39582, GSE41258, GSE44076, GSE44861, GSE68468, GSE83889, GSE87211, GSE106582, GSE131418 and GSE161158). Transcriptome data and clinical information of TCGA-CRC were downloaded from UCSC Xena (http://xena.ucsc.edu/), and single nucleotide variation data (VarScan) were downloaded from the Genomic Data Commons (GDC) portal (https://portal.gdc.cancer.gov/). The transcriptome and clinical information of the GEO dataset were downloaded from the GEO database (https://www.ncbi.nlm.nih.gov/geo/).
Table 2
Source of scRNA-seq and bulk RNA-seq datasets.

<table>
<thead>
<tr>
<th>Deposited Data</th>
<th>Platform</th>
<th>Source</th>
<th>Identifier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human CRC scRNA-seq dataset</td>
<td>10x Genomics</td>
<td>Lee et al., 2020^{44}</td>
<td>GEO: GSE132465</td>
</tr>
<tr>
<td>Human CRC scRNA-seq dataset</td>
<td>10x Genomics</td>
<td>Lee et al., 2020^{44}</td>
<td>GEO: GSE144735</td>
</tr>
<tr>
<td>Human CRC scRNA-seq dataset</td>
<td>10x Genomics</td>
<td>Zhang et al., 2020^{19}</td>
<td>GEO: GSE146771</td>
</tr>
<tr>
<td>Human CRC scRNA-seq dataset</td>
<td>10x Genomics</td>
<td>Che et al., 2021^{61}</td>
<td>GEO: GSE178318</td>
</tr>
<tr>
<td>Human CRC scRNA-seq dataset</td>
<td>10x Genomics</td>
<td>Pelka et al., 2021^{60}</td>
<td>GEO: GSE178341</td>
</tr>
<tr>
<td>Human CRLM scRNA-seq dataset</td>
<td>10x Genomics</td>
<td>Liu et al., 2022^{9}</td>
<td>GEO: GSE164522</td>
</tr>
<tr>
<td>Human CRLM scRNA-seq dataset</td>
<td>10x Genomics</td>
<td>Wu et al., 2022^{18}</td>
<td><a href="http://www.cancerdiversity.asia/scCRLM">http://www.cancerdiversity.asia/scCRLM</a></td>
</tr>
<tr>
<td>Human HCC scRNA-seq dataset</td>
<td>10x Genomics</td>
<td>Sharma et al., 2020^{62}</td>
<td>GEO: GSE156625</td>
</tr>
<tr>
<td>Human CRC bulk RNA-seq dataset</td>
<td>Illumina Hiseq</td>
<td>-</td>
<td>TCGA</td>
</tr>
<tr>
<td>Human CRC bulk RNA-seq dataset</td>
<td>GPL4811</td>
<td>Ki et al., 2007^{63}</td>
<td>GEO: GSE6988</td>
</tr>
<tr>
<td>Human CRC bulk RNA-seq dataset</td>
<td>GPL570</td>
<td>Jorissen et al., 2009^{64}</td>
<td>GEO: GSE14333</td>
</tr>
<tr>
<td>Human CRC bulk RNA-seq dataset</td>
<td>GPL570</td>
<td>Smith et al., 2010^{65}</td>
<td>GEO: GSE17536</td>
</tr>
<tr>
<td>Human CRC bulk RNA-seq dataset</td>
<td>GPL4133</td>
<td>Gaedcke et al., 2010^{66}</td>
<td>GEO: GSE20842</td>
</tr>
<tr>
<td>Human CRC bulk RNA-seq dataset</td>
<td>GPL570</td>
<td>Skrzypczak et al., 2010^{67}</td>
<td>GEO: GSE20916</td>
</tr>
<tr>
<td>Human CRC bulk RNA-seq dataset</td>
<td>GPL570</td>
<td>Marisa et al., 2013^{68}</td>
<td>GEO: GSE39582</td>
</tr>
<tr>
<td>Human CRC bulk RNA-seq dataset</td>
<td>GPL96</td>
<td>Sheffer et al., 2009^{69}</td>
<td>GEO: GSE41258</td>
</tr>
<tr>
<td>Human CRC bulk RNA-seq dataset</td>
<td>GPL13667</td>
<td>Solé et al., 2014^{70}</td>
<td>GEO: GSE44076</td>
</tr>
</tbody>
</table>
Deposited Data

<table>
<thead>
<tr>
<th>Deposited Data</th>
<th>Platform</th>
<th>Source</th>
<th>Identifier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human CRC bulk RNA-seq dataset</td>
<td>GPL3921</td>
<td>Ryan et al., 2014(^71)</td>
<td>GEO: GSE44861</td>
</tr>
<tr>
<td>Human CRC bulk RNA-seq dataset</td>
<td>GPL96</td>
<td>Sheffer et al., 2009(^69)</td>
<td>GEO: GSE68468</td>
</tr>
<tr>
<td>Human CRC bulk RNA-seq dataset</td>
<td>GPL10558</td>
<td>Kwon et al., 2017(^72)</td>
<td>GEO: GSE83889</td>
</tr>
<tr>
<td>Human CRC bulk RNA-seq dataset</td>
<td>GPL13497</td>
<td>Hu et al., 2018(^73)</td>
<td>GEO: GSE87211</td>
</tr>
<tr>
<td>Human CRC bulk RNA-seq dataset</td>
<td>GPL10558</td>
<td>-</td>
<td>GEO: GSE106582</td>
</tr>
<tr>
<td>Human CRC bulk RNA-seq dataset</td>
<td>GPL15048</td>
<td>Kamal et al., 2019(^74)</td>
<td>GEO: GSE131418</td>
</tr>
<tr>
<td>Human CRC bulk RNA-seq dataset</td>
<td>GPL570</td>
<td>Szeglin et al., 2022(^75)</td>
<td>GEO: GSE161158</td>
</tr>
</tbody>
</table>

**Basic analysis workflow for scRNA-seq data**

The R package Seurat (version 4.1.0) converts raw unique molecular identifier (UMI) count matrices to Seurat objects and then filters the data according to the following criteria: (1) filter genes that are expressed in at least 10 cells; (2) filter cells that have unique feature counts over 6,000 or less than 500; and (3) filter cells that have >20% mitochondrial counts. After quality control, "LogNormalize" was used to normalize the expression of all cells with a scale factor of 10,000. The top 2000 highly variable genes were identified based on the mean and dispersion. During scaling of the highly variable genes, the "var.to.regress" option was used to regress percent mitochondrial content. The results are obtained by principal component analysis (PCA) for linear dimensionality reduction. We used the `FindClusters` function on 50 principal components with a resolution of 0.8 for preliminary clustering and annotation. Afterward, the visualization of cell clustering is completed by the nonlinear dimensionality reduction UMAP method\(^76\). Finally, the main cell types were identified and annotated based on canonical marker genes. Subsequent annotation of myeloid cell subsets should follow the same workflow described above for major cell types.

**Annotation of major cell types and myeloid cell subsets**

The main basis for classifying clusters into specific major cell types and myeloid cell subsets was as follows: (1) expression level of key marker genes in clusters; (2) the relative distance between clusters in the UMAP graph; and (3) the distance between clusters evaluated by Cluster tree algorithm.

The following key marker genes of major cell types were used in this study: T cells (CD3D), NK cells (KLRF1), B cells (MS4A1), plasma cells (MZB1), myeloid cells (LYZ), mast cells (TPSAB1), neutrophils.
(FCGR3B), epithelial cells (EPCAM), and stromal cells (DCN).

Myeloid cells were defined in this study based on the similarity of cell clusters rather than strictly following biological definitions. Therefore, in the cluster annotation of major cell types, real myeloid cells such as neutrophils and mast cells were independently named, while pDCs of lymphoid origin\textsuperscript{77}, Kupffer cells derived from embryonic yolk sac\textsuperscript{78}, and even doublets were all classified as myeloid cells.

The myeloid cell subsets were identified based on markers as follows: pDC-LILRA4 (LILRA4), cDC1-CLEC9A (CLEC9A), cDC2-CD1C (CD1C), cDC3-LAMP3 (LAMP3), cDC4-LTB (LTB), Macro-C1QC (C1QC), Macro-FCN1 (FCN1), Macro-MKI67 (MKI67), Macro-SPP1 (SPP1), Doublets-B (CD79A, MS4A1 and MZB1), Doublets-Epi (EPCAM), Doublets-T (CD3D), Monocytes (CD14), and Kupffer cells (MARCO).

Since CRC macrophages were the focus of this study, the major monocyte/macrophage subpopulation in PBMCs of the CRLM dataset was only broadly defined as monocytes. In the datasets without PBMC samples, we did not define the monocyte subset, and because of the high similarity in the transcriptome characteristics of monocytes and Macro-FCN1, the potential monocytes in these datasets are usually classified into the Macro-FCN1 subset. Similarly, in datasets without NL samples, we do not define a subset of Kupffer cells.

Macro-C1QC, Macro-FCN1, Macro-MKI67, Macro-SPP1, monocytes, and Kupffer cells in this study were collectively referred to as monocytes/macrophages; macrophages included Macro-C1QC, Macro-FCN1, Macro-MKI67, and Macro-SPP1 subsets.

**Data integration and batch correction**

To construct the large sample data of CRC myeloid cells, CRC myeloid cell data from five independent datasets (GSE132465, GSE144735, GSE146771, GSE178318, and GSE178341) were combined into CRC-Mix data. After correcting for batch effects using the harmony algorithm\textsuperscript{79}, the CRC-Mix data were reclustered into groups and annotated.

In addition, to compare the differences in SPP1 + macrophage proportions in CRLM and HCC, myeloid cells of GSE156625 and GSE164522 were integrated in the same way, corrected for batch effects by the harmony algorithm, reclustered into groups and annotated.

**Expression difference analysis and representative genes**

The *Findallmarkers* function in Seurat was used to identify marker genes for each cluster or subset. In addition, to explore the functional changes in monocytes/macrophages during CRLM, we also used the *Findmarkers* function to compare the monocyte/macrophage transcriptomes of NC and CRC, as well as CRC and LM. The parameters assert="RNA" and slot="counts" allowed for the inclusion of genes with low variability in the comparison. Genes with significant expression differences (p < 0.05) in the comparison between groups were defined as DEGs. DEGs with the same trend in multiple datasets were defined as common DEGs, and these genes were used for subsequent functional enrichment analysis.
In brief, for myeloid cell subsets in the CRC-Mix dataset, genes that met these criteria were considered representative genes: 1) adjusted P value < 0.01 with all features; 2) log fold-change of the average expression > 0.5; 3) pct.1 (percentage of gene expression detected in the first group) > 0.25; 4) The above three-step filtering process was also performed on the myeloid cell subsets defined before harmony correction, and the intersection genes were selected. These genes were further checked manually to ensure their expression specificity on the corresponding subsets.

**Tissue distribution of myeloid cell subsets**

To understand the distribution preference of each myeloid cell subtype in different tissues, we evaluated the enrichment of each myeloid cell subtype using the OR in seven CRC single-cell datasets. OR > 1 indicates the relative enrichment of the cell subtype in a specific tissue.

**Monocyte/macrophage developmental trajectory in CRLM**

To construct the monocyte/macrophage developmental trajectory of colorectal cancer liver metastases, we used the "Monocle" package (version 2.22.0)\(^8\) to align monocytes/macrophages in pseudotime order. The DDRTree method implemented with the reduceDimension function of Monocle 2 was used for dimensionality reduction and construction of pseudo-temporal order.

**Cell–cell interaction analysis**

We used Python-based (version 3.7) CellphoneDB\(^8\) to evaluate the interactions of macrophages with other cells. The putative ligands and receptors were determined based on whether they were expressed on each cell. To efficiently assess the closeness of cell interactions based on the number of receptor–ligand pairs, we randomly sampled 1,000 cells per population from the Macro-C1QC, Macro-FCN1, Macro-MKI67, Macro-SPP1 and major cell types in the GSE178318 dataset.

**Defining phenotype scores**

To understand the characteristics between different macrophage subsets, scores for different phenotypes were obtained by the AddModuleScore function in the "Seurat" package. These scores were defined by the mean expression of phenotype-related genes.

The related genes of M1 macrophages and M2 macrophages were defined as M1 score and M2 score, respectively. The M1 score considers the expression of CCL5, CCR7, CD40, CD86, CXCL9, CXCL10, CXCL11, IDO1, IL1A, IL1B, IL6, IRF1, and IRF5, while M2 score considers the expression of CCL4, CCL13, CCL18, CCL20, CCL22, CD276, CLEC7A, CTSA, CTSB, CTSC, CTSD, FN1, IL4R, IRF4, LYVE1, MMP9, MMP14, MMP19, MSR1, TGFB1, TGFB2, TGFB3, TNFSF8, TNFSF12, VEGFA, VEGFB, and VEGFC\(^4\). Inflammatory score-related genes included IL1B, IL6, S100A8, and S100A9, while APOE, CD163, MAF, SELENOP, and SEPP1 were used to define the anti-inflammatory score\(^4\).

Glycolysis score-related genes included ALDOA, ALDOB, ALDOC, ENO1, ENO2, ENO3, GAPDH, GPI, HK1, HK2, HK3, LDHA, PFKL, PFKM, PFKP, PGAM1, PGAM4, PGK1, PKLR, PKM, SLC2A1, and TPI\(^8\). Hypoxia
score-related genes included \textit{ACOT7, ADM, ALDOA, CDKN3, ENO1, LDHA, MIF, NDRG1, P4HA1, PGAM1, SLC2A1, TPI1, TUBB6, and VEGFA}^{82}.

EMT score-related genes included \textit{ADAM12, ADAMTS12, ADAMTS2, AEBP1, ANGPTL2, ANTXR1, AP1G1, ATP8B1, AXL, BNC2, CALD1, CDH1, CDH2, CDS1, CGN, CLDN4, CMTM3, CNOT1, CNRIP1, COL10A1, COL1A1, COL1A2, COL3A1, COL5A1, COL5A2, COL6A1, COL6A2, COL6A3, COL8A1, CTNND1, DACT1, DYNC1L2, EMP3, ERBB3, ESRP1, ESRP2, F11R, FAP, FBN1, FN1, FSTL1, GALNT3, GPC6, GPR56, GRHL2, GYPC, HOOK1, HTRA1, INHBA, IRF6, ITGA11, LOXL2, LRRC15, MAP7, MARVELD2, MARVELD3, MMP2, MSRB3, MYO5B, NAP1L3, NID2, OCLN, OLFML2B, PCOLCE, PDGFRB, PMP22, POSTN, PRSS8, SPARC, SPINT1, SPOCK1, SULF1, SYT11, THBS2, VCAN, VIM and ZEB2}^{83}.

Proliferation score-related genes included \textit{AURKA, BUB1, CCNB1, CCND1, CCNE1, DEK, E2F1, FEN1, FOXM1, H2AFZ, HMGB2, MCM2, MCM3, MCM4, MCM5, MCM6, MKI67, MYBL2, PCNA, PLK1, TOP2A, TYMS, and ZWINT}^{84}.

Angiogenesis score-related genes included \textit{ANGPTL4, CXCL8, VCAN, and VEGFA}^{19}. Phagocytosis score-related genes included \textit{C1QB, CD163, MERTK, and MRC1}^{19}.

MHC-II score-related genes included \textit{HLA-DMA, HLA-DMB, HLA-DOA, HLA-DOB, HLA-DPA1, HLA-DPB1, HLA-DQA1, HLA-DQA2, HLA-DQB1, HLA-DQB2, HLA-DRA, HLA-DRB1, and HLA-DRB5}.

**Immune infiltration and macrophage markers**

The immune score calculated by the ESTIMATE algorithm was used to assess the overall immune infiltration of each sample\textsuperscript{85}. The immune cell marker \textit{CD45} was also used to assess the level of immune cell infiltration in tissues. CIBERSORT, a deconvolution algorithm based on transcriptional profiles, was used to calculate the proportions of 22 immune cells, including M0, M1, and M2 macrophages, in each sample\textsuperscript{86}. To understand the clinical features of SPP1 as a specific macrophage marker in CRC, some classic macrophage markers were selected for comparison; these included the pan-macrophage marker \textit{CD68}, M1 macrophage markers \textit{CD86} and \textit{iNOS}, and M2 macrophage markers \textit{CD163} and \textit{CD206}\textsuperscript{34,87}. In addition, the web server TIMER (http://timer.cistrome.org/) was used to evaluate the correlation of \textit{SPP1} with the six immune cell and macrophage markers mentioned above\textsuperscript{88}.

**Tumor mutation burden**

Simple nucleotide variation data from TCGA-CRC were used to calculate TMB, which is defined as the number of mutations per megabase. The visualization of the detailed gene mutation status of the high and low \textit{SPP1/CD68} groups was implemented in the R package "maftools"\textsuperscript{89}.

**Prediction of immunotherapy response**

We used the SubMap algorithm to predict the clinical response to anti-PD1 and anti-CTLA4 immunotherapy. SubMap was used to compare the similarity of different expression profiles; this feature
can reflect the response to treatment. The expression profiles and associated annotation data of 47 melanoma patients used to define high- and low-risk groups were obtained from the Supplementary Materials of Lu et al. In addition, samples in GSE39582 were divided into high and low SPP1 groups based on the median SPP1 expression.

**Survival analysis**

OS and PFI were regarded as the outcome events in the datasets (OS in GSE17536, GSE39582, and GSE17536; OS/PFI in TCGA). Survival analysis was performed based on the Kaplan–Meier algorithm. The log-rank test was used to calculate p values between groups. Univariate Cox models were constructed by Cox proportional hazards regression. These analyses were implemented with "survival" and "survminer" in R.

**Functional and pathway enrichment analysis**

In this study, four different gene sets were used for functional and pathway enrichment analysis, including Gene Ontology (GO) Biological Process (BP), Kyoto Encyclopedia of Genes and Genomes (KEGG), WikiPathways, and Hallmark gene sets.

Metascape (http://metascape.org/), a portal for gene function annotation analysis, was used for enrichment analysis of DEGs in this study (GO BP, KEGG, WikiPathways, and Hallmark gene sets).

Gene set variation analysis (GSVA) implemented in the "GSVA" package (version 1.44.2) was used for gene set enrichment analysis of different macrophage subsets (Hallmark gene sets).

LinkedOmics (http://www.linkedomics.org/) is an online source portal for analyzing multiomics data of 32 cancers in TCGA. In this study, LinkedOmics was used to perform GSEA of SPP1 expression profiles in TCGA-CRC (GO BP).

In addition, GSEA software (version 4.2.3) (https://www.gsea-msigdb.org/gsea/) was used to explore differences in functional enrichment between the L-SPP1/CD68 and H-SPP1/CD68 groups (Hallmark gene sets, permutations = 1000). Normalized enrichment score (NES) $\geq 1.0$, false discovery rate (FDR) $q$-val $\leq 0.25$, and $p$ value $< 0.05$ were considered statistically significant.

**Statistical analysis**

R (version 4.1.3), GraphPad Prism (version 9), Excel, SPSS (version 26), and Python (version 3.7) were used for statistical analysis. The Mann–Whitney U test was used to compare the differences between two groups. The Pearson or Spearman method was used for the correlation test. The log-rank method was used to calculate the P value in survival between groups. A chi-square test was used to compare clinical parameters between the high and low SPP1/CD68 groups. Data in bar plots of cell proportions for single-cell datasets are presented as the mean ± SEM (small sample size), while the mean ± SD is used for the remaining data. $P$ value $< 0.05$ was considered statistically significant.
Declarations

Code availability

The codes used in this study are available on Github (https://github.com/LPC19970117/SPP1-macrophages).

Data availability

The data used in this study are all from public datasets. Details can be found in the Methods section of the manuscript.

Acknowledgments

We would like to thank the researchers contributing to public databases and those providing data related to human medicine. We especially thank Dr. Wu for providing the CRLM data. This work was supported by the National Natural Science Foundation of China (82072655).

Conflict of interest

The authors declare no conflict of interest.

Author contributions

ZYX designed the study. ZYX and LRN collected the data. ZYX, FF and GZZ analyzed the data. ZYX, RKL, HJD, LLD, HZW, GMR and XYD visualized the data. ZYX and LRN drafted the manuscript. JZ and JYZ revised the manuscript. All authors have read and approved the final manuscript. All authors contributed to the article and approved the submitted version.

Supplementary information accompanies the manuscript on the Cell Discovery website http://www.nature.com/celldisc’

References


Figures

Figure 1

**ScRNA-seq revealed the myeloid cell landscape of CRC.**

a. Tissue types of colorectal cancer liver metastases, including CRC, NC, LM, and NL samples and PBMCs. CRC, colorectal cancer; NC, normal colorectum (adjacent colorectum); LM, liver metastases; NL, normal liver (adjacent); PBMC, peripheral blood mononuclear cell. NC and NL samples include both adjacent and healthy tissue.

b. Summary of the seven CRC single-cell datasets in this study. The green box indicates the sample origin of the CRC-Mix dataset. Numbers indicate sample size. Other cells, nonimmune cells including malignant cells and stromal cells; MMR, mismatch repair; MSI, microsatellite instability.

c. Barplots showing comparison of the proportion of myeloid cells to immune cells between the NC and CRC groups in six CRC scRNA-seq datasets. *, p < 0.05; ****, p < 0.0001. Sample size above bar; mean ± SEM; Mann–Whitney U test.

d. Tissue distribution preference of each myeloid subset estimated by OR.

e. UMAP plot of 12 myeloid subsets from the CRC-Mix dataset.
f. Representative genes and functional properties of myeloid subsets of the CRC-Mix dataset.

Figure 2

Changes in monocytes/macrophages during CRLM.

a. Volcano plot showing DEGs in monocytes/macrophages between NC and CRC samples in the GSE164552 dataset.

b. Venn diagrams showing the number of common DEGs that were either up- or downregulated across the six scRNA-seq datasets (NC vs. CRC). Representative common DEGs are shown in the right box.

c. Metascape bar graph showing functional enrichment analysis results for common DEGs (NC vs. CRC). The top 20 enriched terms are displayed. The red, blue, and green backgrounds indicate that the enriched terms are based on GO BP, KEGG, and WikiPathways gene sets, respectively.

d. Volcano plot showing DEGs in monocytes/macrophages between CRC and LM samples in the GSE164552 dataset.

e. Venn diagrams showing the number of common DEGs that were either up- or downregulated across the three scRNA-seq datasets (CRC vs. LM). Representative common DEGs are shown in the right box.

f. Metascape bar graph showing functional enrichment analysis of common DEGs (CRC vs. LM). The top 20 enriched terms are displayed.

g-i shows trajectory analysis of monocytes/macrophages (Wu et al.)

g. Cells were color-coded by monocyte/macrophage subset.

h. Cells were color-coded by tissue type. Arrows indicate the pseudotime direction.

i. Expression levels of monocyte/macrophage marker genes in pseudotime order. The color of the dots shows the tissue types of the cells.

j. Heatmap showing the GSVA score of 50 hallmark pathways of different monocyte/macrophage subsets (Wu et al.)
Figure 3

Tissue distribution and potential mechanisms of SPP1+ macrophages.

a. Bar graphs showing the ratio of SPP1+ macrophages to all macrophages in the six scRNA-seq datasets (NC vs. CRC). *, p < 0.05; **, p < 0.01; ***, p < 0.001; ****, p < 0.0001; ns, not significant. Sample size above bar; mean ± SEM; Mann–Whitney U test.
b. Bar graphs showing the ratio of SPP1+ macrophages to all macrophages in the three scRNA-seq datasets (CRC vs. LM).

c. Bar graphs showing the ratio of SPP1+ macrophages to all macrophages in different tissue types (CRC, NC, LM, NL and HCC) in the GSE164522 and GSE156625 datasets.

d. Pie charts show the proportions of four macrophage subsets in the CRC-Mix dataset.

e-g. The expression of (e) EMT, (f) glycolysis, and (g) hypoxia signatures in the three major macrophage subsets in the CRC-Mix dataset.

h. Heatmap showing the mean expression levels of the top 10 significant DEGs in the four macrophage subsets (CRC-Mix) (left). Exp, z score normalized mean expression. Metascape performs enrichment analysis of DEGs in each subset of macrophages, and the top five enriched terms based on GO BP, KEGG, and WikiPathways gene sets are displayed (right). Lollipop plot showing the expression of inflammatory and anti-inflammatory signatures in the four macrophage subsets (top).

i. The upper and right box plots show the expression of M1 and M2 signatures in the three major macrophage subsets, respectively (CRC-Mix). Scatter plot showing the correlation between the M1 and M2 signatures in macrophages. Pearson correlation and linear regression.

j. Heatmap showing cellular interactions among the four macrophage subsets and six major cell types (GSE178341). The redder the color is, the higher the predicted number of intercellular receptor–ligand pairs.

k. Receptor–ligand interactions among four macrophage subsets and six major cell types (GSE178341). P values are indicated by circle size; colors indicate the average expression levels of ligand and receptor or molecules in the corresponding cell types.
Figure 4

SPP1, a novel macrophage marker in CRC.

a. Heatmap showing the proportion of 22 immune cells (CIBERSORT), immune score (ESTIMATE), expression of markers (CD45, CD68, CD86, iNOS, CD163, CD206, and SPP1), CD68/CD45, and SPP1/CD68 in 10 bulk RNA-seq datasets (NC vs. CRC). The color of the bricks represents significance; the
redder the color is, the higher the value in the CRC compared to NC, and the opposite is true for blue. The expression of *SPP1* in GSE106582 shows an example of brick color meaning. Method (right), dataset source (top), sample size (bottom).

**b.** Correlation analysis between SPP1 expression and the proportion of immune cells in TCGA-CRC (COAD/READ) (CIBERSORT, n=647). Significantly positively and negatively correlated immune cells are shown in red and blue, respectively. Spearman correlation.

**c.** Violin plots showing the expression of *SPP1* in CD45- cells, other immune cells, other myeloid cells, and SPP1+ macrophages (GSE178341).

**d.** UMAP plot showing the distribution of CD45- cells, other immune cells, other myeloid cells, and SPP1+ macrophages (left). UMAP of iterative subsets of cells from the global level to immune cells to myeloid cells showing the enrichment of SPP1 expression levels in different cells (right).

**e.** Images of immunohistochemical staining showing SPP1 expression in NC, CRC, NL and HCC samples. Data from THPA. Scale bar, 200 μm.

**f.** Significantly enriched GO biological process terms in genes coexpressed with SPP1 in the TCGA-CRC cohort. FDR, false discovery rate.

**Figure 5**

**Clinical value of SPP1+ macrophages in CRC (bulk RNA-seq).**

**a.** Heatmap showing the immune score (ESTIMATE) and expression of markers (*CD45, CD68, CD86, iNOS, CD163, CD206, and SPP1*) in five bulk RNA-seq CRC datasets (stage I/II vs. stage III/IV). The color of the bricks represents significance; the redder the color is, the higher the value in the CRC compared to NC, and the bluer the color is, the lower the value. dataset source (top), sample size (bottom). *, GSE14333 uses the Duke stage.

**b.** Heatmap showing the immune score and expression of markers (*CD45, CD68, CD86, iNOS, CD163, CD206, and SPP1*) in three bulk RNA-seq CRC datasets (stage I/II vs. stage III/IV).

**c.** Heatmap showing the immune score and expression of markers (*CD45, CD68, CD86, iNOS, CD163, CD206, and SPP1*) in three bulk RNA-seq CRC datasets (N0 vs. stage N1-3). Six N3 samples were present in GSE39582.

**d.** Heatmap showing the immune score and expression of markers (*CD45, CD68, CD86, iNOS, CD163, CD206, and SPP1*) in three bulk RNA-seq CRC datasets (M0 vs. M1).
e. Heatmap showing the immune score and expression of markers (CD45, CD68, CD86, iNOS, CD163, CD206, and SPP1) in six bulk RNA-seq CRC datasets (CRC vs. LM).

f. Univariate Cox analysis of macrophage markers in four CRC datasets (TCGA, GSE17536, GSE39582, and GSE41258).

g. GSEA was performed between the high and low SPP1/CD68 groups, and the top 10 enriched hallmark terms in the H-SPP1/CD68 group are displayed. CI, confidence interval; GSEA, gene set enrichment analysis; HR, hazard ratio.

**Figure 6**

**Significance of SPP1+ macrophages for CRC treatment.**

a. Kaplan–Meier survival curve showing that SPP1/CD68 were associated with worse OS in TCGA-CRC cohorts (n = 593). Log-rank test.

b-c. Box plots showing the effect of preoperative chemotherapy on SPP1 expression in macrophages in the scRNA-seq cohorts of (B) GSE178318 and (C) Wu et al. R, responders; NR, nonresponders. Mann–Whitney U test.

d-i. The association between SPP1+ macrophages and genome instability and mutations.

d. Comparison of the ratio of SPP1+ macrophages to all macrophages in MSI-H versus MSS CRC (CRC-Mix scRNA-seq cohort).

e. Comparison of the ratio of SPP1+ macrophages to all macrophages in dMMR versus pMMR CRC (CRC-Mix scRNA-seq cohort).

f. Comparison of SPP1/CD68 in MSI-H versus MSS CRC (TCGA-CRC bulk RNA-seq cohort).

g. Comparison of SPP1 expression in MSI-H versus MSS CRC (GSE39582 bulk RNA-seq cohort).

h. The waterfall plot shows the mutation profiles of the top 30 mutated genes in the L-SPP1/CD68 group. Colors of annotations indicate different mutation types (bottom).

i. The waterfall plot shows the mutation profiles of the top 30 mutated genes in the H-SPP1/CD68 group.

j. SubMap analysis in TCGA-CRC showed that the H-SPP1/CD68 group was more sensitive to CTLA4 inhibitors (Bonferroni-corrected P = 0.023) and PD1 inhibitors (Bonferroni-corrected P = 0.021) than the L-SPP1/CD68 group.
SubMap analysis in GSE39582 showed that the H-SPP1 group was more sensitive to the CTLA4 inhibitor (Bonferroni-corrected $P = 0.023$) and PD1 inhibitor (Bonferroni-corrected $P = 0.024$).

The expression of $CSF1R$ in the three major macrophage subsets (CRC-Mix scRNA-seq cohort).

Figure 7

Schematic diagram of the SPP1+ macrophage model in CRC.
a. Classification of macrophage subsets in CRC. Macrophages in the scRNA-seq CRC cohort were divided into FCN1+ macrophages, C1QC+ macrophages, SPP1+ macrophages and MKI67+ macrophages. Inflammation, phagocytosis, malignancy, and proliferation were defined as core features of these four macrophage subsets.

b. Cell origin and tissue distribution of SPP1+ macrophage. Model of the developmental trajectory of monocyte/macrophage lineages in different sample types (PBMCs and NC, CRC, NL, LM and HCC tissues) from CRLM and HCC patients.

c. SPP1+ macrophages as a malignant marker of CRC. The number and proportion of SPP1+ macrophages continued to increase during the occurrence, progression and metastasis of CRC. The sizes of the blue, yellow and red dots represent the numbers of all immune cells, all macrophages, and SPP1+ macrophages in the sample, respectively.

d. Potential mechanisms of SPP1+ macrophages. SPP1+ macrophages exert biological functions through potentially related pathways (left). Red indicates that the signature is specifically elevated in SPP1+ macrophages compared to other macrophages, yellow indicates that the signature is elevated in SPP1+ macrophages compared to C1QC+ macrophages, and blue indicates that the signature is decreased in SPP1+ macrophages compared to C1QC+ macrophages. PD-L1 expression was elevated in SPP1+ macrophages compared to other macrophages. Crosstalk between SPP1+ macrophages and other cell subsets (right).

e. Implications of SPP1+ macrophages for CRC immunotherapy. Patients with a high proportion of SPP1+ macrophages had a worse prognosis, along with signatures related to genome instability and mutations (dMMR, MSI-H, and H-TMB). Immunotherapy has potential to improve outcomes in patients with a high proportion of SPP1+ macrophages, whereas targeting CSF1R is less effective.

Supplementary Files

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

- figS1.tif
- figS2.tif
- figS3.tif
- figS4.tif
- figS5.tif
- TableS1.xlsx
- TableS2.xlsx
- TableS3.xlsx
- TableS4.xlsx