



## FULL LENGTH ARTICLE

# *SPP1*<sup>+</sup> macrophages in colorectal cancer: Markers of malignancy and promising therapeutic targets

Zhenyu Xie <sup>a,1</sup>, Gaozan Zheng <sup>a,1</sup>, Liaoran Niu <sup>a,1</sup>, Kunli Du <sup>a</sup>,  
Ruikai Li <sup>a</sup>, Hanjun Dan <sup>a</sup>, Lili Duan <sup>a</sup>, Hongze Wu <sup>a</sup>,  
Guangming Ren <sup>b</sup>, Xinyu Dou <sup>b</sup>, Songchen Dai <sup>c,d</sup>, Fan Feng <sup>a,\*\*\*</sup>,  
Jian Zhang <sup>e,\*\*</sup>, Jianyong Zheng <sup>a,\*</sup>

<sup>a</sup> Department of Digestive Surgery, Xijing Hospital of Digestive Diseases, Medical University, Xi'an, Shaanxi 710032, China

<sup>b</sup> Xi'an Medical University, Xi'an, Shaanxi 710021, China

<sup>c</sup> Department of Surgical Oncology and General Surgery, The First Hospital of China Medical University, Shenyang, Liaoning 110016, China

<sup>d</sup> Key Laboratory of Precision Diagnosis and Treatment of Gastrointestinal Tumors, Ministry of Education, Shenyang, Liaoning 110016, China

<sup>e</sup> The State Key Laboratory of Cancer Biology, Department of Biochemistry and Molecular Biology, Air Force Medical University, Xi'an, Shaanxi 710032, China

Received 12 October 2023; received in revised form 19 March 2024; accepted 3 April 2024

Available online 30 May 2024

## KEYWORDS

Colorectal cancer;  
Immunotherapy;  
Single-cell RNA  
sequencing;  
Spatial  
transcriptomics;  
*SPP1*<sup>+</sup> macrophages

**Abstract** *SPP1*<sup>+</sup> macrophages have been identified as key players in the colorectal cancer (CRC) tumor microenvironment, but their function remains unclear. This study integrated single-cell and spatial transcriptomics with bulk sequencing to investigate the roles and mechanisms of *SPP1*<sup>+</sup> macrophages in CRC. Our findings revealed a pronounced elevation of *SPP1*<sup>+</sup> macrophages in CRC, especially within tumor territories. These macrophages served as markers for CRC initiation, progression, metastasis, and potential prognosis. Furthermore, they showed heightened transcriptional activity in genes linked to angiogenesis, epithelial–mesenchymal transition, glycolysis, hypoxia, and immunosuppression. *SPP1* protein amplified CRC cell migration and invasion, potentially mediating cellular crosstalk via the *SPP1*-*CD44*, *SPP1*-*PTGER4*, and *SPP1*-*a4b1* complex axes. Patients with a high proportion of *SPP1*<sup>+</sup> macrophages could benefit more from immune checkpoint blockade therapy. Interestingly, *CSF1R*

\* Corresponding author.

\*\* Corresponding author.

\*\*\* Corresponding author.

E-mail addresses: [surgeonfengfan@163.com](mailto:surgeonfengfan@163.com) (F. Feng), [biozhangj@fmmu.edu.cn](mailto:biozhangj@fmmu.edu.cn) (J. Zhang), [zhjy68@163.com](mailto:zhjy68@163.com) (J. Zheng).

Peer review under the responsibility of the Genes & Diseases Editorial Office, in alliance with the Association of Chinese Americans in Cancer Research (ACACR, Baltimore, MD, USA).

<sup>1</sup> These authors contributed equally to this work.

<https://doi.org/10.1016/j.gendis.2024.101340>

2352-3042/© 2024 The Authors. Publishing services by Elsevier B.V. on behalf of KeAi Communications Co., Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

expression was significantly enriched in *C1QC*<sup>+</sup> macrophages versus *SPP1*<sup>+</sup> macrophages, possibly explaining limited anti-CSF1R monotherapy effects. In conclusion, we propose an *SPP1*<sup>+</sup> macrophage model in CRC, highlighting such macrophages as a promising therapeutic target due to their malignancy markers.

© 2024 The Authors. Publishing services by Elsevier B.V. on behalf of KeAi Communications Co., Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

## Introduction

Colorectal cancer (CRC) is the third most common malignancy and the second leading cause of cancer-related death.<sup>1</sup> Following the great success of immune checkpoint blockade in various advanced solid tumors, interest in immunotherapy for CRC is increasing.<sup>2</sup> However, immunotherapy targeting PD-1 (programmed death-1) is only effective in patients with DNA mismatch repair deficiency/high levels of microsatellite instability, which is present in 5% of metastatic CRC cases.<sup>3–5</sup> 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 play a crucial role in the TME and are involved in various aspects of tumor immunity.<sup>6</sup> *In vitro*, M1 (inflammatory/antitumoral) and M2 (anti-inflammatory/protumoral) macrophage polarization systems have been used to classify the activation state of macrophages.<sup>7</sup> However, *in vivo*, macrophages exhibit a more complex phenotype, which contradicts this simplistic *in vitro* classification.<sup>8</sup> In the TME, macrophages are often referred to as tumor-associated macrophages, a heterogeneous cell type that contributes to angiogenesis,<sup>9</sup> extracellular matrix remodeling,<sup>6</sup> promotes epithelial–mesenchymal transition (EMT) of tumor cells,<sup>10</sup> and activates immunosuppression,<sup>11</sup> promoting the progression and metastasis of CRC.<sup>12</sup> Targeting tumor-associated macrophages via immunotherapies, such as those reducing the number of tumor-associated macrophages in the TME by blocking the *CSF1* (colony-stimulating factor 1)-*CSF1R* (colony-stimulating factor 1 receptor) axis,<sup>13–15</sup> has been applied in the clinic, but their efficacy as monotherapy in malignancies has been minimal.<sup>15,16</sup>

Single-cell studies have shown that *SPP1*<sup>+</sup> macrophages play a crucial role in the TME, and are associated with poor prognosis in CRC patients.<sup>17–20</sup> *SPP1* (secreted phosphoprotein 1), also known as osteopontin, is overexpressed in various cancer types<sup>21–23</sup> and is implicated in promoting CRC progression and metastasis through EMT and hypoxia pathways.<sup>24,25</sup> *SPP1* is known to bind to its receptors *CD44* and integrin, regulating several signaling pathways including ERK (extracellular signal-regulated kinase), JNK1 (c-Jun N-terminal kinase 1), and PI3K (phosphoinositide 3-kinase)/Akt (protein kinase B).<sup>26–31</sup> However, the functions and mechanisms of *SPP1*<sup>+</sup> macrophages remain unknown, and a comprehensive investigation of *SPP1*<sup>+</sup> macrophages is lacking.

Recent advances in single-cell RNA-sequencing (scRNA-seq) have provided new insights into the complexity of myeloid cells, particularly macrophages, in different tumor

types.<sup>17,19,32,33</sup> Single-cell evidence suggests that *SPP1*<sup>+</sup> macrophages play a key role in revealing the heterogeneity and functional changes of macrophages during CRC liver metastasis (CRLM).<sup>17,18</sup> Using public CRC single-cell datasets, we constructed the largest CRC myeloid cell atlas to date and conducted a systematic investigation of *SPP1*<sup>+</sup> macrophages. Based on the results, we proposed the *SPP1*<sup>+</sup> macrophage model, which can explain the clinical characteristics and functional changes of macrophages in CRC and help guide the development of macrophage-targeted immunotherapies.

## Materials and methods

### Material

The public datasets used in this study are from Gene Expression Omnibus (GEO) and The Cancer Genome Atlas (TCGA) and the website that stores the scRNA-seq data of Wu et al. These datasets include two spatial transcriptomics datasets, eight scRNA-seq datasets, and fifteen bulk RNA-seq datasets (Table 1). 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 the dataset of Wu et al) and one hepatocellular carcinoma (HCC) dataset (GSE156625) met this criterion. GSE164522 and the dataset of Wu et al were further defined as CRLM datasets based on tissue sample type, including CRC, normal colorectum (NC), liver metastasis (LM), and peripheral blood mononuclear cell (PBMC) samples. The bulk RNA-seq dataset included high-throughput sequencing data from TCGA-CRC and 14 GEO microarray datasets (GSE6988, GSE14333, GSE17536, GSE20842, GSE20916, GSE39582, GSE41258, GSE44076, GSE44861, GSE68468, GSE83889, GSE87211, GSE106582, 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/>).

### Basic analysis workflow for scRNA-seq data

The R package Seurat (version 4.1.0) converted raw unique molecular identifier count matrices to Seurat objects and then filtered the data according to the following criteria: i)

**Table 1** Sources of the ST, scRNA-seq, and bulk RNA-seq datasets.

Deposited data	Platform	Source	Identifier
Human NC ST dataset	10x Genomics	Fawcner-Corbett et al <sup>34</sup>	GEO: GSE158328
Human CRC ST dataset	10x Genomics	Wu et al <sup>18</sup>	<a href="http://www.cancerdiversity.asia/scCRLM">http://www.cancerdiversity.asia/scCRLM</a>
Human LM ST dataset	10x Genomics	Wu et al <sup>18</sup>	<a href="http://www.cancerdiversity.asia/scCRLM">http://www.cancerdiversity.asia/scCRLM</a>
Human CRC scRNA-seq dataset	10x Genomics	Lee et al <sup>35</sup>	GEO: GSE132465
Human CRC scRNA-seq dataset	10x Genomics	Lee et al <sup>35</sup>	GEO: GSE144735
Human CRC scRNA-seq dataset	10x Genomics	Zhang et al. <sup>19</sup>	GEO: GSE146771
Human CRC scRNA-seq dataset	10x Genomics	Che et al. <sup>36</sup>	GEO: GSE178318
Human CRC scRNA-seq dataset	10x Genomics	Pelka et al <sup>37</sup>	GEO: GSE178341
Human CRLM scRNA-seq dataset	10x Genomics	Liu et al <sup>17</sup>	GEO: GSE164522
Human CRLM scRNA-seq dataset	10x Genomics	Wu et al <sup>18</sup>	<a href="http://www.cancerdiversity.asia/scCRLM">http://www.cancerdiversity.asia/scCRLM</a>
Human HCC scRNA-seq dataset	10x Genomics	Sharma et al <sup>38</sup>	GEO: GSE156625
Human CRC bulk RNA-seq dataset	Illumina HiSeq	—	TCGA
Human CRC bulk RNA-seq dataset	GPL4811	Ki et al <sup>39</sup>	GEO: GSE6988
Human CRC bulk RNA-seq dataset	GPL570	Jorissen et al <sup>40</sup>	GEO: GSE14333
Human CRC bulk RNA-seq dataset	GPL570	Smith et al <sup>41</sup>	GEO: GSE17536
Human CRC bulk RNA-seq dataset	GPL4133	Gaedcke et al <sup>42</sup>	GEO: GSE20842
Human CRC bulk RNA-seq dataset	GPL570	Skrzypczak et al <sup>43</sup>	GEO: GSE20916
Human CRC bulk RNA-seq dataset	GPL570	Marisa et al <sup>44</sup>	GEO: GSE39582
Human CRC bulk RNA-seq dataset	GPL96	Sheffer et al <sup>45</sup>	GEO: GSE41258
Human CRC bulk RNA-seq dataset	GPL13667	Sole et al <sup>46</sup>	GEO: GSE44076
Human CRC bulk RNA-seq dataset	GPL3921	Ryan et al <sup>47</sup>	GEO: GSE44861
Human CRC bulk RNA-seq dataset	GPL96	Sheffer et al <sup>45</sup>	GEO: GSE68468
Human CRC bulk RNA-seq dataset	GPL10558	Kwon et al <sup>48</sup>	GEO: GSE83889
Human CRC bulk RNA-seq dataset	GPL13497	Hu et al <sup>49</sup>	GEO: GSE87211
Human CRC bulk RNA-seq dataset	GPL10558	—	GEO: GSE106582
Human CRC bulk RNA-seq dataset	GPL570	Szeglin et al <sup>50</sup>	GEO: GSE161158

Note: ST, spatial transcriptomics; HCC, hepatocellular carcinoma; scRNA-seq, single-cell RNA-sequencing; RNA-seq, RNA-sequencing; CRC, colorectal cancer; NC, normal colorectum (adjacent colorectum); LM, liver metastases; CRLM, CRC liver metastasis.

filtering genes expressed in at least 10 cells; ii) filtering cells with unique feature counts over 6,000 or less than 500; and iii) filtering cells with >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 the percent mitochondrial content. The results were obtained by principal component analysis 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 was completed by the nonlinear dimensionality reduction UMAP method.<sup>51</sup> Finally, the main cell types were identified and annotated based on canonical marker genes. Subsequent annotation of myeloid cell subsets followed the same workflow described above for major cell types.

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

The primary methodology for categorizing clusters into specific major cell types and myeloid cell subsets in this research was based on the expression level of essential marker genes within these clusters.

Myeloid cells were defined based on the similarity of cell clusters rather than strict adherence to biological definitions. As a result, in the cluster annotation of major cell types, real myeloid cells such as neutrophils and mast cells were named independently, while plasmacytoid dendritic cells of lymphoid origin,<sup>52</sup> Kupffer cells derived from embryonic yolk sac,<sup>53</sup> and even doublets were all classified as myeloid cells.

Given that CRC macrophages were the focus of this study, the major monocyte/macrophage subpopulation in PBMCs of the CRLM dataset was broadly defined as monocytes. In datasets without PBMC samples, we did not define the monocyte subset. Due to the high similarity in the transcriptome characteristics of monocytes and Macro-FCN1, potential monocytes in these datasets were usually classified into the Macro-FCN1 subset. Similarly, in datasets without NL samples, we did 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 correction for batch effects using the harmony algorithm,<sup>54</sup> the CRC-Mix data were reclustered into groups and annotated.

In addition, to compare the differences in *SPP1*<sup>+</sup> macrophage proportions in CRLM and HCC, myeloid cells of GSE156625 and GSE164522 were integrated in the same way, subjected to batch effect correction 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 differentially expressed genes (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: i) adjusted  $P$  value  $< 0.01$ ; ii) log fold-change of the average expression  $> 0.5$ ; iii) pct.1 (percentage of gene expression detected in the first group)  $> 0.25$ ; iv) 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 odds ratios in seven CRC single-cell datasets. The odds ratio  $> 1$  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 CRLM, we used the “*Monocle*” package (version 2.22.0)<sup>55</sup> 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<sup>56</sup> to evaluate the interactions of macrophages with other cells.

The putative ligands and receptors were determined based on whether they were expressed in each cell. To efficiently assess the closeness of cell interactions based on the number of receptor–ligand pairs, we randomly sampled 1000 cells per population from the Macro-C1QC, Macro-FCN1, Macro-MKI67, Macro-SPP1, and major cell types in the GSE178318 dataset.

### Definition of 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 for M1 and M2 macrophages were defined as the M1 score and M2 score, respectively<sup>57</sup> (Table S1). Similarly, the inflammatory score, anti-inflammatory score,<sup>35</sup> glycolysis score,<sup>58</sup> hypoxia score,<sup>58</sup> EMT score,<sup>59</sup> proliferation score,<sup>60</sup> angiogenesis score,<sup>19</sup> phagocytosis score,<sup>19</sup> and MHC-II (major histocompatibility complex class II) score were defined based on their respective sets of genes (Table S1).

### Immune infiltration and macrophage markers

The immune score calculated by the ESTIMATE algorithm was used to assess the overall immune infiltration of each sample.<sup>61</sup> The immune cell marker *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.<sup>62</sup> 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 *CD68*, M1 macrophage markers *CD86* and *iNOS* (inducible nitric oxide synthase), and M2 macrophage markers *CD163* and *CD206*.<sup>63,64</sup> In addition, the web server TIMER (<http://timer.cistrome.org/>) was used to evaluate the correlation of *SPP1* with the six immune cells and macrophage markers mentioned above.<sup>65</sup>

### Total RNA extraction and quantitative reverse transcription PCR

Fifty pairs of CRC samples and paratumor tissues were obtained from CRC patients who underwent surgery at Xijing Hospital. Total RNA was extracted from the tissues using Trizol reagent (Invitrogen, Waltham, MA, United States), and cDNA was synthesized by reverse transcription using the PrimeScript RT Reagent Kit (TaKaRa, Tokyo, Japan). Quantitative real-time PCR was performed to confirm the expression levels of *SPP1*, using the SYBR Premix Ex Taq II Kit (TaKaRa, Tokyo, Japan). *GAPDH* was used as an internal control for normalization. The relative mRNA expression was calculated using the  $2^{-\Delta\Delta Ct}$  method. The primer sequences used were as follows: *SPP1* forward: 5'-GAAGTTTCGCAGACCTGACAT-3'; *SPP1* reverse: 5'-GTATG-CACCATTCAACTCCTCG-3'; *GAPDH* forward: 5'-



GACAGTCAGCCGCATCTTCT-3'; *GAPDH* reverse: 5'-GCGCCCAATACGACCAAATC-3'.

## Spatial transcriptomic analysis

The “*SCTransform*” function of Seurat was used to standardize the spatial transcriptomic data of NC, CRC, and LM samples. “*RunPCA*” and “*RunUMAP*” were used for dimensionality reduction and clustering (principal components = 15, resolution = 0.8). After merging similar clusters, we identified spatial transcriptomics-defined regions in NC, CRC, and LM samples. The “*FindAllMarkers*” function was used to analyze the specific genes of each region in the spatial transcriptome (adjusted *P*-value < 0.01, |log<sub>2</sub> fold change| > 0.4). Subsequently, we assessed the distribution of immune cells in spatial transcriptomic regions of CRC and LM samples using multimodal intersection analysis,<sup>66</sup> which determined the cell type enrichment degrees by performing a hypergeometric test on the overlap between the cell type-specific genes of the scRNA-seq data and the region-specific genes of the spatial transcriptomics data (the scRNA-seq and spatial transcriptomics data of CRC and LM samples were both from the dataset of Wu et al). To explore the functional characteristics of tumor regions, region-specific genes in CRC and LM samples were further used for functional enrichment analysis. In addition, we used the “*AddModuleScore*” function of Seurat to evaluate the spatial distribution of glycolysis and hypoxia signatures.

## Immunofluorescence staining

Human tissue specimens were obtained from Xijing Hospital under the approval of the Institutional Review Board. The CRC-paired specimens were collected within 30 min after the tumor resection and fixed in paraformaldehyde for 48 h. Dehydration and embedding in paraffin were carried out using standard procedures. To inhibit endogenous peroxidase activity, the specimens were treated with 3% H<sub>2</sub>O<sub>2</sub> for 25 min. Following preincubation with 10% normal goat serum for blocking nonspecific binding sites for 30 min, the tissue sections were incubated at 4 °C with primary antibodies in a humidified chamber overnight. The primary antibodies used for validating SPP1<sup>+</sup> macrophages were as follows: mouse anti-human CD68 (Proteintech, Cat# 66231-2-Ig, 1:2000) and mouse anti-human SPP1 (Santa Cruz Biotechnology, Cat# sc-21742, 1:1000). After thorough washing, the sections were mounted with an anti-fading reagent and coverslips were applied. Fluorescence microscopy (NIKON ECLIPSE C1) was used to visualize the images, and subsequent analysis was performed using CaseViewer software.

## Tumor mutation burden

Simple nucleotide variation data from TCGA-CRC were used to calculate tumor mutation burden, which was defined as the number of mutations per megabase. The visualization of the detailed gene mutation status of the high and low SPP1/CD68 groups was implemented in the R package “maftools”.<sup>67</sup>

## Prediction of immunotherapy response

We used the SubMap algorithm to predict the clinical response to anti-PD-1 and anti-CTLA4 (cytotoxic T-lymphocyte associated protein 4) immunotherapy. SubMap was used to compare the similarity of different expression profiles. This feature can reflect the response to treatment.<sup>68</sup> 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.<sup>69</sup> In addition, samples in GSE39582 were divided into high and low SPP1 groups based on the median SPP1 expression.

## Survival analysis

Overall survival and progression-free interval were regarded as the outcome events in the datasets (overall survival in GSE17536, GSE39582, and GSE41258; overall survival/progression-free interval 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

Functional and pathway enrichment analysis was conducted using four different gene sets including Gene Ontology (GO) Biological Process (BP), Kyoto Encyclopedia of Genes and Genomes (KEGG), WikiPathways, and Hallmark gene sets.<sup>70</sup>

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).<sup>71</sup> 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).<sup>72</sup> LinkedOmics (<http://www.linkedomics.org/>) is an online source portal for analyzing multiomics data of 32 cancers in TCGA.<sup>73</sup> In this study, LinkedOmics was used to perform Gene Set Enrichment Analysis (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 high (H-SPP1/CD68) and low (L-SPP1/CD68) groups (Hallmark gene sets, permutations = 1000).<sup>74</sup>

## Statistical analysis

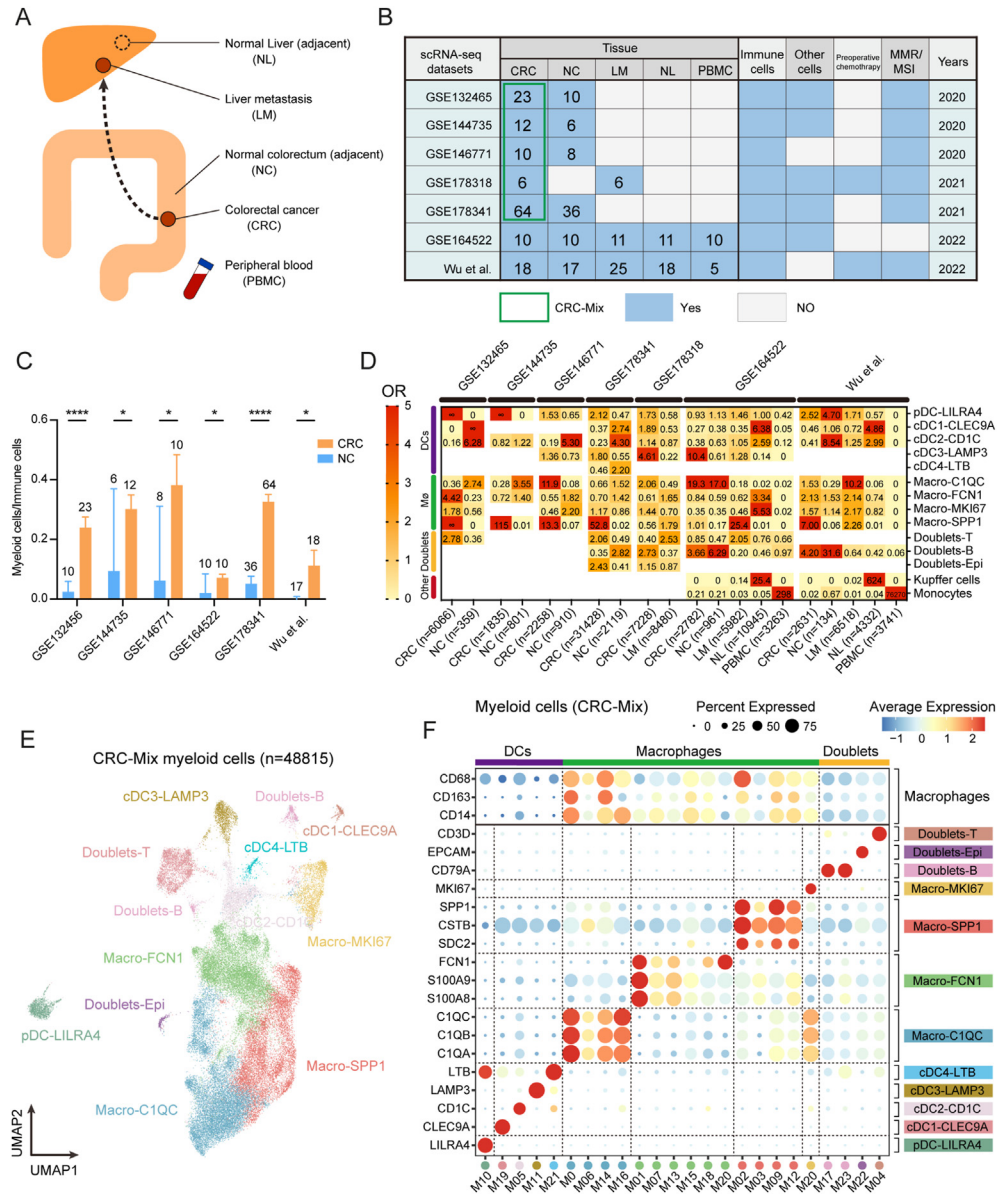
R (version 4.1.3), GraphPad Prism (version 9), Excel, SPSS (version 26), and Python (version 3.7) were used for statistical analysis. The Mann–Whitney U test was used to compare the differences between two groups, and the Wilcoxon test was used for comparing paired samples. The Pearson or Spearman method was used for the correlation test. The log-rank method was used to calculate the *P*-values in survival analysis 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 were presented as mean  $\pm$  standard error of the mean (small sample size), while mean  $\pm$  standard deviation was used for the remaining data.  $P$ -values  $<0.05$  were considered statistically significant.

## Results

### Identification of myeloid cell subsets in CRC

This study involved analysis of tissue types including CRC, NC, LM, NL, and PBMCs (Fig. 1A). CRC single-cell data were



**Figure 1** Identification of myeloid cell subsets in CRC. (A) Tissue types of CRLM, including CRC, NC, LM, and NL samples and PBMCs. 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 refer to non-immune cells including malignant cells and stromal cells. (C) Bar plots of the comparison of the proportion of myeloid cells to immune cells between the NC and CRC groups in six CRC scRNA-seq datasets. The sample size is shown above the bar. \* $P < 0.05$ , \*\*\*\* $P < 0.0001$ ; mean  $\pm$  standard error of the mean; Mann-Whitney U test. (D) Tissue distribution preference of each myeloid subset estimated by odds ratios (ORs). (E) UMAP plot of 12 myeloid subsets from the CRC-Mix dataset. (F) Dot plot of the scaled average expression and percentage of expression for key marker genes per myeloid cluster in the CRC-Mix dataset. CRC, colorectal cancer; CRLM, colorectal cancer liver metastases; NC, normal colorectum (adjacent colorectum); LM, liver metastases; NL, normal liver (adjacent); PBMCs, peripheral blood mononuclear cells; MMR, mismatch repair; MSI, microsatellite instability.

obtained from seven publicly available datasets (GSE132465, GSE144735, GSE146771, GSE164522, GSE178318, GSE178341, and the dataset of Wu et al) (Fig. 1B). Nine major cell types were identified in these datasets, including T cells (*CD3D*), natural killer cells (*KLRF1*), B cells (*MS4A1*), plasma cells (*MZB1*), myeloid cells (*LYZ*), mast cells (*TPSAB1*), neutrophils (*FCGR3B*), epithelial cells (*EPCAM*), and stromal cells (*DCN*) (Fig. S1A, B). Notably, the proportion of myeloid cells among all immune cells was significantly higher (6/6) in CRC samples compared with NC samples (Fig. 1C and Table S2), suggesting an important role for myeloid cells in the CRC tumor immune microenvironment.

We defined 14 distinct myeloid cell subsets, consisting of one plasmacytoid dendritic cell subset (pDC-*LILRA4*, defined by *LILRA4* expression), four conventional DC subsets (cDC1-*CLEC9A*, cDC2-*CD1C*, cDC3-*LAMP3*, and cDC4-*LTB*), four macrophage subsets (Macro-*C1QC*, Macro-*FCN1*, Macro-*MKI67*, and Macro-*SPP1*), and three doublet subsets (Doublets-B, Doublets-Epi, and Doublets-T), as well as the monocyte subset (*CD14*) and Kupffer cells (*MARCO*) (Fig. S1C; Table S2). We then examined the distribution of these myeloid cell subsets in each of the seven cohorts using odds ratios (Fig. 1D). Our analysis revealed that the Macro-*SPP1* subset was enriched in CRC samples.

To create a more accurate model of the CRC myeloid cell lineage, we integrated CRC myeloid cells from five datasets (GSE132465, GSE144735, GSE146771, GSE178318, and GSE178341) into a large cohort ( $n = 48,815$ ) called CRC-Mix (Fig. S1D), which was then reclustered and annotated after batch effect correction. This resulted in the identification of 12 myeloid cell subsets in CRC tissue (Fig. 1E, F; Fig. S1E), along with a summary table of representative genes for aiding other researchers in identifying these subsets (Fig. S1F). Further details on DEGs between subsets can be found in Table S3.

## Changes in monocytes/macrophages during CRLM

We used single-cell sequencing to examine changes in the transcriptional profile and functional pathways of monocytes/macrophages during CRLM. We identified DEGs of monocytes/macrophages between NC and CRC samples (GSE164522) and in six CRC single-cell datasets (NC vs. CRC) (Fig. 2A, B and Table S4). Enrichment analysis of the common DEGs ( $n = 216$ ) revealed angiogenesis-, EMT-, and hypoxia-related pathway changes in monocytes/macrophages during CRC occurrence (Fig. 2C). Similarly, we identified DEGs of monocytes/macrophages between CRC and LM samples (GSE164522) and in three CRC single-cell datasets (CRC vs. LM) (Fig. 2D, E and Table S5), with enrichment analysis of the common DEGs ( $n = 1383$ ) revealing angiogenesis-, EMT-, and hypoxia-related pathway changes in monocytes/macrophages during liver metastasis of CRC (Fig. 2F).

We used Monocle2 to observe trajectories of monocyte/macrophage subsets, marker genes, and tissue types during CRLM (Fig. 2G–I). The results showed that monocytes differentiated into *FCN1*<sup>+</sup> macrophages, which polarized towards *C1QC*<sup>+</sup> and *SPP1*<sup>+</sup> macrophages. The expression of *SPP1* was elevated in CRC and LM-concentrated regions, suggesting their important role in CRC metastasis.

Furthermore, GSVA analysis showed significant enrichment of angiogenesis-, EMT-, glycolysis-, and hypoxia-related genes in *SPP1*<sup>+</sup> macrophages (Fig. 2J; Fig. S2A–F). This enrichment suggests that *SPP1*<sup>+</sup> macrophages may be a key subset contributing to the alteration of angiogenesis-, EMT-, and hypoxia-related pathways in monocytes/macrophages during the CRLM process.

## Characteristics and functions of *SPP1*<sup>+</sup> macrophages in CRC

Our study found that *SPP1*<sup>+</sup> macrophages increased in proportion during CRC development and liver metastases, as shown by scRNA-seq data from multiple datasets (Fig. 3A, B and Table S2). This suggests their critical role in CRLM. We also integrated and reclustered CRLM and HCC macrophage data and found that the proportion of *SPP1*<sup>+</sup> macrophages was significantly lower in HCC than in CRLM samples, indicating that the elevation of *SPP1*<sup>+</sup> macrophages is cancer-type specific (Fig. 3C and Table S2).

We analyzed the CRC-Mix dataset to investigate the function of *C1QC*<sup>+</sup>, *FCN1*<sup>+</sup>, *SPP1*<sup>+</sup>, and *MKI67*<sup>+</sup> macrophages in CRC (Fig. 3D). *MKI67*<sup>+</sup> macrophages were excluded due to their low proportion and well-defined proliferation function (Fig. S3A). *C1QC*<sup>+</sup> macrophages were found to have the highest phagocytosis and MHC-II score (Fig. S3B–E), supporting their role in antigen presentation and phagocytosis of pathogens. *FCN1*<sup>+</sup> macrophages showed higher angiogenesis scores and lower phagocytosis scores than *SPP1*<sup>+</sup> macrophages (Fig. S3B). *SPP1*<sup>+</sup> macrophages exhibited specific elevation of EMT, glycolysis, and hypoxia scores, consistent with our GSVA results (Fig. S3E–G).

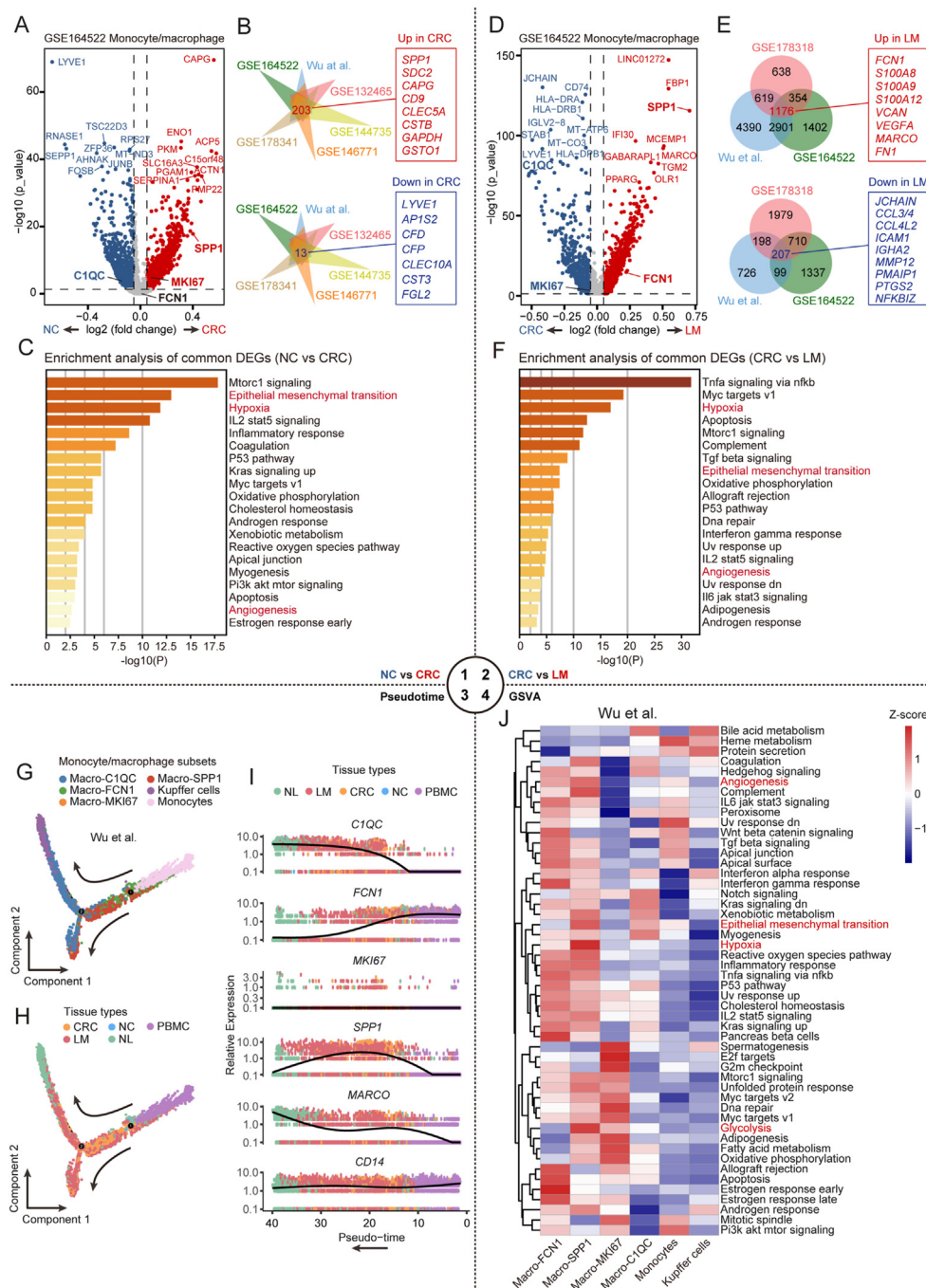
We assessed the inflammatory, anti-inflammatory, M1, and M2 signatures of different macrophage subsets in CRC-Mix and found that *FCN1*<sup>+</sup> macrophages had the highest inflammatory score and the lowest anti-inflammatory score, with higher M1 scores and lower M2 scores than *C1QC*<sup>+</sup> macrophages and *SPP1*<sup>+</sup> macrophages (Fig. S3E, F). In contrast, *C1QC*<sup>+</sup> macrophages and *SPP1*<sup>+</sup> macrophages exhibited more anti-inflammatory features, with *SPP1*<sup>+</sup> macrophages specifically expressing *CD274* (PD-L1, programmed cell death ligand 1) and *HLA-G* (human leukocyte antigen G),<sup>75</sup> suggesting an immunosuppressive characteristic (Fig. S3D–F). Additionally, we observed a positive correlation between the M1 signature and the M2 signature in CRC macrophages (Fig. S3F), consistent with findings in multiple tumor scRNA-seq studies.<sup>76</sup>

CellPhoneDB was used to identify ligand–receptor interactions between *SPP1*<sup>+</sup> macrophages and other major cell types in CRC. *SPP1*<sup>+</sup> macrophages interacted with stromal cells, epithelial cells, and T cells through specific ligand–receptor pairs, including *SPP1-CD44*, *SPP1-PTGER4*, and *SPP1-a4b1* complex (Fig. 3H, I). *SPP1* is a core molecule for their biological functions, and *CD44*, *PTGER4* (prostaglandin E receptor 4), and integrin  $\alpha 4\beta 1$  are enriched in different cell types.

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

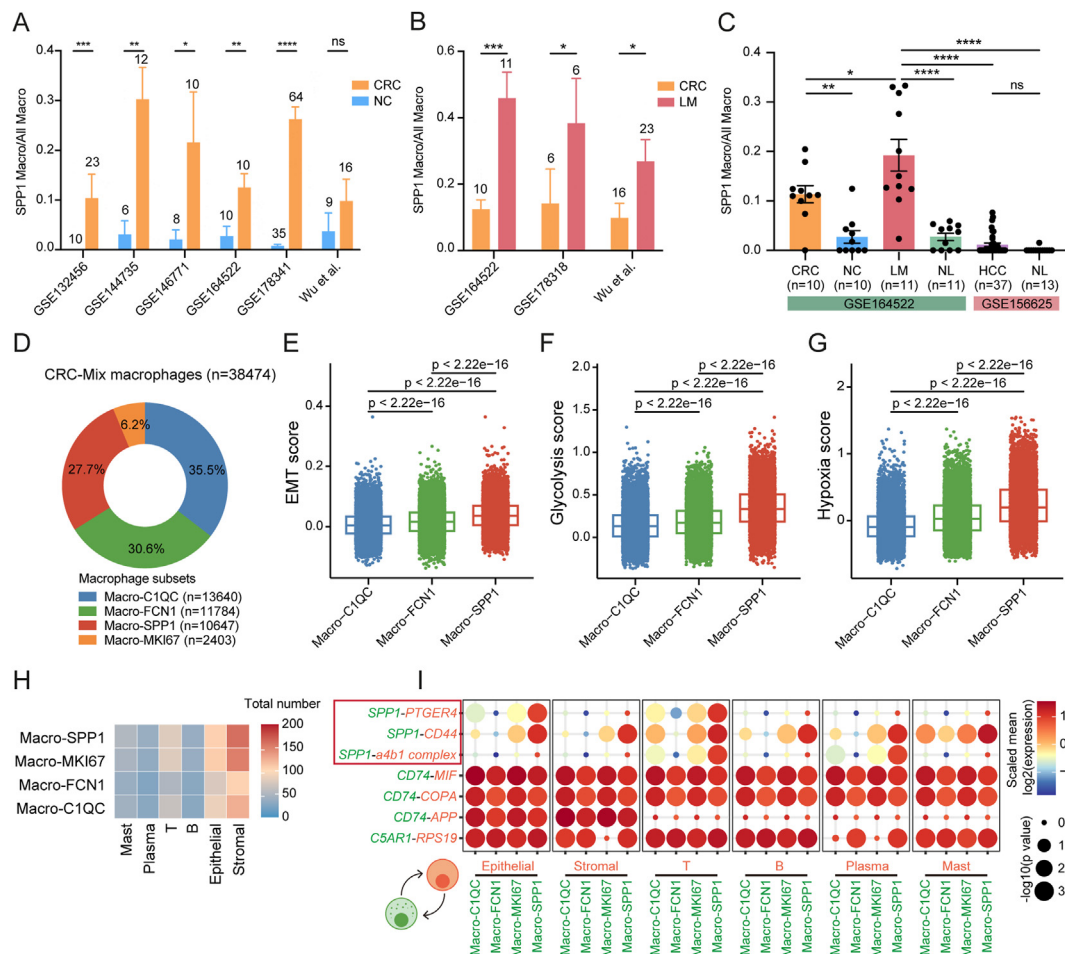
Bulk RNA-seq datasets were used to investigate changes in macrophage proportions and markers during CRC.





**Figure 2** Changes in monocytes/macrophages during CRLM. **(A)** Volcano plot of DEGs in monocytes/macrophages between NC and CRC samples in the GSE164522 dataset. **(B)** Venn diagrams of common DEGs up- or down-regulated across the six scRNA-seq datasets (NC vs. CRC). **(C)** Metascape bar graph of functional enrichment analysis results for common DEGs (NC vs. CRC). The top 20 enriched hallmark terms are displayed. **(D)** Volcano plot of DEGs in monocytes/macrophages between CRC and LM samples in the GSE164522 dataset. **(E)** Venn diagrams of common DEGs up- or down-regulated across the three scRNA-seq datasets (CRC vs. LM). **(F)** Metascape bar graph of functional enrichment analysis results for common DEGs (CRC vs. LM). The top 20 enriched hallmark terms are displayed. **(G–I)** Trajectory analysis of monocytes/macrophages (Wu et al), including color-coded subsets (G), tissue types (H), and gene expression levels (I). **(J)** Heatmap of GSVA (Gene Set Variation Analysis) scores of 50 hallmark pathways for different monocyte/macrophage subsets (Wu et al). CRC, colorectal cancer; CRLM, colorectal cancer liver metastases; NC, normal colorectum (adjacent colorectum); DEGs, differentially expressed genes; LM, liver metastases.





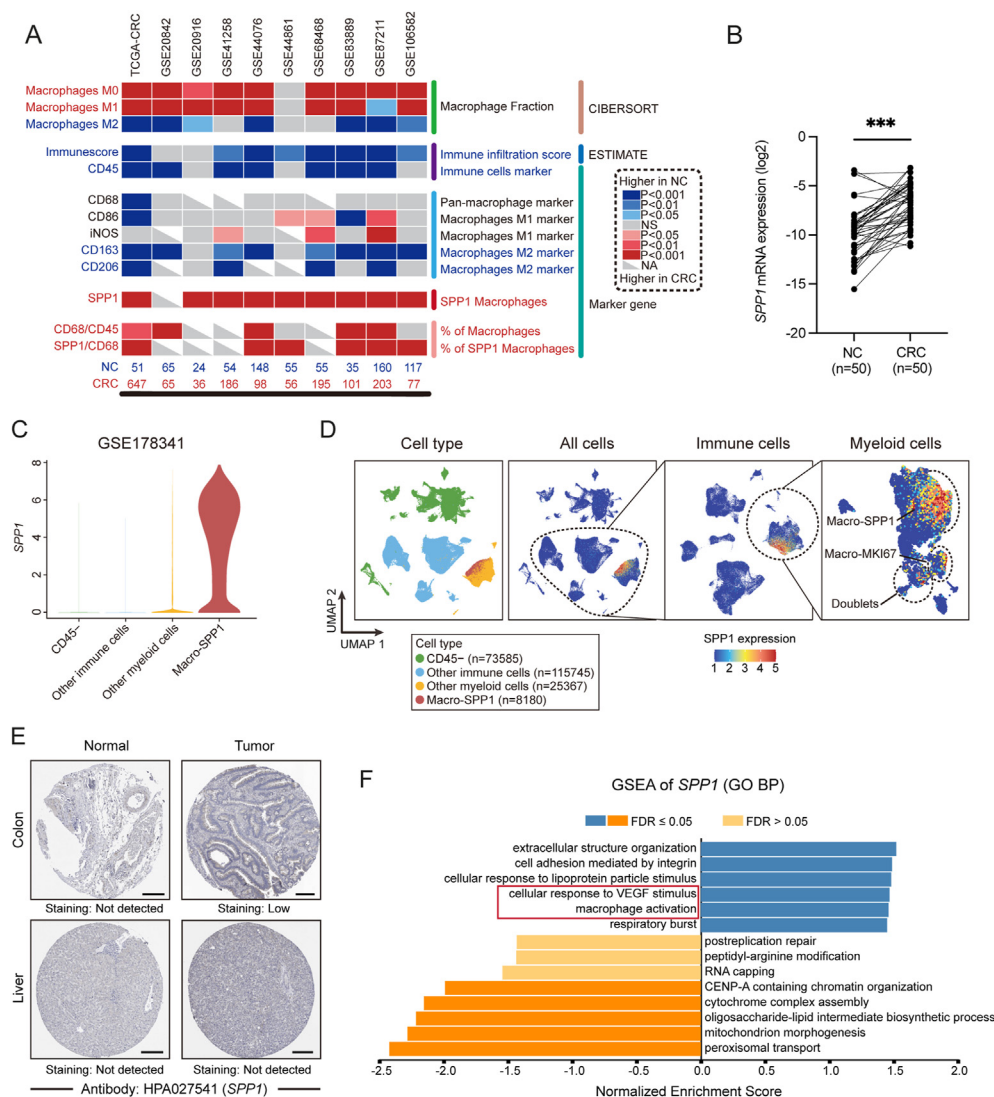
**Figure 3** Tissue distribution and potential mechanisms of SPP1<sup>+</sup> macrophages. (A) Bar graphs of the ratio of SPP1<sup>+</sup> macrophages to all macrophages in the six scRNA-seq datasets (NC vs. CRC). The sample size is shown above the bar. Mean ± standard error of the mean; \**P* < 0.05, \*\**P* < 0.01, \*\*\**P* < 0.001, \*\*\*\**P* < 0.0001; ns, not significant; Mann–Whitney U test. (B) Bar graphs of the ratio of SPP1<sup>+</sup> macrophages to all macrophages in the three scRNA-seq datasets (CRC vs. LM). (C) Bar graphs of the ratio of SPP1<sup>+</sup> macrophages to all macrophages in different tissue types (CRC, NC, LM, NL, and HCC) in the GSE164522 and GSE156625 datasets. (D) Pie charts of 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 of cellular interactions among the four macrophage subsets and six major cell types (GSE178341). (I) 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 ligands and receptors or molecules in the corresponding cell types. SPP1, secreted phosphoprotein 1; CRC, colorectal cancer; NC, normal colorectum (adjacent colorectum); LM, liver metastases; NL, normal liver (adjacent); HCC, hepatocellular carcinoma; EMT, epithelial–mesenchymal transition.

CIBERSORT algorithm analysis revealed a significant increase in the proportions of M0 and M1 macrophages and a decrease in M2 macrophages (NC vs. CRC) (Fig. 4A). Expression levels of classical macrophage markers showed significant down-regulation of M2 markers *CD163* and *CD206*, while pan-macrophage marker *CD68* and M1 markers *CD68* and *iNOS* remained unchanged.

To reconcile the seemingly inconsistent results on changes in the proportion and number of macrophages during CRC development, we used the ESTIMATE algorithm and *CD45* (*PTPRC*) immune cell marker to assess immune infiltration levels in NC and CRC tissues (Fig. 4A). The results indicated a significant decrease in immune infiltration

in CRC tissues, leading to a higher *CD68/CD45* ratio and a significant increase in the proportion of macrophages.

SPP1 and SPP1/*CD68* were found to be significantly up-regulated in CRC (Fig. 4A). Quantitative real-time PCR analysis of 50 paired samples also confirmed an elevated mRNA expression of SPP1 in CRC (*P* < 0.001) (Fig. 4B). Correlation analysis indicated a strong correlation between SPP1 expression and the infiltration of macrophages (Fig. S4B), particularly M2 macrophages, in CRC. Furthermore, single-cell sequencing data demonstrated that SPP1 expression was exclusively enriched in specific myeloid cells in CRC (Fig. 4C, D), indicating that SPP1 is a specific marker for SPP1<sup>+</sup> macrophages in CRC.



**Figure 4** *SPP1* is a novel macrophage marker in CRC. **(A)** Heatmap of the proportion of M0, M1, and M2 macrophages (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 gradient of each tile signifies changes in the marker expression level, where red represents an increase in CRC, blue represents a decrease, and the depth of the color reflects the statistical significance. Additional information provided includes the analysis method (right), dataset source (top), and the sample sizes of NC and CRC for each dataset (bottom). **(B)** Quantitative real-time PCR analysis of *SPP1* mRNA expression between 50 paired NC and CRC samples. Wilcoxon test; \*\*\* $P < 0.001$ . **(C)** Violin plots of the expression of *SPP1* in *CD45*<sup>+</sup> cells, other immune cells, other myeloid cells, and *SPP1*<sup>+</sup> macrophages (GSE178341). **(D)** UMAP plot of the distribution of *CD45*<sup>+</sup> cells, other immune cells, other myeloid cells, and *SPP1*<sup>+</sup> macrophages (left). UMAP of iterative subsets of cells from the global level to immune cells to myeloid cells shows the enrichment of *SPP1* expression levels in different cells (right). **(E)** Images of immunohistochemical staining showing *SPP1* protein expression in NC, CRC, NL, and HCC samples. Data are from The Human Protein Atlas. Scale bar, 200  $\mu$ m. **(F)** Significantly enriched GO BP (Gene Ontology Biological Process) terms in genes coexpressed with *SPP1* in the TCGA-CRC cohort. *SPP1*, secreted phosphoprotein 1; CRC, colorectal cancer; NC, normal colorectum (adjacent colorectum); NL, normal liver (adjacent); HCC, hepatocellular carcinoma; iNOS, inducible nitric oxide synthase; FDR, false discovery rate.

The specificity of *SPP1* as a marker for *SPP1*<sup>+</sup> macrophages in CRC was further confirmed by immunohistochemistry staining from The Human Protein Atlas (THPA) database,<sup>77</sup> which showed specific expression of *SPP1* protein in CRC samples but not in NC, NL, and HCC samples (Fig. 4E). GSEA of *SPP1* coexpressed genes in TCGA-

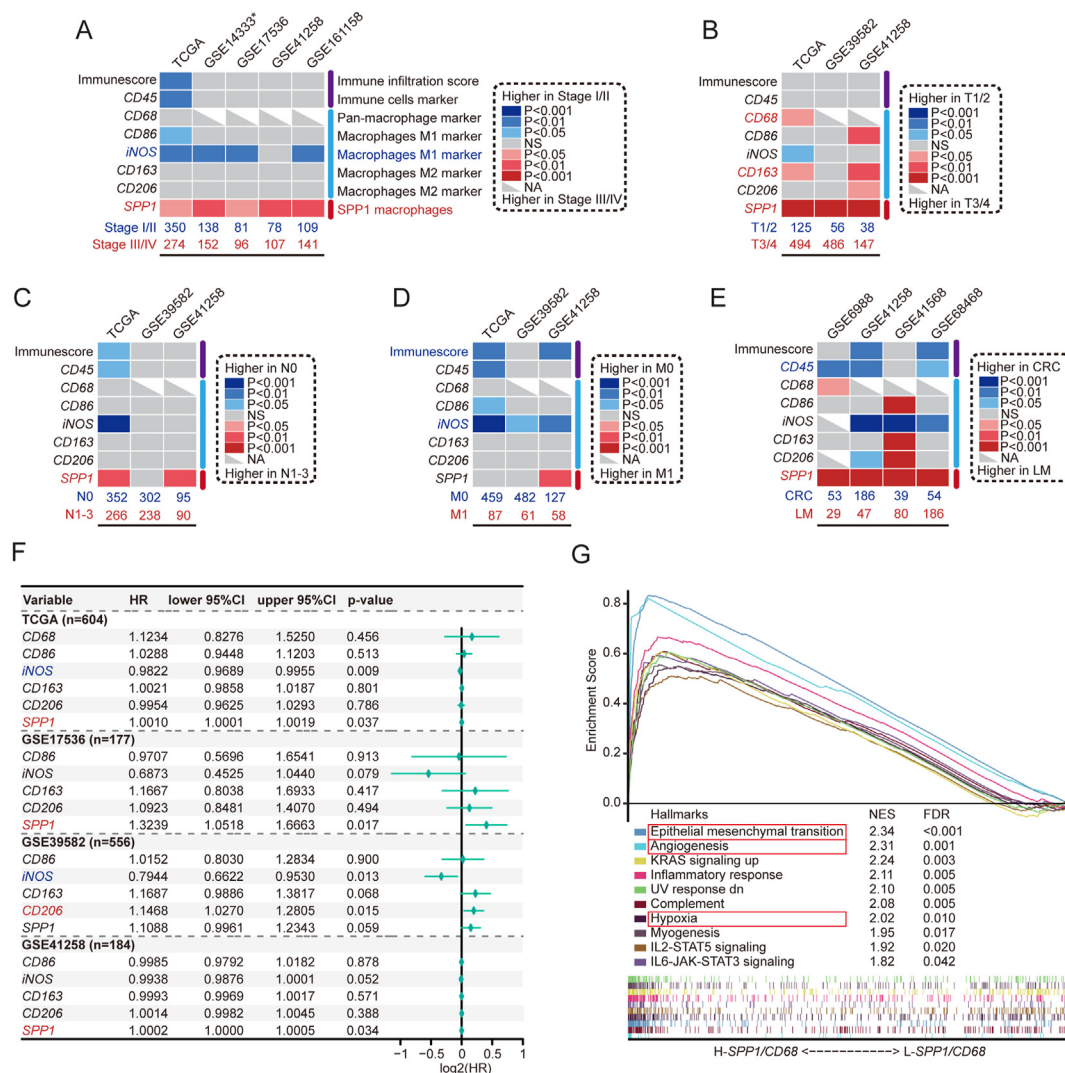
CRC using LinkedOmics revealed that extracellular structure organization, cell adhesion mediated by integrin, cellular response to lipoprotein particle stimulus, cellular response to VEGF stimulus, and macrophage activation were among the top significantly enriched GO BP terms (Fig. 4F).

In summary, *SPP1* is a specific biomarker of *SPP1*<sup>+</sup> macrophages and is significantly elevated in CRC.

### Role of *SPP1*<sup>+</sup> macrophages in CRC progression, metastasis, and prognosis

The analysis of bulk CRC datasets showed that *SPP1* expression was elevated in advanced stage (III/IV) CRC (5/5), T3/4 (3/3), and N1-3 (2/3) samples compared with early stage (I/II), T1/2, and N0 samples (Fig. 5A–C), which suggests that *SPP1*<sup>+</sup> macrophages may play an important role in CRC progression. *SPP1* expression was also significantly higher in LM samples than in CRC samples in four bulk datasets (4/4) (Fig. 5D, E), suggesting the involvement of *SPP1*<sup>+</sup> macrophages in CRC metastasis.

Macrophage enrichment is generally associated with a poor prognosis in tumors, but in CRC, it is often linked with a favorable prognosis.<sup>12,78,79</sup> Univariate Cox analysis of four datasets (TCGA, GSE17536, GSE39582, and GSE41258) showed that *SPP1* expression was significantly associated with poor overall survival in three of the four datasets (Fig. 5F). Traditional macrophage markers, except for *iNOS* (showing a protective effect), were not significantly associated with CRC progression, metastasis, or prognosis. Considering the colinearity between *SPP1*<sup>+</sup> macrophages (*SPP1*) and macrophages (*CD68*), *SPP1/CD68* was used to study the relationship between *SPP1*<sup>+</sup> macrophages and clinical features in CRC (median grouping), revealing that H-*SPP1/CD68* patients had higher mean age, advanced T classification, lymph node metastasis, high levels of microsatellite instability, high tumor mutation burden, and



**Figure 5** Role of *SPP1*<sup>+</sup> macrophages in CRC progression, metastasis, and prognosis. (A–E) Heatmaps of the immune score (ESTIMATE) and expression of markers (*CD45*, *CD68*, *CD86*, *iNOS*, *CD163*, *CD206*, and *SPP1*) in bulk RNA-seq CRC datasets, comparing stage I/II vs. stage III/IV (A), T 1/2 vs. T 3/4 (B), N0 vs. N1-3 (C), M0 vs. M1 (D), and CRC vs. LM (E). Top, dataset source; bottom, sample size. \*GSE14333 uses the Duke stage. (F) Univariate Cox analysis of macrophage markers in four CRC datasets (TCGA, GSE17536, GSE39582, and GSE41258). (G) The top 10 enriched hallmark terms in the H-*SPP1/CD68* group according to GSEA. *SPP1*, secreted phosphoprotein 1; CI, confidence interval; GSEA, Gene Set Enrichment Analysis; HR, hazard ratio; *iNOS*, inducible nitric oxide synthase.

shorter survival time (overall survival and progression-free interval) (Table 2). Furthermore, GSEA revealed the enrichment of hallmark pathways such as EMT, angiogenesis, and hypoxia in the H-*SPP1*/*CD68* group (Fig. 5G).

In summary, *SPP1*<sup>+</sup> macrophages can be used as a biomarker of malignancy to evaluate CRC progression, metastasis, and prognosis.

### Spatial distribution characteristics of *SPP1*<sup>+</sup> macrophages and their potential impact on the CRC TME

We performed spatial transcriptomics analysis of NC, CRC, and LM samples to investigate the spatial distribution

**Table 2** Comparison of clinical parameters between the L-*SPP1*/*CD68* and H-*SPP1*/*CD68* CRC groups.

Clinical parameters	L- <i>SPP1</i> / <i>CD68</i> (n = 309)	H- <i>SPP1</i> / <i>CD68</i> (n = 308)	P value
Age (years)			
Mean ± standard deviation	64.7 ± 12.7	67.8 ± 12.5	0.002 <sup>a</sup>
NA	3	0	
Gender			
Female	140 (45.8%)	146 (47.4%)	0.682 <sup>b</sup>
Male	166 (54.2%)	162 (52.6%)	
NA	3	0	
Location			
Left	180 (61.2%)	160 (53.2%)	0.047 <sup>b</sup>
Right	114 (38.8%)	141 (46.8%)	
NA	15	7	
Pathologic stage			
I	62 (21.3%)	41 (13.5%)	0.078 <sup>b</sup>
II	109 (37.5%)	117 (38.6%)	
III	82 (28.2%)	96 (31.7%)	
IV	38 (13.1%)	49 (16.2%)	
NA	18	5	
T classification			
T1	14 (4.6%)	5 (1.6%)	0.005 <sup>b</sup>
T2	63 (20.7%)	41 (13.4%)	
T3	199 (65.5%)	219 (71.3%)	
T4	28 (9.2%)	42 (13.7%)	
NA	5	1	
N classification			
N0	182 (60.1%)	165 (53.7%)	0.025 <sup>b</sup>
N1	77 (25.4%)	71 (23.1%)	
N2	44 (14.5%)	71 (23.1%)	
NA	6	1	
Metastasis			
M0	216 (85.4%)	236 (82.8%)	0.417 <sup>b</sup>
M1	37 (14.6%)	49 (17.2%)	
NA	56	23	
High levels of microsatellite instability			
Yes	32 (10.5%)	50 (16.8%)	0.024 <sup>b</sup>
No	272 (89.5%)	247 (83.2%)	
NA	5	11	
Tumor mutation burden			
High	32 (11.7%)	50 (20.5%)	0.006 <sup>b</sup>
Low	242 (88.3%)	194 (79.5%)	
NA	35	64	
Overall survival (years)			
Mean (95% confidence interval)	7.97 (6.94–9.01)	6.49 (5.43–7.54)	0.046 <sup>c</sup>
Outcome event	55/299	68/294	
Progression-free interval (years)			
Mean (95% confidence interval)	8.09 (7.08–9.09)	6.03 (4.99–7.06)	0.030 <sup>c</sup>
Outcome event	67/298	86/293	

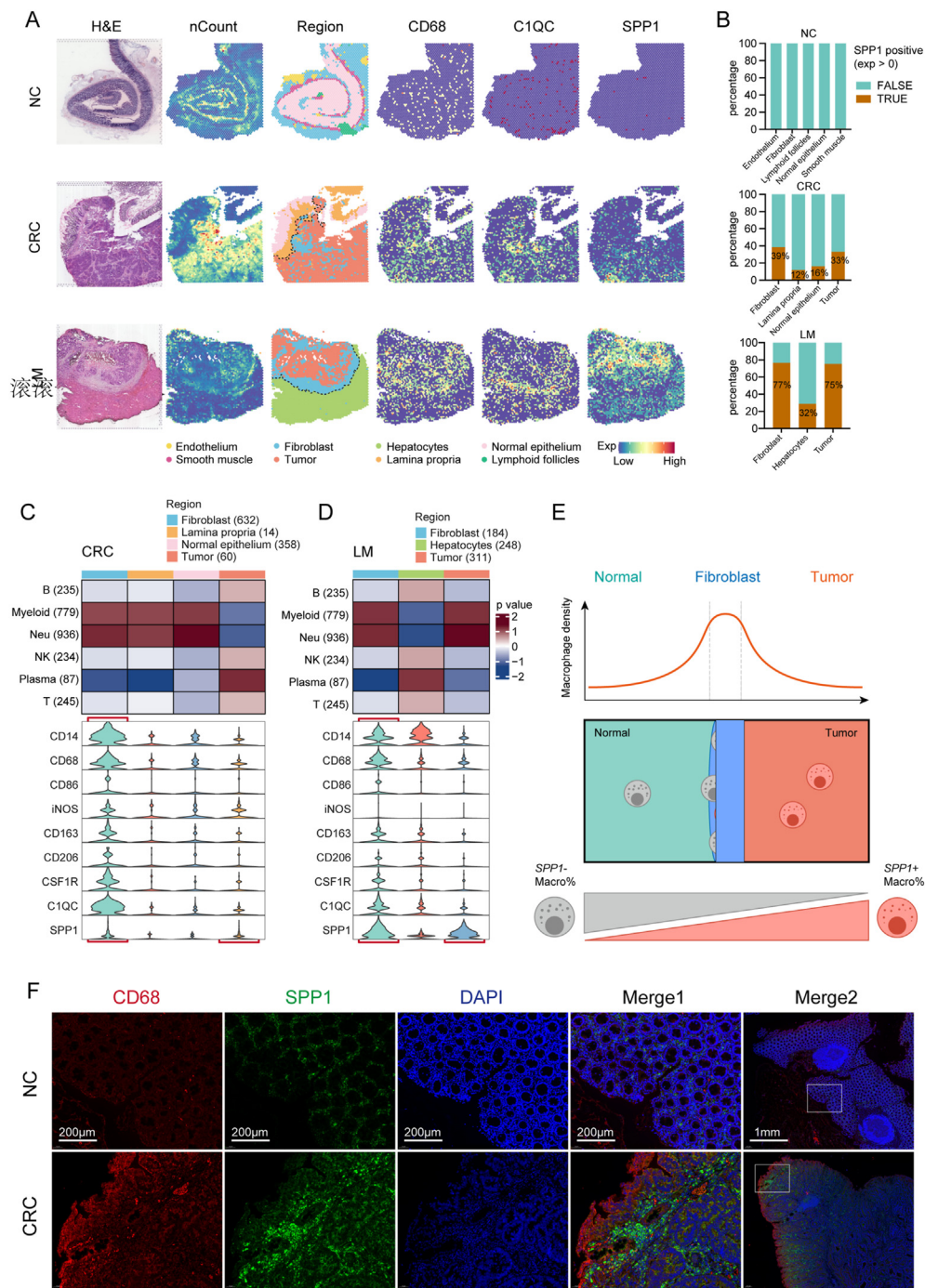
Note: The disease location “Left” includes the splenic flexure to the rectum. The disease location “Right” includes the cecum to the transverse colon.

<sup>a</sup> Mann-Whitney test

<sup>b</sup> chi-square test

<sup>c</sup> Log-rank test. High tumor mutation burden, >10 mutations per megabase; CRC, colorectal cancer; *SPP1*, secreted phosphoprotein 1.





**Figure 6** Spatial distribution characteristics of *SPP1*<sup>+</sup> macrophages and their potential impact on the CRC TME. **(A)** Spatial plots of hematoxylin-eosin staining (column 1), cell count (column 2), region (column 3), and the expression of three macrophage marker genes (column 4–6) in NC (top), CRC (middle), and LM (bottom) samples. **(B)** Stacked plots of the percentage of spots expressing *SPP1* in each region in NC, CRC, and LM samples. **(C)** MIA map of scRNA-seq-identified immune cell types and spatial transcriptomics (ST)-defined regions in CRC (upper). Red indicates enrichment and blue indicates depletion. The violin plot shows the expression of nine macrophage marker genes in ST-defined regions in CRC (lower). **(D)** MIA map of scRNA-seq-identified immune cell types and ST-defined regions in LM samples (upper). The violin plot shows the expression of nine macrophage marker genes in ST-defined regions in LM (lower). **(E)** Schematic diagram of the spatial distribution of macrophages. The fibroblast-enriched area at the tumor border had the highest macrophage density, while the tumor area had the highest proportion of *SPP1*<sup>+</sup> macrophages. **(F)** Immunofluorescence staining of human CRC tissue and paired NC tissue. *CD68* (red), *SPP1* (green), and DAPI (blue) in individual and merged channels are shown. CRC, colorectal cancer; NC, normal colorectum (adjacent colorectum); LM, liver metastases; MIA, multimodal intersection analysis; *SPP1*, secreted phosphoprotein 1; TME, tumor microenvironment.

characteristics of *SPP1*<sup>+</sup> macrophages (Fig. 6A; Fig. S5A–C). Both CRC and LM samples contained tumor regions, border regions (fibroblast-enriched), and normal tissue regions to ensure representative results. The spots in all regions of the NC samples lacked positive expression of *SPP1*. However, the tumor region and tumor border region (fibroblast region) in CRC and LM samples had significantly higher *SPP1*-positive rates than adjacent normal tissue regions (Fig. 6A, B). The multimodal intersection analysis results of CRC and LM samples showed significant enrichment of myeloid cells in the fibroblast region, and analysis of the expression of nine macrophage markers revealed high macrophage enrichment in the fibroblast region of both CRC and LM samples (Fig. 6C, D). *SPP1* was highly expressed in the tumor regions, unlike other macrophage markers that were only highly expressed in the fibroblast region. These results indicate that although the fibroblast region had the highest macrophage density, the tumor region had the highest proportion of *SPP1*<sup>+</sup> macrophages (Fig. 6E).

The immunofluorescence results of paired CRC and NC samples confirmed the enrichment of *SPP1*<sup>+</sup> macrophages in CRC (Fig. 6F). Additionally, these macrophages were found to be distributed closer to the core of the tumor, unlike other macrophages that predominantly concentrated at the boundary region of CRC (Fig. 6F).

We performed enrichment analysis (Metascape) of tumor region-specific genes in CRC and LM samples, which revealed elevated expression of hypoxia- and glycolysis-related pathway genes in tumor regions (Fig. S5D, E). This was further confirmed by the high expression of hypoxia and glycolysis signatures in the tumor region of LM samples (Fig. S5F, G).

In summary, the spatial transcriptomics and immunofluorescence results indicate that *SPP1*<sup>+</sup> macrophages are enriched in tumor regions, potentially contributing to the hypoxia and glycolysis characteristics of the CRC TME.

## Implications of *SPP1*<sup>+</sup> macrophages for CRC immunotherapy

Kaplan–Meier analysis of the TCGA-CRC dataset revealed a shorter overall survival in the high *SPP1/CD68* group compared with the low *SPP1/CD68* group ( $P < 0.001$ ) (Fig. 7A). ScRNA-seq data showed that preoperative chemotherapy reduced *SPP1* expression in macrophages (Fig. 7B, C), especially in responders. This suggests that preoperative chemotherapy may reduce *SPP1*<sup>+</sup> macrophages in CRC TME and may benefit patients. A high proportion of *SPP1*<sup>+</sup> macrophages was associated with genome instability and mutation (DNA mismatch repair deficiency, high levels of microsatellite instability, and high tumor mutation burden) in CRC (Fig. S6A–F), indicating that they could benefit from immunotherapy. SubMap algorithms predicted that the H-*SPP1/CD68* group was more likely to respond to anti-PD-1 therapy and anti-CTLA-4 therapy (Fig. 7D, E). *CSF1R* expression was significantly enriched in *C1QC*<sup>+</sup> macrophages but not *SPP1*<sup>+</sup> macrophages (Fig. 7F–H; Fig. S7A–D), suggesting that blocking *CSF1R* would preferentially deplete *C1QC*<sup>+</sup> macrophages over *SPP1*<sup>+</sup> macrophages.

## Discussion

This study provides a comprehensive analysis of *SPP1*<sup>+</sup> macrophages in CRC, including their origin, distribution, clinical value, functional pathways, and implications for treatment (Fig. 8A–E). Based on the results, the authors propose the *SPP1*<sup>+</sup> macrophage model theory, which explains the changes in macrophages during CRLM and guides clinical diagnosis and treatment.

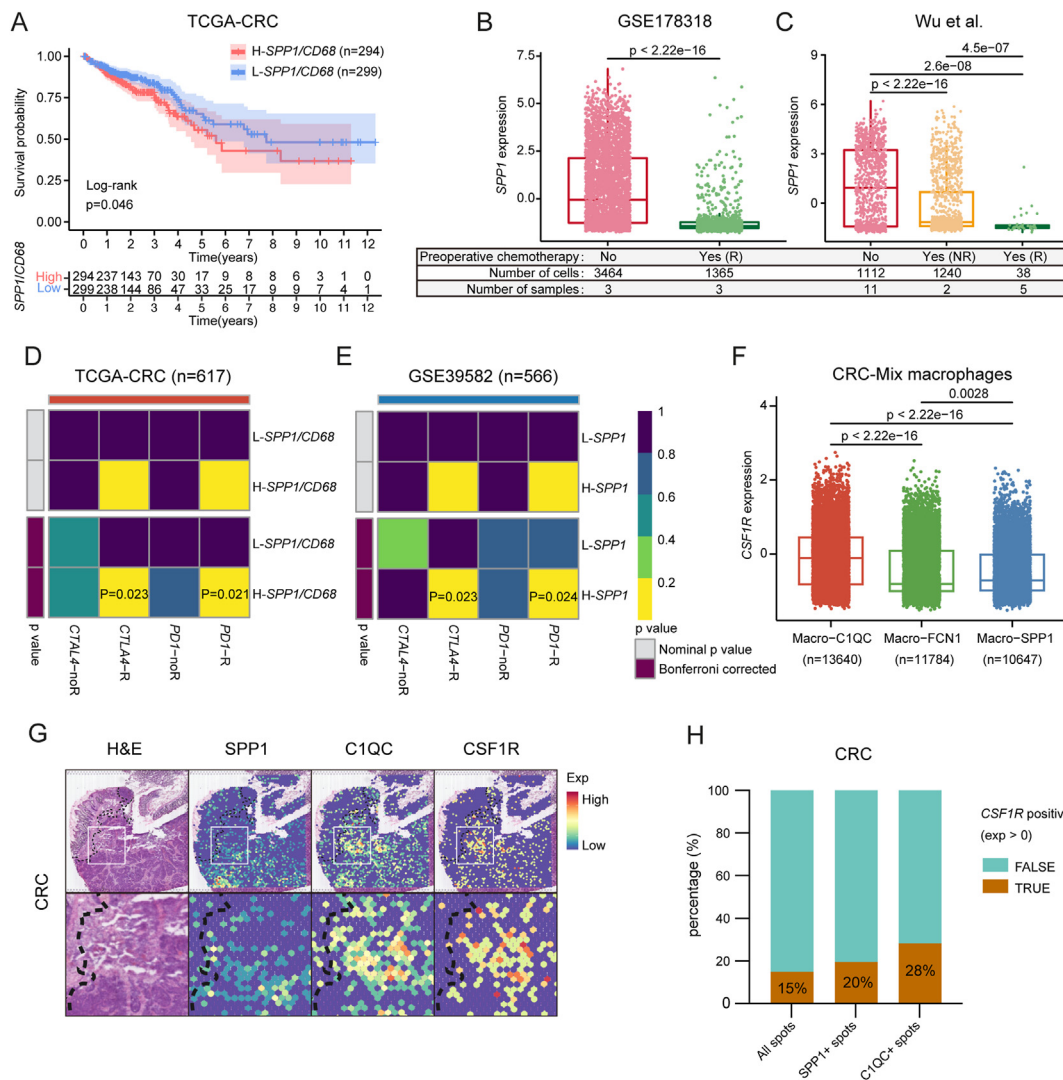
Four macrophage subsets were identified from CRC tissue, including *FCN1*<sup>+</sup> macrophages, *C1QC*<sup>+</sup> macrophages, *SPP1*<sup>+</sup> macrophages, and *MKI67*<sup>+</sup> macrophages (Fig. 8A). Enrichment analysis and phenotypic signatures indicated that inflammation, phagocytosis, malignancy, and proliferation were the most prominent features related to these four macrophage subsets, respectively.

The *SPP1*<sup>+</sup> macrophages were specifically enriched in primary foci and liver metastases of CRC, and evidence from trajectory and tissue distribution analysis supports the hypothesis that they originate from monocytes recruited into colonic tissues and rapidly differentiate into macrophages, which then differentiate into monocyte-like *FCN1*<sup>+</sup> macrophages that polarize toward *C1QC*<sup>+</sup> macrophages and *SPP1*<sup>+</sup> macrophages (Fig. 8B).

Myeloid cells play an active role in shaping the CRC TME,<sup>35</sup> with a significant increase in the proportion of myeloid cells and macrophages confirmed in our study. Interestingly, the number of macrophages in tissues did not change significantly during CRC tumorigenesis, but the decrease in overall immune cells resulted in a significant increase in the proportion of macrophages (Fig. 8C). This suggests that anti-tumor immunity weakens as CRC progresses, while the importance of macrophages in the TME increases.

*SPP1*<sup>+</sup> tumor-associated macrophages have been associated with CRC malignancy and liver metastasis,<sup>19,20,32,33</sup> and *SPP1* is a diagnostic and prognostic marker for CRC.<sup>22,28,80,81</sup> Our study confirms *SPP1*<sup>+</sup> macrophages as a comprehensive marker for CRC diagnosis, disease severity assessment, and prognosis evaluation, with their number and proportion constantly increasing during CRC occurrence, progression, and metastasis, indicating a poor prognosis (Fig. 8C). These results suggest that *SPP1*<sup>+</sup> macrophages play an important driving role in the occurrence, progression, and metastasis of CRC.

The *in vitro* M1/M2 polarization model has limitations in describing macrophages in the complex phenotype of the TME.<sup>8</sup> Coexpression of M1 and M2 signatures in macrophage subsets in cancer, including CRC, has been reported by single-cell studies.<sup>19,32,57,76</sup> In our study, we also found an increase in the proportion of M1 macrophages and a decrease in M2 macrophages in CRC, contradicting the M1/M2 polarization theory. The biological function of *SPP1*<sup>+</sup> macrophages in CRC remains unclear, but our study suggests they may promote CRC through EMT, hypoxia, glycolysis, and immunosuppressive pathways (Fig. 8D). *SPP1* is a key functional molecule of *SPP1*<sup>+</sup> macrophages, and these macrophages interact with other cells through the *SPP1-CD44*, *SPP1-PTGER4*, and *SPP1-a4b1* complex axes. This suggests that *SPP1*<sup>+</sup> macrophages may promote CRC progression and metastasis through these interactions (Fig. 8D).



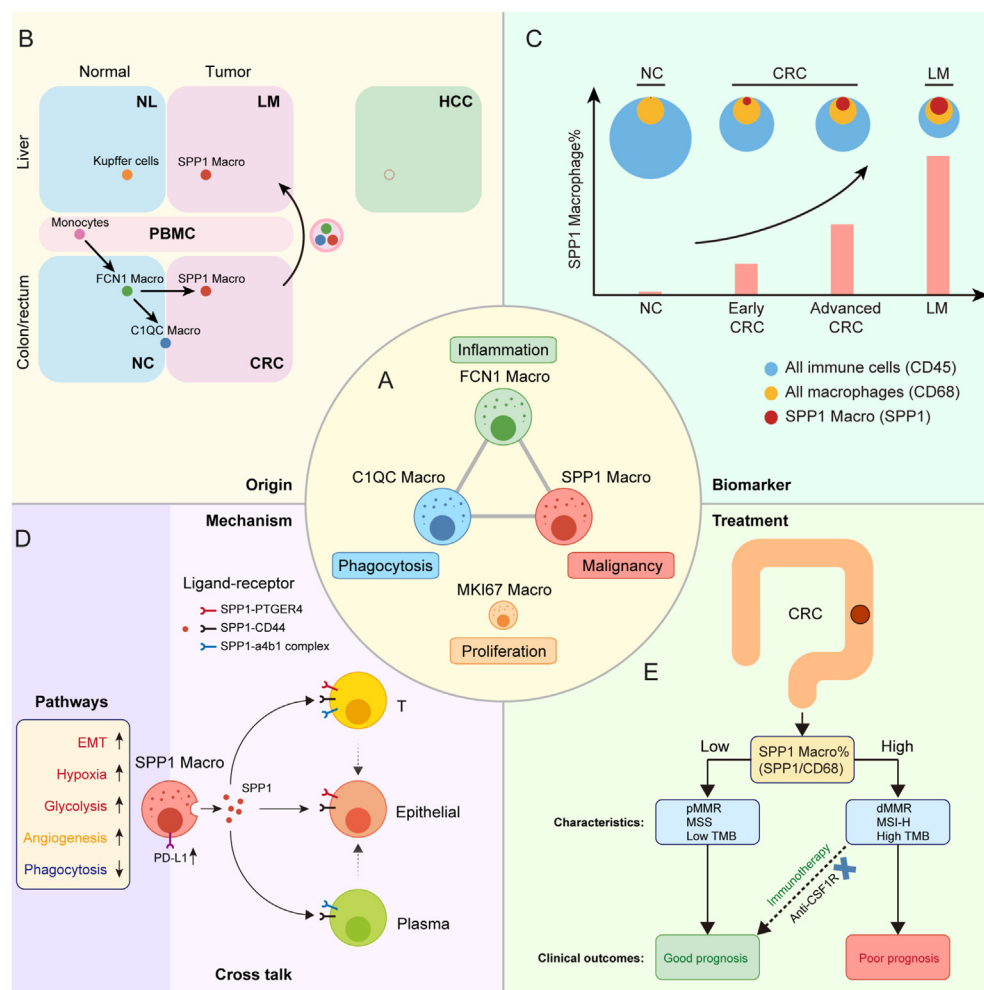
**Figure 7** Implications of *SPP1*<sup>+</sup> macrophages for CRC immunotherapy. **(A)** The Kaplan–Meier survival curve showed the high expression of *SPP1/CD68* was associated with worse overall survival (OS) in TCGA-CRC cohorts ( $n = 593$ ). Log-rank test. **(B, C)** Box plots of the effect of preoperative chemotherapy on *SPP1* expression in macrophages in the scRNA-seq cohorts of **(B)** GSE178318 and **(C)** Wu et al. R represents responders and NR represents non-responders. Mann–Whitney U test. **(D)** SubMap analysis in TCGA-CRC showed that the H-*SPP1/CD68* group was more sensitive to CTLA4 inhibitors (Bonferroni-corrected  $P = 0.023$ ) and PD-1 inhibitors (Bonferroni-corrected  $P = 0.021$ ) than the L-*SPP1/CD68* group. **(E)** SubMap analysis in GSE39582 showed that the H-*SPP1* group was more sensitive to the CTLA4 inhibitor (Bonferroni-corrected  $P = 0.023$ ) and PD-1 inhibitor (Bonferroni-corrected  $P = 0.024$ ). **(F)** The expression of *CSF1R* in the three major macrophage subsets (CRC-Mix). **(G)** Spatial feature plots of the expression of *SPP1* (column 2), *C1QC* (column 3), and *CSF1R* (column 4) in the CRC sample. **(H)** Stacked plots of the percentage of spots expressing *CSF1R* in regions of all spots ( $n = 3138$ ), *SPP1*<sup>+</sup> spots, and *C1QC*<sup>+</sup> spots in the CRC sample. SPP1, secreted phosphoprotein 1; CRC, colorectal cancer; CTLA4, cytotoxic T-lymphocyte associated protein 4; PD-1, programmed death-1; CSF1R, colony-stimulating factor 1 receptor; C1QC, complement C1q C chain.

The understanding of *SPP1*<sup>+</sup> macrophages presents new potential for CRC treatment. Chemotherapy-induced macrophage phenotype switching has been reported in breast cancer,<sup>82</sup> and our findings suggest that preoperative chemotherapy can significantly reduce *SPP1* expression in CRC macrophages, especially in responders. This suggests that effective chemotherapy may inhibit *SPP1*<sup>+</sup> macrophages specifically and raise the possibility of immunotherapy targeting *SPP1*<sup>+</sup> macrophages. Importantly, patients with high *SPP1*<sup>+</sup> macrophage levels show higher levels of

genomic instability and mutation characteristics that predict better responses to immune checkpoint blockade (Fig. 8E). Our findings are supported by the enrichment of *PD-L1* and *HLA-G* in *SPP1*<sup>+</sup> macrophages and prediction results of the SubMap analysis. Several studies have shown that targeted blockade of *SPP1* can inhibit tumor progression and metastasis.<sup>21,25,83–85</sup> Thus, targeting *SPP1*<sup>+</sup> macrophages may offer a viable option for CRC immunotherapy.

Notably, our *SPP1*<sup>+</sup> macrophage theory also reveals a possible problem in current macrophage-targeted





**Figure 8** Schematic diagram of the *SPP1*<sup>+</sup> macrophage model in CRC. **(A)** Classification of macrophage subsets in CRC, including *FCN1*<sup>+</sup>, *C1QC*<sup>+</sup>, *SPP1*<sup>+</sup>, and *MKI67*<sup>+</sup> macrophages, defined by their core features of inflammation, phagocytosis, malignancy, and proliferation, respectively. **(B)** Model of the developmental trajectory of monocyte/macrophage lineages in different sample types from CRLM and HCC patients. This model shows the cell origin and tissue distribution of *SPP1*<sup>+</sup> macrophages. **(C)** The number and proportion of *SPP1*<sup>+</sup> macrophages increased during CRC occurrence, progression, and metastasis, making them a marker of CRC malignancy. Blue, yellow, and red dots represent the numbers of all immune cells, all macrophages, and *SPP1*<sup>+</sup> macrophages in the sample, respectively. **(D)** *SPP1*<sup>+</sup> macrophages had distinct signatures compared with other macrophages (left). Red: signature specifically elevated in *SPP1*<sup>+</sup> macrophages. Yellow: signature elevated in *SPP1*<sup>+</sup> macrophages compared with *C1QC*<sup>+</sup> macrophages. Blue: signature decreased in *SPP1*<sup>+</sup> macrophages compared with *C1QC*<sup>+</sup> macrophages. The right shows the crosstalk between *SPP1*<sup>+</sup> macrophages and other cell subsets through the *SPP1* protein. **(E)** Patients with a high proportion of *SPP1*<sup>+</sup> macrophages have a worse prognosis, along with signatures related to genome instability and mutations. Immunotherapy has the potential to improve outcomes in patients with a high proportion of *SPP1*<sup>+</sup> macrophages, whereas targeting *CSF1R* is less effective. *SPP1*, secreted phosphoprotein 1; *CSF1R*, colony-stimulating factor 1 receptor; CRC, colorectal cancer; CRLM, colorectal cancer liver metastases; HCC, hepatocellular carcinoma; *C1QC*, complement C1q C chain.

immunotherapy research. Inhibition of the *CSF1-CSF1R* axis to reduce tumor-associated macrophages is a major research direction of macrophage-targeted immunotherapy.<sup>6,13–15</sup> However, we found that *CSF1R* was significantly enriched in *C1QC*<sup>+</sup> macrophages compared with other macrophages, suggesting that *CSF1R* blockade may preferentially deplete the protective *C1QC*<sup>+</sup> macrophage subset while sparing malignant *SPP1*<sup>+</sup> macrophages. Anti-*CSF1R* therapy may not be sufficient to

deplete all tumor-promoting *SPP1*<sup>+</sup> macrophages, which may contribute to the poor efficacy of anti-*CSF1R* monotherapy in the Renca mouse tumor model and human cancer patients.<sup>15,16,19</sup>

In summary, we propose an *SPP1*<sup>+</sup> macrophage model to explain macrophage characteristics and changes in CRC. *SPP1*<sup>+</sup> macrophages are a marker of malignancy and targeting them would serve as a promising CRC therapeutic strategy.



## Ethics declaration

The research was approved by the Medical Ethics Committee of the First Affiliated Hospital of the Fourth Military Medical University (Approval No.: KY20212211–C-1). The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

## CRedit authorship contribution statement

Z.Y.X. designed the study. Z.Y.X. and L.R.N. collected the data. Z.Y.X., F.F., and G.Z.Z. analyzed the data. Z.Y.X., K.L.D., R.K.L., H.J.D., L.L.D., H.Z.W., G.M.R., and X.Y.D. visualized the data. Z.Y.X. and L.R.N. drafted the manuscript. Z.Y.X., G.Z.Z., J.Z., S.C.D., and J.Y.Z. revised the manuscript. All authors read and approved the final manuscript. All authors contributed to the manuscript and approved the submitted version.

## Conflict of interests

The authors declared no conflict of interests.

## Funding

This work was supported by the National Natural Science Foundation of China (No. 82072655), the Scientific and technological innovation team of Shaanxi Innovation Capability Support Plan (China) (No. 2023-CX-TD-67), and the Key R&D Plan of Shaanxi Province, China (No. 2022SF-603).

## Data availability

We utilized publicly available datasets in this study, and their details can be found in the “Methods” section of the manuscript. The codes used in this study are accessible on GitHub (<https://github.com/LPC19970117/SPP1-macrophages>). Further inquiries can be directed to the corresponding authors.

## Ethical approval

Human tissue specimens were obtained from Xijing Hospital under the approval of the Institutional Review Board and have obtained informed consent from the patients.

Review approval document of the medical ethics committee of the first affiliated hospital of the air force medical university (Xijing Hospital).

Approval number: KY20212211–C-1.

## Acknowledgements

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 manuscript has been released as a pre-print at *Research Square*.<sup>86</sup>

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.gendis.2024.101340>.

## References

1. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2022. *CA A Cancer J Clin*. 2022;72(1):7–33.
2. Fan A, Wang B, Wang X, et al. Immunotherapy in colorectal cancer: current achievements and future perspective. *Int J Biol Sci*. 2021;17(14):3837–3849.
3. Zaanen A, Shi Q, Taieb J, et al. Role of deficient DNA mismatch repair status in patients with stage III colon cancer treated with FOLFOX adjuvant chemotherapy: a pooled analysis from 2 randomized clinical trials. *JAMA Oncol*. 2018;4(3):379–383.
4. Nordholm-Carstensen A, Krarup PM, Morton D, Harling H, Group DCC. Mismatch repair status and synchronous metastases in colorectal cancer: a nationwide cohort study. *Int J Cancer*. 2015;137(9):2139–2148.
5. Le DT, Uram JN, Wang H, et al. PD-1 blockade in tumors with mismatch-repair deficiency. *N Engl J Med*. 2015;372(26):2509–2520.
6. DeNardo DG, Ruffell B. Macrophages as regulators of tumour immunity and immunotherapy. *Nat Rev Immunol*. 2019;19(6):369–382.
7. Vogel DYS, Glim JE, Stavenhagen AWD, et al. Human macrophage polarization *in vitro*: maturation and activation methods compared. *Immunobiology*. 2014;219(9):695–703.
8. Ginhoux F, Schultze JL, Murray PJ, Ochando J, Biswas SK. New insights into the multidimensional concept of macrophage ontogeny, activation and function. *Nat Immunol*. 2016;17(1):34–40.
9. Khantakova D, Brioschi S, Molgora M. Exploring the impact of TREM2 in tumor-associated macrophages. *Vaccines*. 2022;10(6):943.
10. Ding J, Yang C, Zhang Y, et al. M2 macrophage-derived G-CSF promotes trophoblasts EMT, invasion and migration via activating PI3K/Akt/Erk1/2 pathway to mediate normal pregnancy. *J Cell Mol Med*. 2021;25(4):2136–2147.
11. Wang Z, Du Z, Sheng H, et al. Polarization of intestinal tumour-associated macrophages regulates the development of schistosoma colorectal cancer. *J Cancer*. 2021;12(4):1033–1041.
12. Cortese N, Soldani C, Franceschini B, et al. Macrophages in colorectal cancer liver metastases. *Cancers*. 2019;11(5):633.
13. Pyonteck SM, Akkari L, Schuhmacher AJ, et al. CSF-1R inhibition alters macrophage polarization and blocks glioma progression. *Nat Med*. 2013;19(10):1264–1272.
14. Lee JH, Chen TWW, Hsu CH, et al. A phase I study of pexidartinib, a colony-stimulating factor 1 receptor inhibitor, in Asian patients with advanced solid tumors. *Invest N Drugs*. 2020;38(1):99–110.
15. Ries CH, Cannarile MA, Hoves S, et al. Targeting tumor-associated macrophages with anti-CSF-1R antibody reveals a strategy for cancer therapy. *Cancer Cell*. 2014;25(6):846–859.
16. Papadopoulos KP, Gluck L, Martin LP, et al. First-in-human study of AMG 820, a monoclonal anti-colony-stimulating factor 1 receptor antibody, in patients with advanced solid tumors. *Clin Cancer Res*. 2017;23(19):5703–5710.
17. Liu Y, Zhang Q, Xing B, et al. Immune phenotypic linkage between colorectal cancer and liver metastasis. *Cancer Cell*. 2022;40(4):424–437.e5.

18. Wu Y, Yang S, Ma J, et al. Spatiotemporal immune landscape of colorectal cancer liver metastasis at single-cell level. *Cancer Discov.* 2022;12(1):134–153.
19. Zhang L, Li Z, Skrzypczynska KM, et al. Single-cell analyses inform mechanisms of myeloid-targeted therapies in colon cancer. *Cell.* 2020;181(2):442–459.e29.
20. Qi J, Sun H, Zhang Y, et al. Single-cell and spatial analysis reveal interaction of FAP<sup>+</sup> fibroblasts and SPP1<sup>+</sup> macrophages in colorectal cancer. *Nat Commun.* 2022;13(1):1742.
21. Feng S, Yuan W, Sun Z, et al. SPP1 as a key gene in the lymph node metastasis and a potential predictor of poor prognosis in head and neck carcinoma. *J Oral Pathol Med.* 2022;51(7):620–629.
22. Liu Y, Ye G, Dong B, et al. A pan-cancer analysis of the oncogenic role of secreted phosphoprotein 1 (SPP1) in human cancers. *Ann Transl Med.* 2022;10(6):279.
23. Zhao K, Ma Z, Zhang W. Comprehensive analysis to identify SPP1 as a prognostic biomarker in cervical cancer. *Front Genet.* 2022;12:732822.
24. Xu C, Sun L, Jiang C, et al. SPP1, analyzed by bioinformatics methods, promotes the metastasis in colorectal cancer by activating EMT pathway. *Biomed Pharmacother.* 2017;91:1167–1177.
25. Wei J, Chen Z, Hu M, et al. Characterizing intercellular communication of pan-cancer reveals SPP1<sup>+</sup> tumor-associated macrophage expanded in hypoxia and promoting cancer malignancy through single-cell RNA-seq data. *Front Cell Dev Biol.* 2021;9:749210.
26. Rao G, Wang H, Li B, et al. Reciprocal interactions between tumor-associated macrophages and CD44-positive cancer cells via osteopontin/CD44 promote tumorigenicity in colorectal cancer. *Clin Cancer Res.* 2013;19(4):785–797.
27. Senbanjo LT, Chellaiah MA. CD44: a multifunctional cell surface adhesion receptor is a regulator of progression and metastasis of cancer cells. *Front Cell Dev Biol.* 2017;5:18.
28. Assidi M, Gomaa W, Jafri M, et al. Prognostic value of osteopontin (SPP1) in colorectal carcinoma requires a personalized molecular approach. *Tumour Biol.* 2019;41(9):1010428319863627.
29. Pietras A, Katz AM, Ekström EJ, et al. Osteopontin-CD44 signaling in the glioma perivascular niche enhances cancer stem cell phenotypes and promotes aggressive tumor growth. *Cell Stem Cell.* 2014;14(3):357–369.
30. Pang X, Gong K, Zhang X, Wu S, Cui Y, Qian BZ. Osteopontin as a multifaceted driver of bone metastasis and drug resistance. *Pharmacol Res.* 2019;144:235–244.
31. Clay R, Siddiqi SA. *Recent Advances in Molecular Diagnostics and Therapeutic Targets for Pancreatic Cancer. Theranostic Approach for Pancreatic Cancer.* Amsterdam: Elsevier; 2019:325–367.
32. Cheng S, Li Z, Gao R, et al. A pan-cancer single-cell transcriptional atlas of tumor infiltrating myeloid cells. *Cell.* 2021;184(3):792–809.e23.
33. Ren X, Zhang L, Zhang Y, Li Z, Siemers N, Zhang Z. Insights gained from single-cell analysis of immune cells in the tumor microenvironment. *Annu Rev Immunol.* 2021;39:583–609.
34. Fawcner-Corbett D, Antanaviciute A, Parikh K, et al. Spatio-temporal analysis of human intestinal development at single-cell resolution. *Cell.* 2021;184(3):810–826.e23.
35. Lee HO, Hong Y, Etlioglu HE, et al. Lineage-dependent gene expression programs influence the immune landscape of colorectal cancer. *Nat Genet.* 2020;52(6):594–603.
36. Che LH, Liu JW, Huo JP, et al. A single-cell atlas of liver metastases of colorectal cancer reveals reprogramming of the tumor microenvironment in response to preoperative chemotherapy [published correction appears in *Cell Discov.* 2023 Mar 22;9(1):31]. *Cell Discov.* 2021;7(1):80.
37. Pelka K, Hofree M, Chen JH, et al. Spatially organized multicellular immune hubs in human colorectal cancer. *Cell.* 2021;184(18):4734–4752.e20.
38. Sharma A, Seow JJW, Dutertre CA, et al. Onco-fetal reprogramming of endothelial cells drives immunosuppressive macrophages in hepatocellular carcinoma. *Cell.* 2020;183(2):377–394.e21.
39. Ki DH, Jeung HC, Park CH, et al. Whole genome analysis for liver metastasis gene signatures in colorectal cancer. *Int J Cancer.* 2007;121(9):2005–2012.
40. Jorissen RN, Gibbs P, Christie M, et al. Metastasis-associated gene expression changes predict poor outcomes in patients with dukes stage B and C colorectal cancer. *Clin Cancer Res.* 2009;15(24):7642–7651.
41. Smith JJ, Deane NG, Wu F, et al. Experimentally derived metastasis gene expression profile predicts recurrence and death in patients with colon cancer. *Gastroenterology.* 2010;138(3):958–968.
42. Gaedcke J, Grade M, Jung K, et al. Mutated KRAS results in overexpression of DUSP4, a MAP-kinase phosphatase, and SMYD3, a histone methyltransferase, in rectal carcinomas. *Genes Chromosomes Cancer.* 2010;49(11):1024–1034.
43. Skrzypczak M, Goryca K, Rubel T, et al. Modeling oncogenic signaling in colon tumors by multidirectional analyses of microarray data directed for maximization of analytical reliability. *PLoS One.* 2010;5(10):e13091.
44. Marisa L, de Reyniès A, Duval A, et al. Gene expression classification of colon cancer into molecular subtypes: characterization, validation, and prognostic value. *PLoS Med.* 2013;10(5):e1001453.
45. Sheffer M, Bacolod MD, Zuk O, et al. Association of survival and disease progression with chromosomal instability: a genomic exploration of colorectal cancer. *Proc Natl Acad Sci USA.* 2009;106(17):7131–7136.
46. Solé X, Crous-Bou M, Cordero D, et al. Discovery and validation of new potential biomarkers for early detection of colon cancer. *PLoS One.* 2014;9(9):e106748.
47. Ryan BM, Zanetti KA, Robles AI, et al. Germline variation in NCF4, an innate immunity gene, is associated with an increased risk of colorectal cancer. *Int J Cancer.* 2014;134(6):1399–1407.
48. Kwon Y, Park M, Jang M, et al. Prognosis of stage III colorectal carcinomas with FOLFOX adjuvant chemotherapy can be predicted by molecular subtype. *Oncotarget.* 2017;8(24):39367–39381.
49. Hu Y, Gaedcke J, Emons G, et al. Colorectal cancer susceptibility loci as predictive markers of rectal cancer prognosis after surgery. *Genes Chromosomes Cancer.* 2018;57(3):140–149.
50. Szeglin BC, Wu C, Marco MR, et al. A SMAD4-modulated gene profile predicts disease-free survival in stage II and III colorectal cancer. *Cancer Rep.* 2022;5(1):e1423.
51. McInnes L, Healy J. UMAP: uniform manifold approximation and projection for dimension reduction. *arXiv:1802.03426*; 2018. <http://arxiv.org/abs/1802.03426>.
52. Mulder K, Patel AA, Kong WT, et al. Cross-tissue single-cell landscape of human monocytes and macrophages in health and disease. *Immunity.* 2021;54(8):1883–1900.e5.
53. Guillems M, Scott CL. Does niche competition determine the origin of tissue-resident macrophages? *Nat Rev Immunol.* 2017;17(7):451–460.
54. Korsunsky I, Millard N, Fan J, et al. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat Methods.* 2019;16(12):1289–1296.
55. Trapnell C, Cacchiarelli D, Grimsby J, et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol.* 2014;32(4):381–386.

56. Efremova M, Vento-Tormo M, Teichmann SA, Vento-Tormo R. CellPhoneDB: inferring cell–cell communication from combined expression of multi-subunit ligand–receptor complexes. *Nat Protoc.* 2020;15(4):1484–1506.
57. Azizi E, Carr AJ, Plitas G, et al. Single-cell map of diverse immune phenotypes in the breast tumor microenvironment. *Cell.* 2018;174(5):1293–1308.e36.
58. Wei J, Huang K, Chen Z, et al. Characterization of glycolysis-associated molecules in the tumor microenvironment revealed by pan-cancer tissues and lung cancer single cell data. *Cancers.* 2020;12(7):1788.
59. Mak MP, Tong P, Diao L, et al. A patient-derived, pan-cancer EMT signature identifies global molecular alterations and immune target enrichment following epithelial-to-mesenchymal transition. *Clin Cancer Res.* 2016;22(3):609–620.
60. Whitfield ML, George LK, Grant GD, Perou CM. Common markers of proliferation. *Nat Rev Cancer.* 2006;6(2):99–106.
61. Yoshihara K, Shahmoradgoli M, Martínez E, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun.* 2013;4:2612.
62. Newman AM, Liu CL, Green MR, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods.* 2015;12(5):453–457.
63. Wu K, Lin K, Li X, et al. Redefining tumor-associated macrophage subpopulations and functions in the tumor microenvironment. *Front Immunol.* 2020;11:1731.
64. Zhao Y, Ge X, Xu X, Yu S, Wang J, Sun L. Prognostic value and clinicopathological roles of phenotypes of tumour-associated macrophages in colorectal cancer. *J Cancer Res Clin Oncol.* 2019;145(12):3005–3019.
65. Li T, Fan J, Wang B, et al. TIMER: a web server for comprehensive analysis of tumor-infiltrating immune cells. *Cancer Res.* 2017;77(21):e108–e110.
66. Moncada R, Barkley D, Wagner F, et al. Author Correction: integrating microarray-based spatial transcriptomics and single-cell RNA-seq reveals tissue architecture in pancreatic ductal adenocarcinomas. *Nat Biotechnol.* 2020;38(12):1476.
67. Mayakonda A, Lin DC, Assenov Y, Plass C, Koeffler HP. Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res.* 2018;28(11):1747–1756.
68. Hoshida Y, Brunet JP, Tamayo P, Golub TR, Mesirov JP. Subclass mapping: identifying common subtypes in independent disease data sets. *PLoS One.* 2007;2(11):e1195.
69. Lu X, Jiang L, Zhang L, et al. Immune signature-based subtypes of cervical squamous cell carcinoma tightly associated with human papillomavirus type 16 expression, molecular features, and clinical outcome. *Neoplasia.* 2019;21(6):591–601.
70. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* 2015;1(6):417–425.
71. Zhou Y, Zhou B, Pache L, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun.* 2019;10(1):1523.
72. Hänzelmann S, Castelo R, Guinney J. GSEA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinf.* 2013;14:7.
73. Vasaikar SV, Straub P, Wang J, Zhang B. LinkedOmics: analyzing multi-omics data within and across 32 cancer types. *Nucleic Acids Res.* 2018;46(D1):D956–D963.
74. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA.* 2005;102(43):15545–15550.
75. Loustau M, Anna F, Dréan R, Lecomte M, Langlade-Demoyen P, Caumartin J. HLA-G neo-expression on tumors. *Front Immunol.* 2020;11:1685.
76. Zhang Y, Zhang Z. The history and advances in cancer immunotherapy: understanding the characteristics of tumor-infiltrating immune cells and their therapeutic implications. *Cell Mol Immunol.* 2020;17(8):807–821.
77. Pontén F, Jirstrom K, Uhlen M. The human protein atlas: a tool for pathology. *J Pathol.* 2008;216(4):387–393.
78. Li J, Li L, Li Y, et al. Tumor-associated macrophage infiltration and prognosis in colorectal cancer: systematic review and meta-analysis. *Int J Colorectal Dis.* 2020;35(7):1203–1210.
79. Bruni D, Angell HK, Galon J. The immune contexture and Immunoscore in cancer prognosis and therapeutic efficacy. *Nat Rev Cancer.* 2020;20(11):662–680.
80. Amilca-Seba K, Tan TZ, Thiery JP, et al. Osteopontin (OPN/SPP<sub>1</sub>), a mediator of tumor progression, is regulated by the mesenchymal transcription factor slug/SNAI2 in colorectal cancer (CRC). *Cells.* 2022;11(11):1808.
81. Tu Y, Chen C, Fan G. Association between the expression of secreted phosphoprotein-related genes and prognosis of human cancer. *BMC Cancer.* 2019;19(1):1230.
82. Liu J, Lao L, Chen J, et al. The IRENA lncRNA converts chemotherapy-polarized tumor-suppressing macrophages to tumor-promoting phenotypes in breast cancer. *Nat Can (Ott).* 2021;2(4):457–473.
83. Klement JD, Poschel DB, Lu C, et al. Osteopontin blockade immunotherapy increases cytotoxic T lymphocyte lytic activity and suppresses colon tumor progression. *Cancers.* 2021;13(5):1006.
84. Deng G, Zeng F, Su J, et al. BET inhibitor suppresses melanoma progression via the noncanonical NF- $\kappa$ B/SPP<sub>1</sub> pathway. *Theranostics.* 2020;10(25):11428–11443.
85. Farrokhi V, Chabot JR, Neubert H, Yang Z. Assessing the feasibility of neutralizing osteopontin with various therapeutic antibody modalities. *Sci Rep.* 2018;8(1):7781.
86. Zheng J, Xie Z, Niu L, et al. SPP1<sup>+</sup> macrophages: a malignant macrophage subset in the colorectal cancer microenvironment revealed by single-cell sequencing. *Research Square.* 2022.