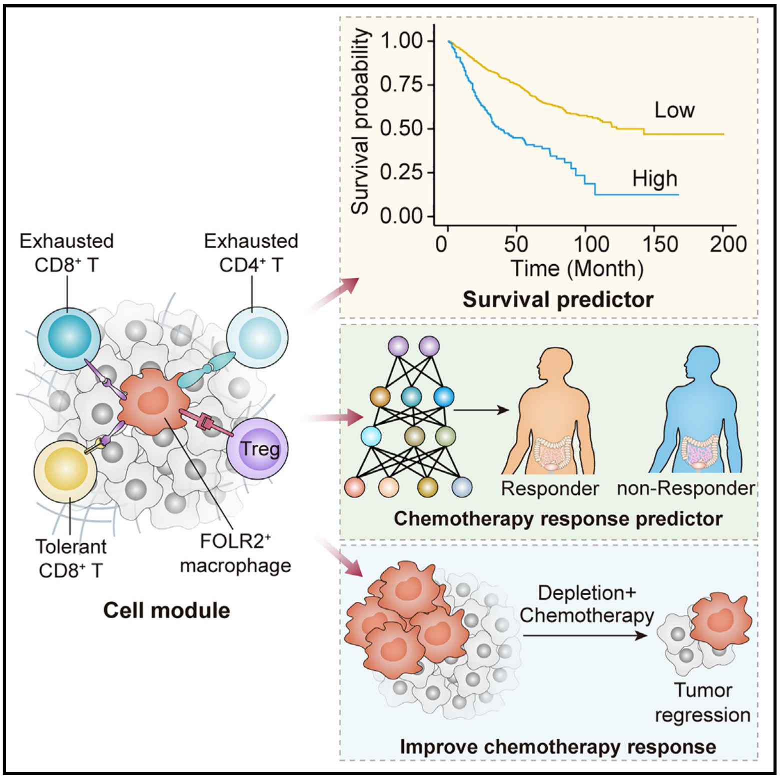
Cell Reports
Medicine

Article

A multiomics analysis-assisted deep learning model identifies a macrophage-oriented module as a potential therapeutic target in colorectal cancer

Graphical abstract



Highlights

- A FOLR2$^+$ macrophage-oriented cell module is uncovered in colorectal cancer
- Deep learning models utilize the macrophage-oriented spatial interaction map
- Targeting FOLR2$^+$ resident macrophages enhances the response to chemotherapy

Authors

Xuanwen Bao, Qiong Li, Dong Chen, ..., Jianpeng Sheng, Jian Ruan, Peng Zhao

Correspondence

xuanwen.bao@zju.edu.cn (X.B.),
shengjp@zju.edu.cn (J.S.),
software233@zju.edu.cn (J.R.),
zhaop@zju.edu.cn (P.Z.)

In brief

Bao et al. utilize a deep learning model supported by multiomics analysis to establish a FOLR2$^+$ macrophage-oriented cell module. This module serves as both a therapeutic target and a prognostic predictor in colorectal cancer. Additionally, the spatial interaction map not only predicts chemotherapy response but also identifies potential therapeutic targets.

Bao et al., 2024, Cell Reports Medicine *5*, 101399
February 20, 2024 © 2024 The Author(s).
https://doi.org/10.1016/j.xcrm.2024.101399

CellPress

Cell Reports Medicine

CellPress
OPEN ACCESS

Article

A multiomics analysis-assisted deep learning model identifies a macrophage-oriented module as a potential therapeutic target in colorectal cancer

Xuanwen Bao,[1,14,*] Qiong Li,[1,14] Dong Chen,[2,14] Xiaomeng Dai,[1,14] Chuan Liu,[1,14] Weihong Tian,[4] Hangyu Zhang,[1] Yuzhi Jin,[1] Yin Wang,[3] Jinlin Cheng,[5] Chunyu Lai,[1] Chanqi Ye,[1] Shan Xin,[6] Xin Li,[7] Ge Su,[3] Yongfeng Ding,[1] Yangyang Xiong,[8] Jindong Xie,[9] Vincent Tano,[10] Yanfang Wang,[11] Wenguang Fu,[12] Shuiguang Deng,[3] Weijia Fang,[1] Jianpeng Sheng,[13,*] Jian Ruan,[1,12,*] and Peng Zhao[1,15,*]

[1]Department of Medical Oncology, The First Affiliated Hospital, School of Medicine, Zhejiang University, Hangzhou, Zhejiang Province 310003, China

[2]Department of Colorectal Surgery, The First Affiliated Hospital, School of Medicine, Zhejiang University, Hangzhou, Zhejiang Province 310003, China

[3]College of Computer Science and Technology, Zhejiang University, Hangzhou, Zhejiang Province 310003, China

[4]Department of Immunology, School of Medicine, Jiangsu University, Zhenjiang, Jiangsu 212013, China

[5]State Key Laboratory for Diagnosis and Treatment of Infectious Diseases, The First Affiliated Hospital, School of Medicine, Zhejiang University, Hangzhou, Zhejiang Province 310003, China

[6]Department of Genetics, Yale School of Medicine, New Haven, CT 06510, USA

[7]Department of Chronic Inflammation and Cancer, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany

[8]Department of Gastroenterology, The First Affiliated Hospital, School of Medicine, Zhejiang University, Hangzhou, Zhejiang Province 310003, China

[9]Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Guangzhou 510060, China

[10]Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore 637551, Republic of Singapore

[11]Ludwig-Maximilians-Universität München (LMU), 80539 Munich, Germany

[12]Department of Hepatobiliary Surgery, The Affiliated Hospital of Southwest Medical University, Luzhou, Sichuan Province 646000, China

[13]Zhejiang Provincial Key Laboratory of Pancreatic Disease, The First Affiliated Hospital, School of Medicine, Zhejiang University, Hangzhou, Zhejiang Province 310003, China

[14]These authors contributed equally

[15]Lead contact

*Correspondence: xuanwen.bao@zju.edu.cn (X.B.), shengjp@zju.edu.cn (J.S.), software233@zju.edu.cn (J.R.), zhaop@zju.edu.cn (P.Z.)

https://doi.org/10.1016/j.xcrm.2024.101399

SUMMARY

Colorectal cancer (CRC) is a common malignancy involving multiple cellular components. The CRC tumor microenvironment (TME) has been characterized well at single-cell resolution. However, a spatial interaction map of the CRC TME is still elusive. Here, we integrate multiomics analyses and establish a spatial interaction map to improve the prognosis, prediction, and therapeutic development for CRC. We construct a CRC immune module (CCIM) that comprises FOLR2$^+$ macrophages, exhausted CD8$^+$ T cells, tolerant CD8$^+$ T cells, exhausted CD4$^+$ T cells, and regulatory T cells. Multiplex immunohistochemistry is performed to depict the CCIM. Based on this, we utilize advanced deep learning technology to establish a spatial interaction map and predict chemotherapy response. CCIM-Net is constructed, which demonstrates good predictive performance for chemotherapy response in both the training and testing cohorts. Lastly, targeting FOLR2$^+$ macrophage therapeutics is used to disrupt the immunosuppressive CCIM and enhance the chemotherapy response *in vivo.*

INTRODUCTION

Colorectal cancer (CRC) is one of the most common malignancies worldwide. Mortality from CRC has decreased slightly over the past 30 years due to earlier diagnosis and developments of chemotherapy, targeted therapy and immunotherapy. XELOX (combination of oxaliplatin and capecitabine), FOLFOX (combination of 5-fluorouracil, leucovorin, and oxaliplatin), and FOLFIRI (combination of 5-fluorouracil, leucovorin, and irinotecan) are the first-line chemotherapy regimens in late-stage CRC. Cetuximab and bevacizumab exhibit good efficacy, mainly in combination with systematic chemotherapy.[1,2] In addition, patients with microsatellite instability-high (MSI-H) or mismatch repair-deficient (dMMR) CRC are susceptible to immune checkpoint inhibitor (ICI)-based immunotherapy with objective clinical responses,[3] providing an alternative treatment for a subset of CRC patients with advanced disease.

Nevertheless, a heterogeneous treatment response has been found in CRC patients, possibly due to heterogeneity of the tumor microenvironment (TME). The TME consists of distinctive and interacting cell populations, including myeloid cells, T cells, B cells, natural killer (NK) cells, and other cell types. Studies of the diversity and reprogramming of the TME have shown its potential to influence immunotherapy response and prognosis.[4–11] The development of a suppressive TME limits the response to chemotherapy and immunotherapy treatment and is characterized by increased levels of pro-tumor macrophages, exhausted T cells, and regulatory T (Treg) cells. A high level of macrophages is associated with a poor prognosis in various tumor types. Targeting macrophages is a promising strategy for cancer therapy.[12,13] The colony-stimulating factor 1 (CSF1)/CSF1 receptor (CSF1R) axis has gained the most attention in the context of macrophage-targeting therapies. A variety of monoclonal antibodies and small molecules targeting CSF1R or its ligand CSF1 are in phase I studies as monotherapies or in combination with chemotherapy as well as other cancer immunotherapy approaches.[8] Nevertheless, anti-CSF1R therapy has shown limited therapeutic effects and significant toxicity in several tumor types, including CRC,[8] which warrants the exploration of the mechanism underlying the resistance to macrophage-targeting therapy.

The CRC TME has been characterized well at single-cell resolution.[14–18] However, most of the previous studies focused on the identification of different cell populations. A spatial interaction map of the CRC TME is still elusive. Moreover, the generation of high-throughput bulk and single-cell transcriptome data relies on fresh tumor tissues rather than paraffin sections. Transcriptomics data-driven patient stratification is not pathology friendly. The prognosis and the therapeutic sensitivity prediction of CRC patients are also not established well by pathological methods. To solve these problems, in this study, we integrated multiomics data (bulk transcriptome, single-cell transcriptome, and multiplex immunohistochemistry [mIHC] staining) to uncover an immunospatial interaction map of CRC. Our work highlights that the FOLR2$^+$ macrophage-mediated CRC immune module (CCIM) orchestrates a suppressive CRC TME and leads to therapy resistance in CRC. A seven-color CCIM mIHC panel was designed. Deep learning models were applied to the CCIM mIHC panel to develop a CCIM-Net for artificial intelligence (AI)-assisted diagnosis, which showed the properties of being efficient, pathology friendly, and cost effective. Last, we evaluated the possibility of the multicellular module being a promising therapeutic target.

RESULTS

Deeping learning-assisted multiomics integration workflow for CRC

The detailed workflow of our research is shown in Figure 1A. We performed the integration analysis on single-cell RNA sequencing (scRNA-seq) (public, n = 18; in house, n = 8), bulk transcriptome (total, n = 2,081; n = 1,673 for the integrative cohort, and n = 408 for four independent cohorts), and mIHC staining (n = 226) to construct a cell module for prognosis stratification and chemotherapy prediction. A multicellular module was identified as the CCIM based on the integrative analysis of scRNA-seq and bulk RNA-seq. With the established CCIM, a CCIM score was calculated to stratify patients with a poor prognosis in the training (n = 1,405) and 5 testing cohorts (total, n = 682) (total n = 2,081). Furthermore, we developed an mIHC imaging panel based on the CCIM to spatially detect the CCIM pattern. A CCIM deep learning model (CCIM-Net) based on the previous established CCIM pattern was developed with two cohorts from different centers (training cohort, n = 181; testing cohort, n = 45) to facilitate precision chemotherapy. *In vivo* experiments were performed for validation.

We first explored the TME heterogeneity and molecular signature for the cellular components in CRC using scRNA-seq (Figures 1B and 1C). T-distributed stochastic neighbor embedding (t-SNE) was performed on variably expressed genes across all cells for dimensionality reduction and cell clustering. The cells were obtained from 18 patients by integrating the public dataset and our in-house cohort (Figure S1). To define the identity of each cell cluster, the gene signature was determined by performing differential gene expression analysis (Figure 1C). Cell type signatures were used to identify cell clusters using the "SingleR" package with manual assistance. For instance, the expression of *NKG7*, *CCL5*, *CD3D*, and *CD3E* was significantly higher in T cells than in other cell clusters. Myeloid cells highly expressed *S100A8*, *S100A9*, and *PLAUR*. *CD79A* and *CD79B* were used to annotate B cells. In total, we found six cell types, which are shown in Figure 1C. Altogether, the integrated scRNA-seq landscape covers the major cellular components of the CRC TME, including various immune cells from all stages of CRC tumor tissues, based on which we applied a deep learning-assisted multiomics integration workflow for CRC.

Myeloid subpopulations and their immune activities in CRC

Next, we determined the myeloid populations in CRC. The tumor myeloid populations were further subclustered into dendritic cells (DCs), mast cells, FCN1$^+$ monocytes,[19] OAS$^+$ macrophages,[20] OLR1$^+$ macrophages,[21] HSPA6$^+$ monocytes,[22] and FOLR2$^+$ macrophages (see Figure 2A and Table S1 for the featured genes in each population). The key lineage markers are shown in Figure 2B. For instance, *CD1C* is highly expressed in DCs, while *KIT* is expressed on mast cells. *CD14*, *CD16*, and *CD68* were also investigated for monocyte and macrophage populations. With the help of the lineage markers, the DC and mast cell populations were identified. In the next step, we investigated the subpopulations of monocytes and macrophages. Monocytes could be divided into FCN1$^+$ and HSPA6$^+$ monocytes, respectively, characterized by expression of classic monocyte markers (*S100A8/9/12*), pro-inflammatory cytokines (*IL1B*, *IL6*, *CCL3*, and *CCL4*), and heat shock proteins (*HSPA6* and *HSPA1A/B*).[22] FCN1$^+$ monocytes have been identified in bronchoalveolar immune cells,[23] which fuel inflammation during severe disease. We also identified FCN1$^+$ and HSPA6$^+$ monocyte subpopulations (Figure S2A). *OLR1* was more highly expressed in pro-inflammatory M1-like macrophages and DCs than M2-like macrophage

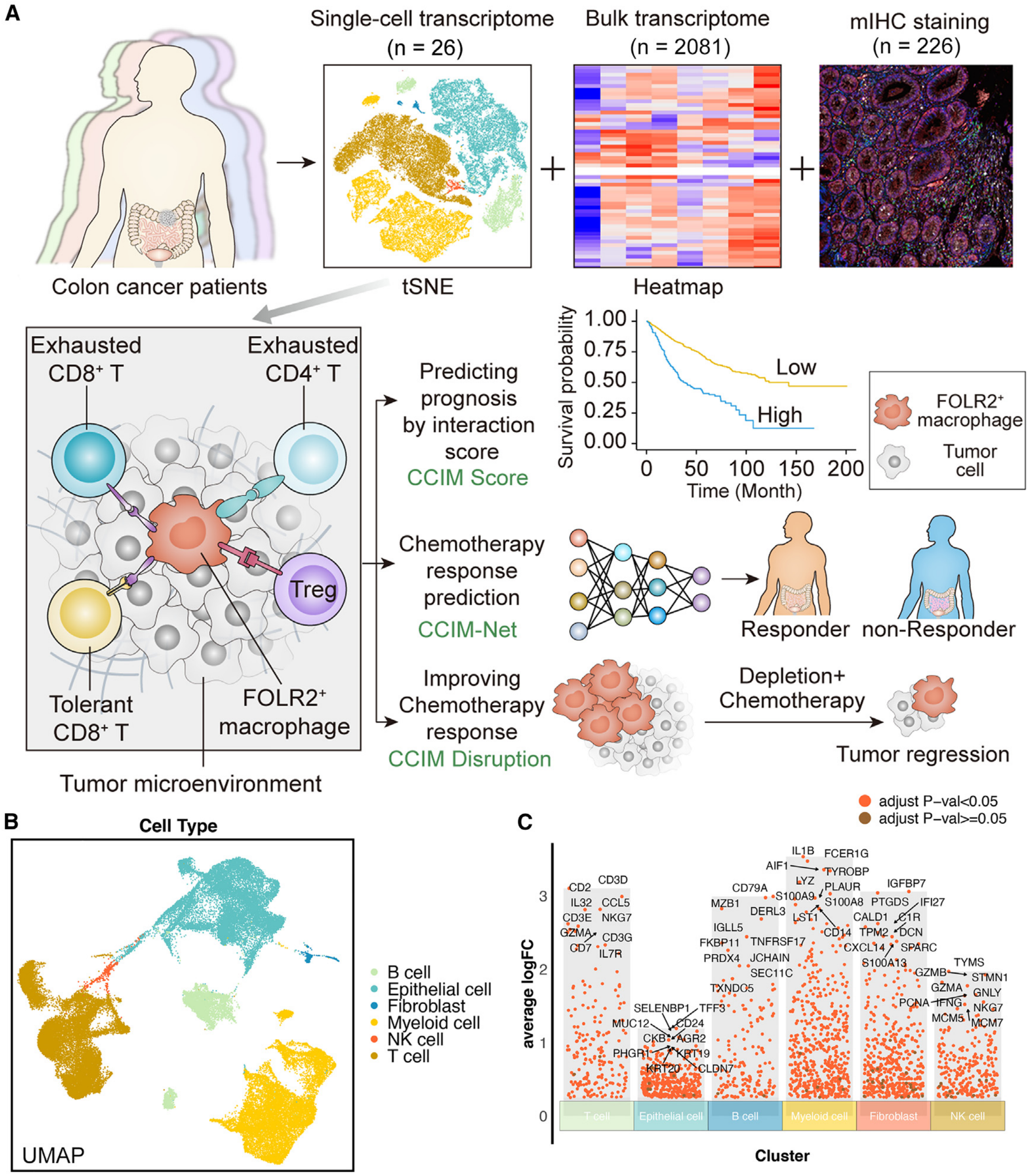


Figure 1. Depiction of the workflow for multiomics-assisted deep learning analysis

(A) Schematic of this research.

(B) The uniform manifold approximation and projection (UMAP) plot identifies the main cell types in colorectal cancer (CRC).

(C) Diverse cell types of CRC and the expression of markers.

subpopulations.[21] *OAS1* is upregulated by type I interferons (IFNs) and is also involved in the innate immune response. Therefore, we named the two macrophage subpopulations that highly express *OLR1* or *OAS1* OLR1$^+$ macrophages and OAS$^+$ macrophages, respectively (Figure S2A). Resident macrophages featured high expression of *FOLR2*, *C1QA*, *C1QB*, and *APOE*.[24] Previous studies have shown that *FOLR2* is highly expressed in resident macrophages and related to the activation of resident macrophages.[25–27] We therefore considered that FOLR2$^+$ macrophage exhibited a resident phenotype.

In the next step, Gene Ontology (GO) analysis was performed to characterize the underlying biological processes of each monocyte/macrophage population. FOLR2$^+$ macrophages were enriched in lipid metabolism-related processes, while OLR1$^+$ macrophages were enriched in G protein-coupled receptor and growth factor receptor binding (Figure 2C). Gene

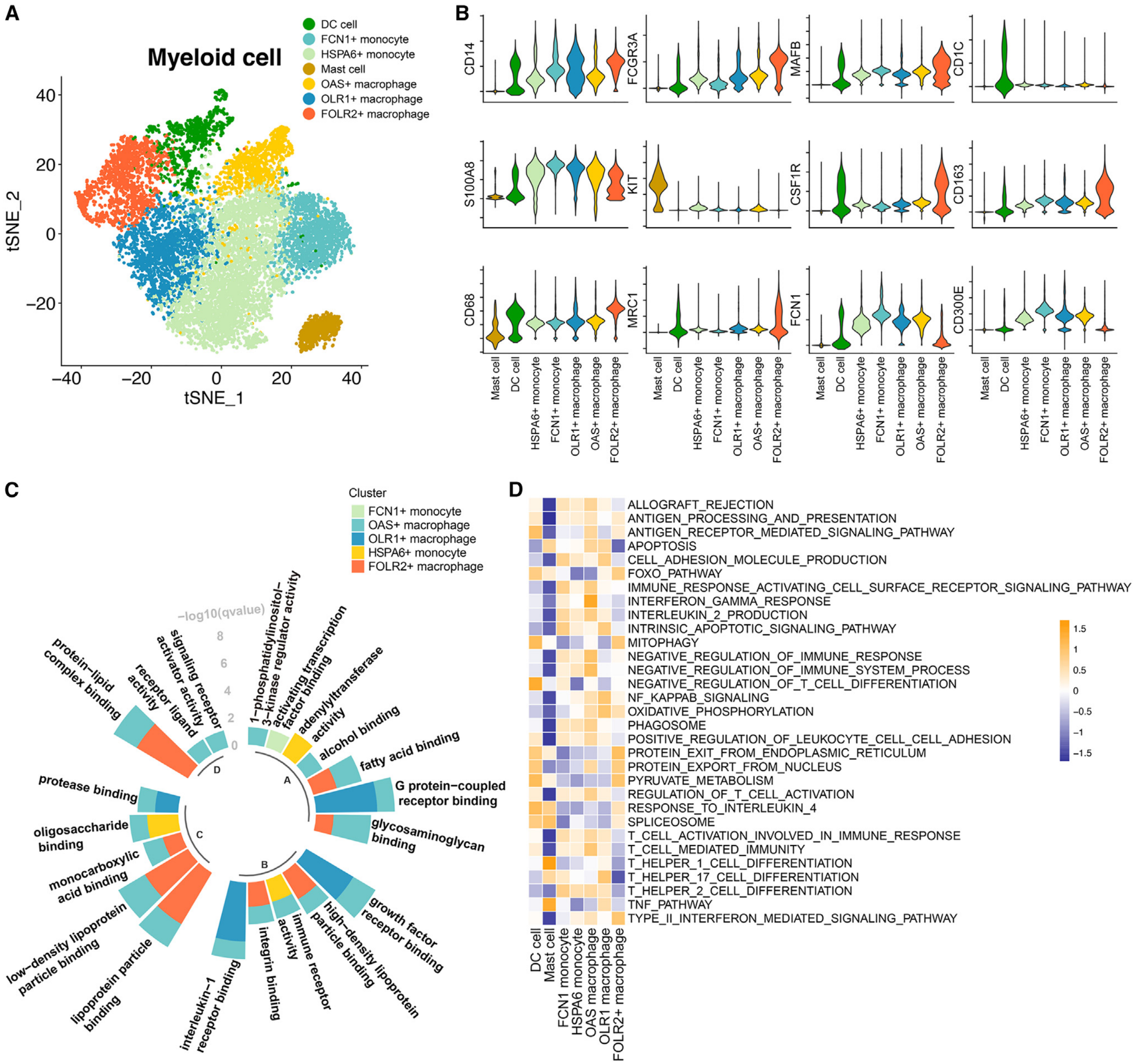


Figure 2. The properties of myeloid subpopulations

(A) The t-distributed stochastic neighbor embedding (t-SNE) shows seven subclusters of myeloid cells (patients, n = 26).

(B) Stacked violin plot showing the scaled expression level of markers for each population.

(C) The enriched Gene Ontology (GO) terms for each monocyte/macrophage population. The GO analysis was performed with the gene signature for each monocyte and macrophage subpopulation. The height of each column represents the $-\log_{10}(p)$ value of enriched biological processes for each subpopulation.

(D) Heatmap showing the immune pathway activities in myeloid populations. Purple indicates low values, while yellow indicates high values.

set variation analysis (GSVA) was performed with an immune gene set[28] on the identified myeloid populations (Figure 2D), which indicated significantly lower anti-tumor immunity activities in FOLR2$^+$ macrophage populations than in the other monocyte and macrophage populations (Figure 2D). To determine the effects of myeloid cell populations on the prognosis of CRC patients, we combined bulk transcriptome and scRNA-seq data for further analysis. An integrated cohort that included 1,673 CRC patients was selected. With the gene signature of each myeloid subpopulation identified from scRNA-seq, the association of the myeloid population and recurrence-free survival (RFS) of CRC patients was analyzed. We observed that patients with enrichment of FCN$^+$ monocytes, DCs, mast cells, OLR1$^+$ macrophages, HSPA6$^+$ monocytes, and FOLR2$^+$ macrophages exhibited worse RFS survival (Figure S2B). Taken together, the myeloid populations and their

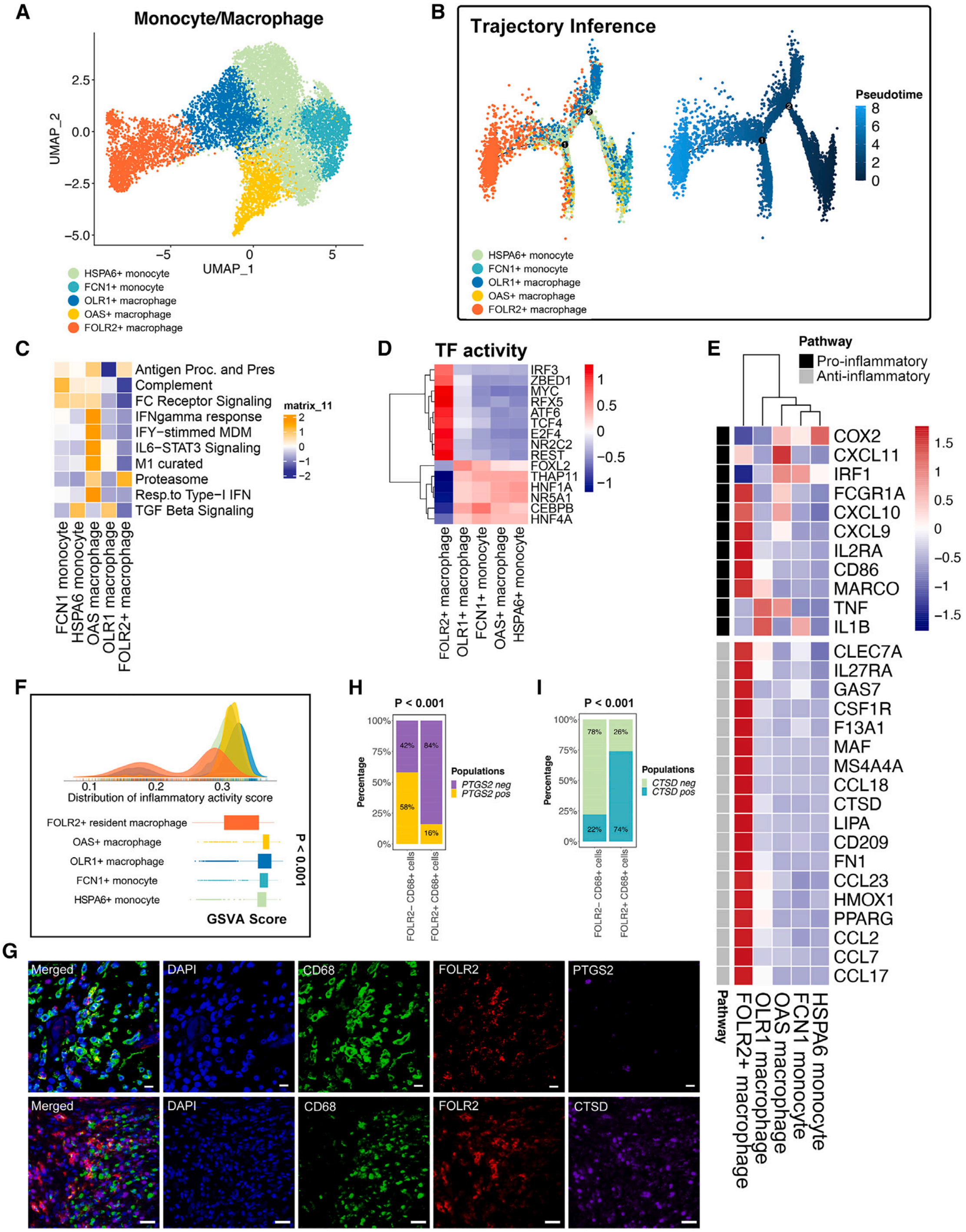


Figure 3. The anti-inflammatory properties of FOLR2$^+$ macrophages

(A) The UMAP plot shows five monocyte/macrophage subpopulations (patients, n = 26).

(B) Pseudotime analysis of five subclusters of monocytes/macrophages (patients, n = 26). The color key from deep blue to light blue indicates the pseudotime score from low to high.

(C) Heatmap of the single-sample gene set enrichment analysis (ssGSEA) scores for gene sets in monocyte/macrophage subtypes, calculated according to single-cell RNA sequencing (scRNA-seq) data.

(D) Transcription factor (TF) activity in monocyte/macrophage subpopulations.

(E) Heatmap of the expression of M1-related and M2-related genes from the scRNA-seq data.

(legend continued on next page)

immune activities were identified in CRC patients, which helped to identify the CCIM in our study.

Functionality difference of FOLR2$^+$ macrophages compared with other macrophage populations

To further analyze the functionality difference of the identified macrophage populations (Figure 3A), we performed the analyses below. Single-cell trajectory analysis revealed the pseudotime differentiation trajectory of monocytes/macrophages. In particular, the FOLR2$^+$ macrophage subpopulation is located at the terminal differentiation site (Figure 3B). Distinct macrophage-related signatures were associated with aspects of macrophage activity, including antigen presentation, the M1 pathway, complement, the proteasome, FC receptor signaling, and interferon (IFN)-gamma response function, and were observed in the different monocyte/macrophage populations (Figure 3C). M1 pathway, complement, FC receptor signaling and IFN-gamma response functions were downregulated in FOLR2$^+$ macrophages compared with other macrophage populations (Figure 3C). Furthermore, the transcription factor (TF) analysis suggested significantly distinct TF activities among FOLR2$^+$ macrophages and other monocyte/macrophage subpopulations (Figure 3D). For instance, *CEBPB* is a TF that plays crucial roles in macrophages differentiating into the pro-inflammatory spectrum.[29] FOLR2$^+$ macrophages showed significantly less expression of *CEBPB* compared with the other populations. In contrast, macrophage *TCF4*, which acts as a crucial Wnt pathway regulator and contributes hypoinflammation M2 polarization,[30] was significantly more expressed in FOLR2$^+$ macrophages (Figure 3D). In parallel, distinct pro-inflammatory and anti-inflammatory expression patterns were observed in the monocyte/macrophage populations with a previously defined gene set (Figure 3E). We therefore confirmed that FOLR2$^+$ macrophages showed a more M2-like phenotype than other macrophage subpopulations (Figure 3E). M2-like macrophages exhibit anti-inflammatory characteristics, while M1-like macrophages feature pro-inflammatory characteristics. We investigated the inflammatory properties in each monocyte/macrophage subpopulation. The density plot showed that a large proportion of FOLR2$^+$ macrophages exhibited lower inflammatory response GSVA scores than the other monocyte and macrophage populations, indicating a hypoinflammatory status of FOLR2$^+$ macrophages compared with other macrophage populations (Figure 3F). To validate the results from the single-sample gene set enrichment analysis (ssGSEA), we selected one typical inflammation marker, *PTGS2* (commonly known as COX2), as well as one anti-inflammation marker, *CTSD*, for validation using mIHC technology. FOLR2$^+$ macrophages exhibited lower levels of the inflammatory markers *PTGS2* while showing higher levels of the anti-inflammatory markers *CTSD* compared with other monocytes/macrophages (Figures 3G–3I). Taken together, the terminal differentiation trajectory location, distinct TF activities, gene expression pattern, and macrophage-related pathway alterations in the FOLR2$^+$ macrophage population suggested their special properties, which makes it possible to develop new therapeutics that employ tumor-associated macrophage (TAM) subset targeting. We then focused on FOLR2$^+$ macrophages and their relationship with T cell populations for CCIM construction.

The components and pathway heterogeneity of T cells in CRC

We next determined the T cell subpopulations and key genes expressed in T cells. CD4$^+$ T cells were further subclustered into 5 cell populations: exhausted CD4$^+$ T cells, memory CD4$^+$ T cells, naive CD4$^+$ T cells, helper T cell 17 (Th17) cells, and Treg cells (Figure 4A; see Table S2 for the featured genes in each population). Treg cells showed high expression of *FOXP3*. Naive CD4$^+$ T cells exhibited high *CCR7* and *SELL* expression. Th17 cells expressed GIMAP-family genes, and exhausted T cells were identified to have high *PDCD1* and *TIGIT* expression. CD8$^+$ cells were subclustered into tolerant CD8$^+$ T cells, effector memory CD8$^+$ T cells, exhausted CD8$^+$ T cells, exhausted CD8$^+$ NK T cells, and tissue-resident CD8$^+$ T cells (Figure 4B; see Table S3 for the featured genes in each population). In terms of the feature genes, effector memory CD8$^+$ T cells showed high expression of *GZMK* and *GIMAP4*. Exhausted NK T cells highly expressed *LAG3* and *NCR3*. Tissue-resident CD8$^+$ T cells exhibited high *S100A4* expression. Great expression of *DKK3* was found on tolerant CD8$^+$ T cells and *TMIGD2*, and *PDCD1* was identified as the feature gene for exhausted CD8$^+$ T cells.

Next, the key genes involved in T cell regulation pathways, such as cytolytic, cytokine, and inhibitory pathways, were identified (Figure 4C), revealing the heterogeneity of T cell activation. For instance, *ZNF683* acted as a TF and was mainly expressed in tissue-resident CD8$^+$ T cells. *GZMA*, *GZMB*, *NKG7*, and other cytolytic genes were mainly expressed in CD8$^+$ T cell subpopulations. Exhausted NK T cells and tissue-resident CD8$^+$ T cells still had the capacity to release *IFNG*. The C7 immunological signature GSVA score alterations further revealed a suppressive TME in CRC and the modulation of T cell exhaustion (Figure 4D). Last, the pseudotime analysis suggested that *PDCD1* was highly expressed in CD8$^+$ T cells that underwent the exhaustion process (Figures 4E–4G). In parallel, high expression of *PDCD1* was also observed during the exhaustion process of CD4$^+$ T cells (Figures S3A–S3C). Taken together, the main T cell populations were identified in our RNA-seq analysis, which showed distinct gene expression pattern and pathway activation. We

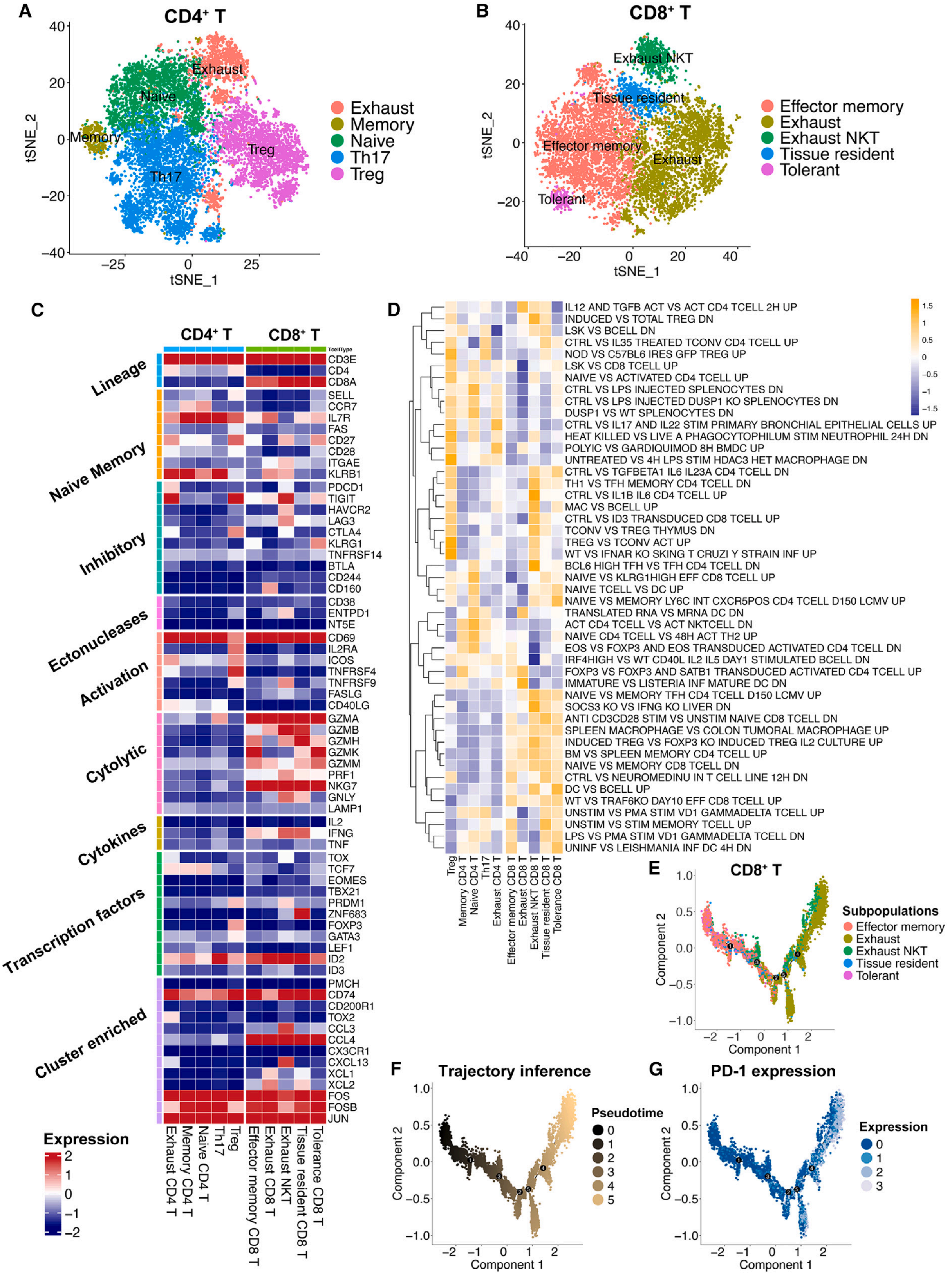
---

(F) Density plot indicating the density distribution of the inflammatory response activity score in monocyte and macrophage populations and boxplot indicating the inflammatory response activity score in monocyte and macrophage populations. The inflammatory response activity score was calculated with the ssGSEA method using the inflammatory response signature on the single-cell transcriptome. The p value was calculated with a Kruskal-Wallis test.

(G) Multiplex immunohistochemistry (mIHC) staining for CD68 (green), FOLR2 (red), PTGS2 (purple), CTSD (purple), and DAPI (blue). Scale bar, 10 μm. Biological replicates, n = 5.

(H) The percentage of FOLR2$^+$ CD68$^+$ cells and FOLR2$^-$ CD68$^+$ cells expressing PTGS2. A total of 100 cells were included for analysis. Significance was evaluated by a chi-square test.

(I) The percentage of FOLR2$^+$ CD68$^+$ cells and FOLR2$^-$ CD68$^+$ cells expressing CTSD. Significance was evaluated by a chi-squared test.



(legend on next page)

therefore investigated the relationship between the T cell subpopulations and FOLR2$^+$ macrophages.

The FOLR2$^+$ macrophage-oriented CCIM is composed of immunosuppressive T cell populations

The complex cell crosstalk was investigated to reveal the regulation of the suppressive CRC TME. The cell-cell interactions (CCIs) between each cell type were identified.[31] FOLR2$^+$ macrophages showed strong CCIs activities with multiple cell populations (Figure 5A). We next focused on the interaction between macrophage populations and T cell populations. FOLR2$^+$ macrophages showed the greatest number of CCIs with both CD4$^+$ T and CD8$^+$ T cells (Figures 5B and 5C; Table S4). In terms of the T cell subpopulations, exhausted CD4$^+$ T cells, Treg cells, tolerant CD8$^+$ T cells, and exhausted CD8$^+$ T cells showed the most CCIs with FOLR2$^+$ macrophages than other T cells. Taken together, these findings indicate that FOLR2$^+$ macrophages have abundant interactions with exhausted CD4$^+$ cells, exhausted CD8$^+$ T cells, tolerant CD8$^+$ T cells, and Treg cells, which may affect the formation of a suppressive CRC TME. The topological pattern of the FOLR2$^+$ macrophages, exhausted CD4$^+$ cells, exhausted CD8$^+$ T cells, tolerant CD8$^+$ T cells, and Treg cells in the CRC TME was validated by mIHC staining (Figure 5D). Large proportions of exhausted CD4$^+$ cells (20.3%), exhausted CD8$^+$ T cells (25.2%), tolerant CD8$^+$ T cells (16.6%), and Treg cells (24.1%) cells were found in the area around FOLR2$^+$ macrophages (Figure 5E), which further confirmed the spatial interaction of these cells. Last, the Voronoi plot suggested aggregation of FOLR2$^+$ macrophages with Tregs cells, exhausted CD4$^+$ T cells, tolerant CD8$^+$ T cells, and exhausted CD8$^+$ T cells in CRC tissues (Figure 5F). Therefore, a CCIM was defined as a module that contained at least one FOLR2$^+$ macrophage, one tolerant CD8$^+$ T cell, one Treg cell, and one exhausted CD4$^+$ T cell (Figure 5G). The high density of the CCIM numbers implied the important influence of CCIM in the suppressive CRC TME (Figure 5H). The interaction between FOLR2$^+$ macrophages and the other CCIM cell populations was identified, showing the suppressive interaction effects from FOLR2$^+$ macrophages on others (Figure 5I). Significant associations of FOLR2$^+$ macrophages with other CCIM cell components were identified, showing the aggregation effects of CCIM (Figures 5J–5L).

The weighted gene co-expression network analysis (WGCNA) method infers the co-expression network and related biological mechanisms in high-throughput data. We therefore applied WGCNA to CRC bulk transcriptome data to explore the potential regulation roles underlying CCIM. A power of $\beta = 3$ was chosen as the optimal soft threshold to ensure scale-free co-expression (Figures S4A–S4C). The FOLR2$^+$ macrophage score was calculated with the GSVA method for each patient. The unsupervised clustering method was used for obtaining the clustered gene modules. Eleven non-gray (meaningful) gene modules were identified, and the yellow module showed the most significant correlation with the FOLR2$^+$ macrophage GSVA score (Figure S4D). Therefore, the genes involved in the yellow module were identified as the key genes associated with the FOLR2$^+$ macrophage score. A total of 1,052 genes were identified, which suggested a significant association between the module membership score (the central point score for each module calculated by the WCGNA method) and gene significance (the Pearson's coefficient between the target gene expression and the FOLR2$^+$ macrophage score in the cohort), indicating the importance of those genes in the FOLR2$^+$ macrophage regulating network (Figures S4E and S4F). Using those genes as input for Kyoto Encyclopedia of Genes and Genomes (KEGG) and GO analysis, we further found that T cell-related pathways were altered, which confirmed the immunosuppressive regulation potential of FOLR2$^+$ macrophages (Figures S4G and S4H). Briefly, we utilized the scRNA-seq technology to identify a CCIM cell module whose function was further confirmed by bulk transcriptome analysis.

Lastly, we developed a CCIM scoring system to stratify CRC patients with a distinct prognosis. The integrated cohort (GSE39582, GSE17536, GSE17537, GSE14333, GSE56699, GSE37892, and GSE33113) was used as the training cohort. The top 20 featured genes in each CCIM cell component and the ligand-receptor pairs between FOLR2$^+$ macrophage and other CCIM cell components (Figure 5I) were used as inputs for the least absolute shrinkage and selection operator (LASSO) Cox regression model. A total of 17 features (genes) were selected by the regression model to build the CCIM scoring system (Figure S5A). With the coefficients (Figure S5B) and the distinct expression pattern of the 17 features, the CCIM scores were calculated for each patient. The results indicated that patients with high CCIM scores exhibited poorer survival probability than patients with low CCIM scores (Figure 5M; best-cutoff method for threshold of high and low score). Multivariate Cox analysis suggested that the CCIM score is an independent risk factor for patients with CRC in The Cancer Genome Atlas (TCGA) colon adenocarcinoma (COAD) cohort (Figures S5C and S5D). Furthermore, five independent cohorts (GSE28722, GSE12945, GSE72970, GSE5851, and TCGA-CRC) validated the robustness of the CCIM score in stratifying CRC patients (Figures S5E–S5I). Pooled estimates revealed that a lower CCIM

Figure 4. The subpopulations of T cells showed distinct gene expression patterns and pathway activation levels

(A) The t-SNE plot identifies five subclusters for CD4$^+$ T cells (patients, n = 26).

(B) The t-SNE plot identifies five subclusters for CD8$^+$ T cells (patients, n = 26).

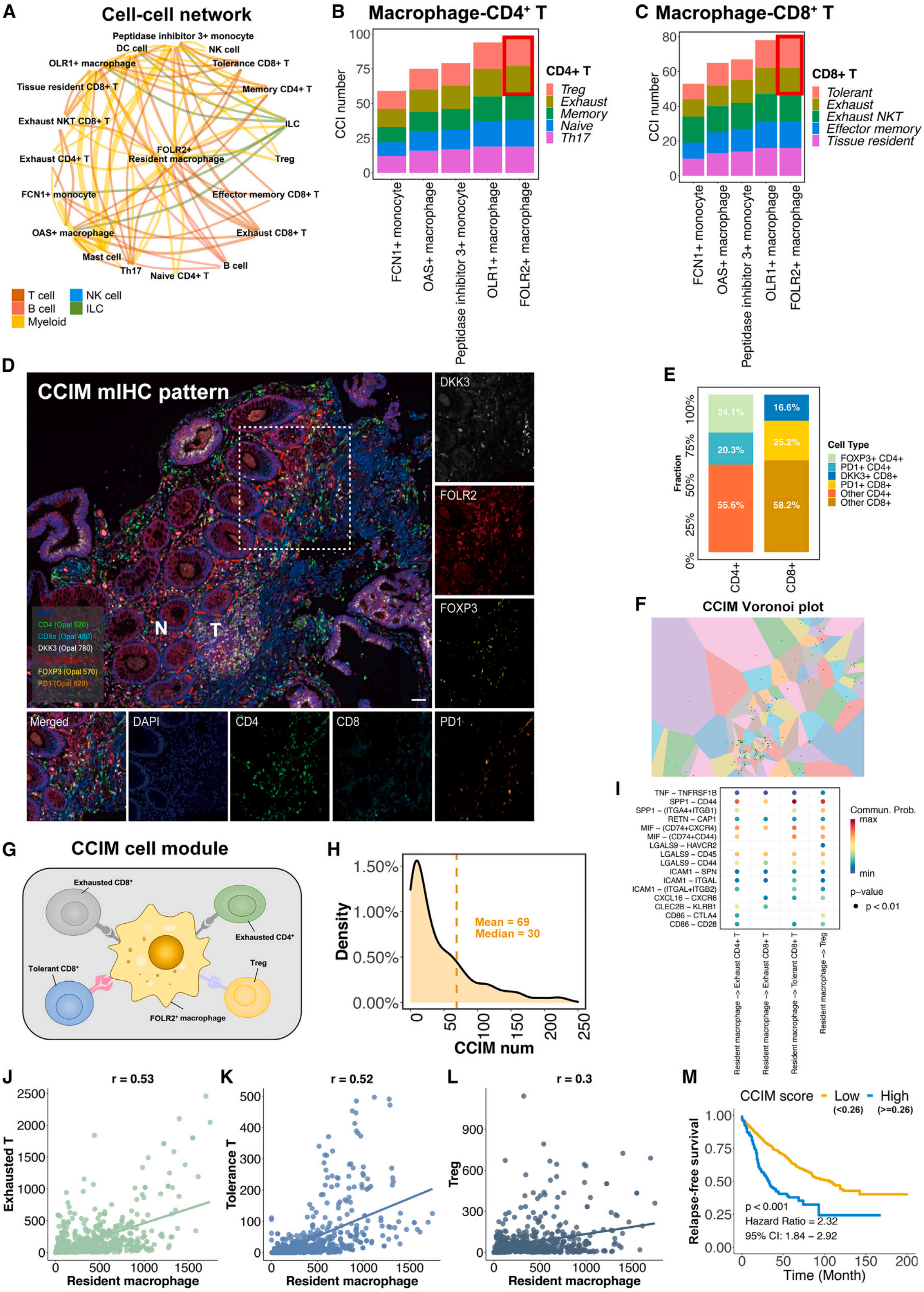
(C) Heatmap showing gene expression in key pathways regulating T cell activities. The color key from blue to red indicates relative expression levels from low to high.

(D) Heatmap representation of the gene set variation analysis (GSVA) results based on immunologic signature C7. The color key from purple to yellow indicates the relative GSVA score from low to high.

(E) Trajectory inference by monocle2 method on five subclusters of CD8$^+$ T cells.

(F) Trajectory inference with the pseudotime score for five subclusters of CD8$^+$ T cells. The color key from black to gold indicates the pseudotime score from low to high.

(G) Trajectory inference with *PDCD1* expression for five subclusters of CD8$^+$ T cells. The color key from blue to white indicates *PDCD1* expression from low to high.



(legend on next page)

score was significantly associated with improved survival probability (hazard ratio [HR] 2.04, 95% confidence interval [CI] 1.56–2.67, $p < 0.001$; Figure S5J). In addition, the GSVA algorithm showed significantly increased anti-PD1 resistance in patients with greater CCIM scores by calculating the anti-PD1 resistance-related gene set activities[32] (Figure S5K). Through the integration analysis of scRNA-seq and bulk RNA-seq, we identified a cell pattern consisting of FOLR2$^+$ macrophages, Treg cells, exhausted CD4$^+$ T cells, tolerant CD8$^+$ T cells, and exhausted CD8$^+$ T cells. We named this cell pattern the CCIM module, with its features depicted through mIHC and related image analysis.

The FOLR2$^+$ macrophage-oriented CCIM module contributes to AI-guided precision chemotherapy

Several studies have shown the capacity of AI-assisted analysis of transcriptome data to predict therapy response, molecular subtyping, and pathway activation. Nevertheless, the efficiency and accuracy were unstable among different cohorts.[33–35] We next assessed the potential relationship between FOLR2$^+$ macrophages and the chemotherapy response. With the application of a neural network, we built an FOLR2-RM model based on the FOLR2$^+$ macrophage signature to predict the chemotherapy response in the training cohort and then assessed the prediction efficiency of the FOLR2-RM model in an external cohort (Figure S6A). The confusion matrix and receiver operating characteristic (ROC) plot revealed an acceptable predictive efficiency of the FOLR2-RM model with a relative high area under the curve (AUC) value (AUC = 0.797), which also indicated potential important roles of FOLR2$^+$ macrophages in the response to chemotherapy in CRC patients (Figures S6B and S6C).

However, the AUC value of FOLR2-RM is not as great enough to help with clinical decisions. We then assessed the potential application of the CCIM module to precision chemotherapy. Deep learning technology enabled the extraction of hidden information directly from image patterns and provided potentially clinically useful information. We therefore tested several advanced deep learning models for predicting the response to systemic first-line chemotherapy by CCIM image pattern. To further utilize the CCIM module and improve the prediction efficiency, we built a diagram using the AI-guided CCIM imaging pattern (Figure 6A). We established two medical center cohorts as training and validation cohorts (n = 181, 80% for the training cohort and 20% for the validation cohort) and an independent testing cohort (n = 45). The tumor tissues were obtained by colonoscopy and subjected to mIHC staining depicting the CCIM module before systemic chemotherapy. Then, the patients were treated with systemic chemotherapy, and the chemotherapy response was assessed according to the Response Evaluation Criteria in Solid Tumors (RECIST) standard. Typical computed tomography (CT) scans before and after chemotherapy and CCIM mIHC staining images are shown in Figure 6B. The tumor areas in the mIHC slides were aligned with the tumor areas in the corresponding HE staining slides. Only the tumor areas were used for the subsequent analysis in the CCIM pipeline. The non-responders exhibited greater enrichment of the CCIM per area than the responders among the total patients (Figure 6C). We trained 8 models (ResNet18, ResNet34, ResNet50, ResNet101, Xception, DenseNet121, and ResNext50) in the training and validation cohorts. The predictive accuracy of each model was tested in the independent testing cohort. The AUC value varied from 0.77 to 0.99 in the independent testing cohort (Figure 6D). Finally, the ResNext50 model was selected as the CCIM-Net. CCIM-Net showed the greatest AUC value (AUC = 0.99) and highest accuracy rate (98%) by the confusion matrix (Figures 6E and 6F). The detailed work flow of CCIM-Net construction is depicted in Figure S6D, which shows the four stages from input of the CCIM module to output of decision. Thus, we confirmed that the CCIM-Net is a robust AI-guided tool for clinical chemotherapy decision-making by integrating two independent cohorts.

Disruption of the CCIM by targeting FOLR2$^+$ macrophages improved the chemotherapy response

We next assessed the effect of the abundance of FOLR2$^+$ macrophages on chemotherapy outcomes by an *in vivo* model (Figure 7A). We developed a FOLR2-EGFP (enhanced GFP)-diphtheria toxin receptor (DTR) mouse strain to deplete the FOLR2$^+$ macrophages (Figure S7). The DTR-EGFP fusion sequence was integrated into the region between the 6th exon

Figure 5. Construction of the CRC immune module (CCIM)

(A) The cell-cell interaction (CCI) network for all cell subpopulations in CRC. The color of edges indicates the type of ligand cell, and the size of edges represents the CCI numbers between cells.

(B) The number of interactions between monocytes/macrophages and CD4$^+$ T cells inferred from single-cell RNA sequencing (scRNA-seq).

(C) The number of interactions between monocytes/macrophages and CD8$^+$ T cells.

(D) mIHC staining for FOLR2 (red), DKK3 (white), FOXP3 (yellow), PDCD1 (orange), CD4 (green), CD8 (light blue), and DAPI (blue) reveals the CCIM image pattern. Scale bar, 50 μm. The dotted line indicates the margin between tumor tissue and para-tumor tissue.

(E) The proportion of cells from CCIM and other cell types around FOLR2$^+$ macrophages. The statistics were determined with the area of 800 pixels around FOLR2$^+$ macrophages. Biological replicates, n = 5.

(F) Voronoi plot for the CCIM topological pattern. The green dot represents the center of each part.

(G) Depiction of the CCIM module. A CCIM module contains at least one FOLR2$^+$ macrophage, one tolerant CD8$^+$ T cell, one Treg cell, and one exhausted CD4$^+$ T cell.

(H) The density plot for the CCIM in CRC.

(I) Heatmap of cell-type-specific receptor-ligand interactions inferred by SingleCellSignalR. Shown are inferred interactions between FOLR2$^+$ macrophages and Treg cells, exhausted CD4$^+$ T cells, exhausted CD4$^+$ T cells, and tolerant CD8$^+$ T cells. Circle color indicates the inferred interaction score, and circle size indicates the mean expression of receptor and ligand genes for each pair.

(J–L) The association of exhausted T cells ($p < 0.01$) (J), tolerant T cells ($p < 0.01$) (K), and Treg cells ($p < 0.01$) (L) with FOLR2$^+$ macrophages. The p value was calculated with Pearson's coefficient, and $p < 0.05$ was considered to indicate significance.

(M) The CCIM score stratifies CRC patients with poor prognosis (patients, n = 1,405).

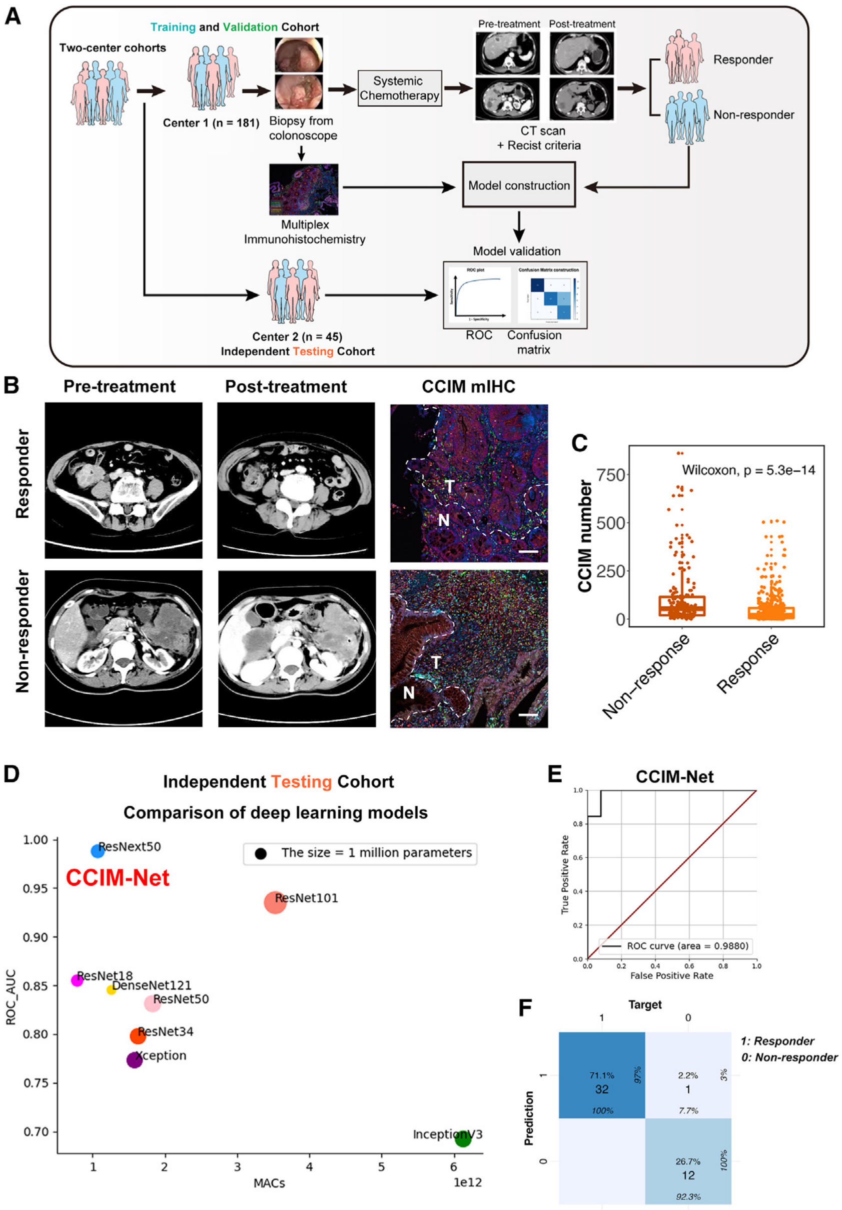


Figure 6. Artificial intelligence (AI)-guided precision chemotherapy based on CCIM imaging pattern

(A) Schematic of the AI-guided precision chemotherapy process.

(B) Representative computed tomography (CT) scan and CCIM mIHC pattern images for patients who responded or did not respond to chemotherapy. mIHC staining of FOLR2 (red), DKK3 (white), FOXP3 (yellow), PDCD1 (orange), CD4 (green), CD8 (light-blue) and DAPI (blue) was applied to portray the CCIM. Scale bar, 50 μm. The dotted line indicates the margin between tumor tissue and para-tumor tissue.

(C) Boxplot showing the CCIM numbers in the non-responder group and responder group in the training and validation cohorts. The significance was calculated by a Wilcoxon test, with $p < 0.05$ considered significant.

(D) Comparison of eight deep learning models in the independent testing cohort. The y axis indicates the area under the curve (AUC) value of each model in the independent testing cohort.

(E) The receiver operating characteristic (ROC) analysis shows the sensitivity and specificity of the CCIM-Net in the independent testing cohort.

(F) The confusion matrix for predicting the chemotherapy response by the CCIM-Net in the independent testing cohort. Columns indicate the predicted response, while rows represent the actual response.

and the noncoding region of the FOLR2 gene in a homologous recombination way after the genomic cutting from recombinant Cas9 nucleases and designed guide RNAs. The FOLR2-DTR-Apc^{Min} mice were bred by crossing FOLR2-DTR-EGFP mice with B6/JGpt-$Apc^{em1Cin(Min)}$/Gpt mice. Four groups were designated as control (Ctrl), 5-Fu treatment, diphtheria toxin (DT) treatment, and 5-Fu + DT treatment, and the 5-Fu + DT treatment group exhibited the lowest tumor burden (Figures 7B and 7C). H&E staining and Ki-67 IHC staining were used to assess the properties of the CRC tumor model (Figure 7D). Ki-67 IHC staining suggested reduced proliferation in the 5-Fu + DT treatment group compared with the other groups (Figure 7E). Precise ablation of FOLR2$^+$ (EGFP$^+$) macrophages could be observed after DT injection (Figure 7F). Taken together, these results indicated that anti-FOLR2$^+$ macrophage therapeutics led to disruption of the immunosuppressive CCIM module and enhanced chemotherapy response.

DISCUSSION

In this study, we comprehensively analyzed the molecular and cellular signature of CRC components through depiction of the single-cell landscape and integration of the bulk transcriptome and mIHC. Importantly, we determined the importance of FOLR2$^+$ macrophages, which showed distinct properties compared with other macrophage populations, in the formation of the immunosuppressive module affecting prognosis. A CCIM consisting of FOLR2$^+$ macrophages, exhausted CD4$^+$ T cells, exhausted CD8$^+$ T cells, tolerant CD8$^+$ T cells, and Treg cells was identified. We therefore developed a CCIM scoring system to serve as a clinically tractable tool for survival prediction and treatment guidance in patients with CRC. With the CCIM spatial interaction map-based mIHC pattern, we developed an AI-guided precision chemotherapy model, which showed robust prediction efficiency for chemotherapy sensitivity. The *in vivo* model further indicated that targeting FOLR2$^+$ macrophages disrupted the CCIM and improved the chemotherapy response; thus, we considered that disruption of the FOLR2$^+$ macrophage-mediated CCIM could be a promising therapy for CRC.

Zhang et al.[36] highlighted that myeloid populations and conventional dendritic cell subsets are key mediators of cellular crosstalk in the CRC TME.[36] In our study, we observed that the majority of DCs were type 2 conventional DCs (cDC2s), as evidenced by high expression levels of *CD1c* and *CLEC10A*. cDC2s were associated with Treg cell abundance and promoted Treg cell-mediated immunosuppression in hepatocellular carcinoma through a hypoxia-dependent mechanism.[37] Regarding CRC, cDC2s were found to be heterogeneous and could be further subdivided into several subpopulations.[38] Specifically, C1QC$^+$ cDC2s were associated with worse overall survival in TCGA COAD and rectum adenocarcinoma (READ) datasets.[38] The functions and heterogeneity of cDC2s in CRC warrant further exploration. TAMs are one of the main populations in myeloid cells and have been a promising target of immunotherapy. Macrophages in tumor tissues are traditionally classified into two subsets, the M1 and M2 populations, based on their distinguished molecular features and pathway differences.[39,40] M1 macrophages show proinflammatory properties and are considered an antitumor subset. In contrast, M2 macrophages, which feature anti-inflammatory and proangiogenic properties, promote tumor progression in various solid tumor types.[41–43] Nevertheless, we consider that the M1 and M2 classification may not be suitable in CRC considering the heterogeneity of CRC immune populations. A variety of TAMs in the TME confer both protection and vulnerability.[44] Targeting protumorigenic macrophages is a priority in cancer treatment, but in most tumor types, it is still unclear which TAM subsets have a protumorigenic role and when they should be targeted. Identifying the protumorigenic TAM subsets and their biological functions in CRC is thus crucial to develop new therapeutics that employ TAM subset targeting. We therefore explored the monocyte/macrophage populations in CRC by integrated scRNA-seq. FCN1$^+$ monocyte, HSPA6$^+$ monocyte, OAS$^+$ macrophage, OLR1$^+$ macrophage, and FOLR2$^+$ macrophage populations with distinct gene expression signatures were identified.

Macrophages are typically distributed throughout the villi of the lamina propria.[45] In our study, we obtained tumor tissues that contained both the stromal area and intraepithelial layer. The majority of macrophages we obtained were from the stromal area, which explains the high abundance of macrophages in our dataset. Resident macrophages within the stromal area exhibit a persistent anti-inflammatory phenotype, even in inflammatory settings, which may be essential for mucosal healing.[46,47] In agreement with this, depletion of resident macrophages has been shown to exacerbate experimental colitis.[48] In our dataset, we also observed the low inflammatory properties of FOLR2$^+$ macrophages. In tumor tissues, it is still not clear what the exact functions of resident macrophages in the TME are. One study showed that recruitment of macrophages in CCR2$^{-/-}$ mice to pancreatic ductal adenocarcinoma tumors was reduced.[49] Nevertheless, pancreatic ductal tumor growth is not affected by a reduction in macrophages, implying that pancreatic ductal tumor progression may be regulated by a CCR2$^+$ macrophage-independent mechanism. They then found that resident macrophages are able to self-renew in tumors with the aid of tumor-derived CSF-1 and that this process can promote tumor progression.[49] Nalio Ramos et al.[50] revealed that FOLR2$^+$ macrophages reside in a perivascular niche in the tumor stroma of breast cancer. The density of FOLR2$^+$ macrophages positively correlate with T cell infiltration and better prognosis. In our study, we found distinct TF, pathway activation, and expression patterns of FOLR2$^+$ macrophages compared with other macrophage populations. Active lipid metabolism-related pathways were enriched in FOLR2$^+$ macrophages. Given that M2-phenotype macrophages are macrophages that have anti-inflammatory properties and significant lipid metabolism activation, we considered that FOLR2$^+$ macrophages exhibited an M2-like resident phenotype in CRC and were correlated with a poor prognosis of CRC. Regarding the contradictory results between breast cancer and CRC, we investigated our signature of FOLR2$^+$ macrophages in breast cancer and found a positive association between FOLR2$^+$ macrophages and prognosis. Several factors may contribute to the observed inconsistencies between the two cancers. First, in CRC, FOLR2$^+$ macrophages exhibit numerous cell communication pairs with Treg, exhausted T, and tolerant T cells, which may act as immunosuppressive factors. In contrast, FOLR2$^+$ macrophages in breast cancer demonstrate the functional ability to activate CD8$^+$ T cells. Additionally, the distinct pathological features of breast cancer and CRC may influence the spatial distribution of FOLR2$^+$ macrophages and other immune cells. Specifically, FOLR2$^+$ macrophages reside in a perivascular niche in the tumor stroma of breast cancer, whereas the observed FOLR2$^+$ macrophages in this study predominantly originate from the stromal area. Therefore, FOLR2$^+$ macrophages from different organs may have distinct effects on the TME and prognosis of tumors of different origins.

In the next step, we detected abundant FOLR2$^+$ macrophages, Treg cells, exhausted CD4$^+$ T cells, exhausted CD8$^+$ T cells, and tolerant CD8$^+$ T cells in CRC tissues, which constitute a cell module named CCIM. Colocalization of FOLR2$^+$ macrophages, Treg cells, exhausted CD4$^+$ T cells, exhausted CD8$^+$ T cells, and tolerant CD8$^+$ T cells was observed in tumor tissues, indicating the potential roles of CCIM topology in reprogramming the suppressive TME in CRC. Most previous studies infer interaction networks between cell populations based on the expression of receptor-ligand pairs. However, the full spectrum of network relationships cannot be captured by discrete cell clusters considering the spatial distribution. Recently, imaging-based studies have highlighted CCI networks based on the colocalization of neighboring cell populations.[51] The so-called TME hub (or module) showed the capacity to regulate the suppressive immune TME and predict the immunotherapy response. For instance, Leader et al.[52] revealed a cellular module consisting of PDCD1$^+$CXCL13$^+$ activated T cells, immunoglobulin G (IgG)$^+$ plasma cells, and SPP1$^+$ macrophages, referred to as the lung cancer activation module (LCAM).[52] High baseline LCAM scores were correlated with enhanced immunotherapy response even in patients with above-median tumor mutation burden (TMB), suggesting that immune cell composition, while correlated with TMB, may be a nonredundant biomarker of the response to immunotherapy.[52] Pelka et al.[18] discovered

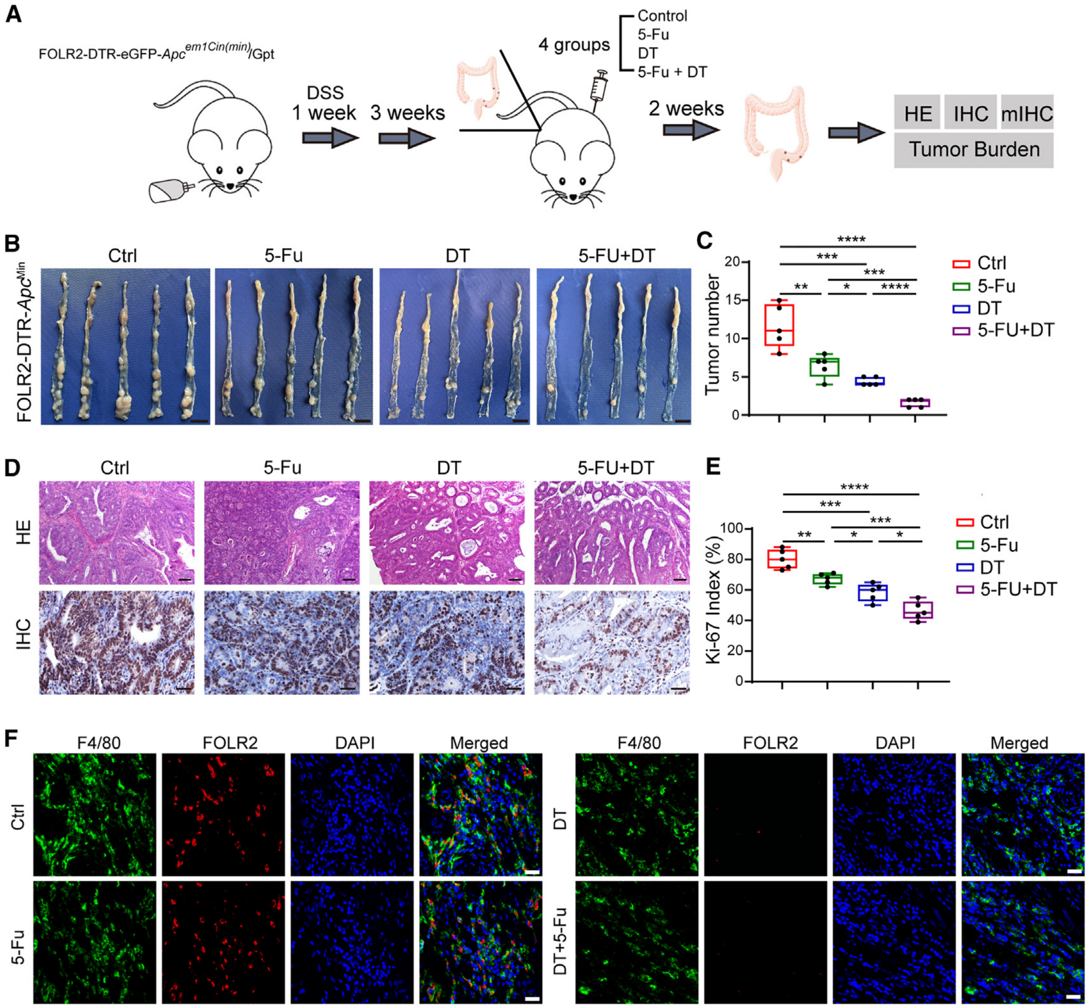


Figure 7. FOLR2$^+$ macrophage depletion disrupts the immunosuppressive tumor microenvironment (TME)

(A) The workflow of the *in vivo* experiment.

(B) The tumor burden from mice of four groups was assessed (n = 5).

(C) The statistics for tumor burden in each group. Significance was evaluated by a Kruskal-Wallis test (n = 5).

(D) H&E and Ki-67 IHC staining for the tumor tissues from four groups of mice (biological replicates, n = 5). Scale bars, 100 μm (H&E) and 40 μm (IHC)

(E) The statistics for Ki-67-positive cells in each group. Significance was evaluated by a Kruskal-Wallis test (n = 5).

(F) Multiplex immunohistochemistry staining of F4/80 (green) and FOLR2 (red) on tumor tissues in the *in vivo* model.

a myeloid cell-attracting hub at the tumor-luminal interface associated with tissue damage. In contrast, a dMMR-enriched immune hub within the tumor was identified, featuring activated T cells together with malignant and myeloid cells expressing T cell-attracting chemokines.[18] Nevertheless, most of the previous TME module research focused on depiction of the image pattern.

In recent years, AI-guided risk stratification and precision treatment for cancer have been promising methods for personized medicine.[4,6,53–56] Considering the special topological pattern of CCIM shown by the Voronoi plot, we developed a workflow for building an scRNA-seq-assisted and CCIM-based precision chemotherapy model. We applied cutting-edge deep learning technology to train 8 AI models and finally built a CCIM-Net, which showed high prediction accuracy in the testing cohort. Therefore, we demonstrate that the CCIM-Net based on the seven-colour mIHC panel is a robust tool for helping oncologists predict chemotherapy response. The CCIM-Net did not rely on high-throughput data generated from fresh tumor tissues and showed the properties of being efficient, cost effective, and pathology friendly. The AI-assisted tool may aid in the development of precise treatment strategies for CRC.

The anti-CSF1R antibody is a decent way to target TAMs and lead to their depletion.[57] Nevertheless, various solid tumors showed limited response to the CSF1R antibody. Zhang et al.[36] identified the persistence of the anti-inflammatory TAM population, and the loss of proinflammatory TAM populations is involved in resistance to anti-CSF1R therapy. FOLR2$^+$ macrophages showed a hypoinflammatory property in CRC that was not observed for other macrophage populations. Thus, targeting the distinct FOLR2$^+$ macrophage population may be more advantageous to overcome the drawback of general TAM blockade therapy. As expected, the depletion of the FOLR2$^+$ macrophages *in vivo* also showed a considerable antitumor effect. Therefore, we considered that targeting FOLR2$^+$ macrophages may be a promising method in combination with chemotherapy for CRC.

Taken together, our study reveals the CCIM topological pattern and uncovers the importance of CCIM components in reprogramming the suppressive CRC TME. The CCIM score serves as a clinically tractable tool for survival prediction. The CCIM-Net acts as a promising and effective opportunity for therapeutic intervention and precision chemotherapy. Targeting FOLR2$^+$ macrophages can alter the enrichment of the CCIM and holds promise as a therapeutic strategy. Therefore, we believe that the immune spatial interaction map of CCIM can serve as a more direct measurement of the immunosuppressive TME and provide information for clinical decision-making.

Limitations of the study

There are still some limitations to our study. First, background differences caused by the biopsy methods and length of data sequencing from the in-house cohort and the public cohort may exist. Second, two in-house samples make a predominant contribution to the myeloid and epithelial cell populations, which may affect the generalizability of these findings, although we applied large series of mIHC sections for validation. Last, we proposed that our AI-guided precision chemotherapy model contained information for clinical decision-making and obtained samples from our clinic to further validate this system. Multicenter cohort validation and further clinical trials should be considered.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - ○ Lead contact
  - ○ Materials availability
  - ○ Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - ○ Animal objects
  - ○ Human subjects
- METHOD DETAILS
  - ○ scRNA-seq processing
  - ○ Functional enrichment analysis
  - ○ Cell−cell interaction (CCI) analysis using SingleCellSignalR
  - ○ Weighted gene co-expression network analysis (WGCNA)
  - ○ Multiplex immunohistochemistry (mIHC) staining of tissue sections
  - ○ AI-guided chemotherapy model construction
  - ○ Haematoxylin and eosin (HE) staining
  - ○ Immunohistochemistry (IHC) staining
  - ○ Tandem mass tag (TMT)-based proteomic analysis
  - ○ Construction of the CCIM scoring system
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.xcrm.2024.101399.

ACKNOWLEDGMENTS

This work was supported in part by National Natural Science Foundation of China grants 82101830 (to X.B.), 81874173 (to J.R.), 81472346 (to P.Z.), 82074208 (to P.Z.), and 82102817 (to X.D.) and by Natural Science Foundation of Zhejiang Province grants Y23H160106 (to X.B.), LY20H160033 (to P.Z.), LQ22H160041 (to X.D.), and LY22H160019 (to J.R.). The results here, whole or in part, are based on data generated by the TCGA Research Network (https://www.cancer.gov/tcga).

AUTHOR CONTRIBUTIONS

Conceptualization, X.B.; methodology, X.B., Q.L., D.C., X.D., H.Z., Y.J., C.L., Y.W., C.Y., S.X., X.L., W.T., Y.D., Y.X., V.T., Y.W., W. Fu, S.D., and W. Fang; investigation, X.B., Q.L., D.C., X.D., and H.Z.; visualization, X.B., Q.L., X.D., and H.Z.; supervision, X.B., J.S., J.R., and P.Z.; writing – original draft, X.B.; writing – review & editing, X.B., J.S., J.R., and P.Z.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: March 6, 2023
Revised: January 2, 2024
Accepted: January 8, 2024
Published: February 1, 2024

REFERENCES

1. Alberts, S.R., Sargent, D.J., Nair, S., Mahoney, M.R., Mooney, M., Thibodeau, S.N., Smyrk, T.C., Sinicrope, F.A., Chan, E., Gill, S., et al. (2012). Effect of oxaliplatin, fluorouracil, and leucovorin with or without cetuximab on survival among patients with resected stage III colon cancer: a randomized trial. JAMA *307*, 1383–1393.
2. de Gramont, A., Van Cutsem, E., Schmoll, H.-J., Tabernero, J., Clarke, S., Moore, M.J., Cunningham, D., Cartwright, T.H., Hecht, J.R., Rivera, F., et al. (2012). Bevacizumab plus oxaliplatin-based chemotherapy as adjuvant treatment for colon cancer (AVANT): a phase 3 randomised controlled trial. Lancet Oncol. *13*, 1225–1233.
3. Dudley, J.C., Lin, M.-T., Le, D.T., and Eshleman, J.R. (2016). Microsatellite instability as a biomarker for PD-1 blockade. Clin. Cancer Res. *22*, 813–820.
4. Bao, X., Shi, R., Zhao, T., Wang, Y., Anastasov, N., Rosemann, M., and Fang, W. (2021). Integrated analysis of single-cell RNA-seq and bulk RNA-seq unravels tumour heterogeneity plus M2-like tumour-associated macrophage infiltration and aggressiveness in TNBC. Cancer Immunol. Immunother. *70*, 189–202.
5. Liu, L., Zhang, R., Deng, J., Dai, X., Zhu, X., Fu, Q., Zhang, H., Tong, Z., Zhao, P., and Fang, W. (2021). Construction of TME and Identification of crosstalk between malignant cells and macrophages by SPP1 in hepatocellular carcinoma. Cancer Immunol. Immunother. *71*, 121–136.

6. Bao, X., Zhang, H., Wu, W., Cheng, S., Dai, X., Zhu, X., Fu, Q., Tong, Z., Liu, L., Zheng, Y., et al. (2020). Analysis of the molecular nature associated with microsatellite status in colon cancer identifies clinical implications for immunotherapy. J. Immunother. Cancer *8*, e001437.
7. DeBerardinis, R.J. (2020). Tumor microenvironment, metabolism, and immunotherapy. N. Engl. J. Med. *382*, 869–871.
8. Frankel, T., Lanfranca, M.P., and Zou, W. (2017). The role of tumor microenvironment in cancer immunotherapy. Adv. Exp. Med. Biol. *1036*, 51–64.
9. Chen, D., Bao, X., Zhang, R., Ding, Y., Zhang, M., Li, B., Zhang, H., Li, X., Tong, Z., Liu, L., et al. (2021). Depiction of the genomic and genetic landscape identifies CCL5 as a protective factor in colorectal neuroendocrine carcinoma. Br. J. Cancer *125*, 994–1002.
10. Bao, X., Li, Q., Chen, J., Chen, D., Ye, C., Dai, X., Wang, Y., Li, X., Rong, X., and Cheng, F. (2022). Molecular Subgroups of Intrahepatic Cholangiocarcinoma Discovered by Single-Cell RNA Sequencing-Assisted Multi-Omics Analysis. Cancer Immunol. Res. *10*, 811–828.
11. Bao, X., Wang, D., Dai, X., Liu, C., Zhang, H., Jin, Y., Tong, Z., Li, B., Tong, C., Xin, S., et al. (2023). An immunometabolism subtyping system identifies S100A9+ macrophage as an immune therapeutic target in colorectal cancer based on multiomics analysis. Cell Rep. Med. *4*, 100987.
12. Geeraerts, X., Bolli, E., Fendt, S.-M., and Van Ginderachter, J.A. (2017). Macrophage metabolism as therapeutic target for cancer, atherosclerosis, and obesity. Front. Immunol. *8*, 289.
13. Barkal, A.A., Brewer, R.E., Markovic, M., Kowarsky, M., Barkal, S.A., Zaro, B.W., Krishnan, V., Hatakeyama, J., Dorigo, O., Barkal, L.J., and Weissman, I.L. (2019). CD24 signalling through macrophage Siglec-10 is a target for cancer immunotherapy. Nature *572*, 392–396.
14. Bian, S., Hou, Y., Zhou, X., Li, X., Yong, J., Wang, Y., Wang, W., Yan, J., Hu, B., Guo, H., et al. (2018). Single-cell multiomics sequencing and analyses of human colorectal cancer. Science *362*, 1060–1063.
15. Roerink, S.F., Sasaki, N., Lee-Six, H., Young, M.D., Alexandrov, L.B., Behjati, S., Mitchell, T.J., Grossmann, S., Lightfoot, H., Egan, D.A., et al. (2018). Intra-tumour diversification in colorectal cancer at the single-cell level. Nature *556*, 457–462.
16. Zhang, L., Yu, X., Zheng, L., Zhang, Y., Li, Y., Fang, Q., Gao, R., Kang, B., Zhang, Q., Huang, J.Y., et al. (2018). Lineage tracking reveals dynamic relationships of T cells in colorectal cancer. Nature *564*, 268–272.
17. Lee, H.-O., Hong, Y., Etlioglu, H.E., Cho, Y.B., Pomella, V., Van den Bosch, B., Vanhecke, J., Verbandt, S., Hong, H., Min, J.-W., et al. (2020). Lineage-dependent gene expression programs influence the immune landscape of colorectal cancer. Nat. Genet. *52*, 594–603.
18. Pelka, K., Hofree, M., Chen, J.H., Sarkizova, S., Pirl, J.D., Jorgji, V., Bejnood, A., Dionne, D., Ge, W.H., Xu, K.H., et al. (2021). Spatially organized multicellular immune hubs in human colorectal cancer. Cell *184*, 4734–4752.e20.
19. Xue, D., Tabib, T., Morse, C., Yang, Y., Domsic, R., Khanna, D., and Lafyatis, R. (2021). Expansion of FCGR3A+ macrophages, FCN1+ mo-DC, and plasmacytoid dendritic cells associated with severe skin disease in systemic sclerosis. Arthritis Rheumatol. *74*, 329.
20. Li, J., Wang, Y., Wang, X., Ye, L., Zhou, Y., Persidsky, Y., and Ho, W. (2013). Immune activation of human brain microvascular endothelial cells inhibits HIV replication in macrophages. Blood, The Journal of the American Society of Hematology *121*, 2934–2942.
21. Chou, C.-H., Jain, V., Gibson, J., Attarian, D.E., Haraden, C.A., Yohn, C.B., Laberge, R.-M., Gregory, S., and Kraus, V.B. (2020). Synovial cell crosstalk with cartilage plays a major role in the pathogenesis of osteoarthritis. Sci. Rep. *10*, 10868.
22. Wauters, E., Van Mol, P., Garg, A.D., Jansen, S., Van Herck, Y., Vanderbeke, L., Bassez, A., Boeckx, B., Malengier-Devlies, B., Timmerman, A., et al. (2021). Discriminating mild from critical COVID-19 by innate and adaptive immune single-cell profiling of bronchoalveolar lavages. Cell Res. *31*, 272–290.
23. Liao, M., Liu, Y., Yuan, J., Wen, Y., Xu, G., Zhao, J., Cheng, L., Li, J., Wang, X., Wang, F., et al. (2020). Single-cell landscape of bronchoalveolar immune cells in patients with COVID-19. Nat. Med. *26*, 842–844.
24. Dick, S.A., Wong, A., Hamidzada, H., Nejat, S., Nechanitzky, R., Vohra, S., Mueller, B., Zaman, R., Kantores, C., Aronoff, L., et al. (2022). Three tissue resident macrophage subsets coexist across organs with conserved origins and life cycles. Sci. Immunol. *7*, eabf7777.
25. Sharma, A., Seow, J.J.W., Dutertre, C.-A., Pai, R., Blériot, C., Mishra, A., Wong, R.M.M., Singh, G.S.N., Sudhagar, S., Khalilnezhad, S., et al. (2020). Onco-fetal reprogramming of endothelial cells drives immunosuppressive macrophages in hepatocellular carcinoma. Cell *183*, 377–394.e21.
26. Cresswell, G.M., Wang, B., Kischuk, E.M., Broman, M.M., Alfar, R.A., Vickman, R.E., Dimitrov, D.S., Kularatne, S.A., Sundaram, C.P., Singhal, S., et al. (2021). Folate Receptor Beta Designates Immunosuppressive Tumor-Associated Myeloid Cells That Can Be Reprogrammed with Folate-Targeted Drugs. Cancer Res. *81*, 671–684.
27. Puig-Kröger, A., Sierra-Filardi, E., Domínguez-Soto, A., Samaniego, R., Corcuera, M.T., Gómez-Aguado, F., Ratnam, M., Sánchez-Mateos, P., and Corbí, A.L. (2009). Folate receptor β is expressed by tumor-associated macrophages and constitutes a marker for M2 anti-inflammatory/regulatory macrophages. Cancer Res. *69*, 9395–9403.
28. Zou, Y., Ye, F., Kong, Y., Hu, X., Deng, X., Xie, J., Song, C., Ou, X., Wu, S., Wu, L., et al. (2023). The Single-Cell Landscape of Intratumoral Heterogeneity and The Immunosuppressive Microenvironment in Liver and Brain Metastases of Breast Cancer. Adv. Sci. *10*, 2203699.
29. Ruffell, D., Mourkioti, F., Gambardella, A., Kirstetter, P., Lopez, R.G., Rosenthal, N., and Nerlov, C. (2009). A CREB-C/EBPβ cascade induces M2 macrophage-specific gene expression and promotes muscle injury repair. Proc. Natl. Acad. Sci. USA *106*, 17475–17480.
30. Cosín-Roger, J., Ortiz-Masiá, D., Calatayud, S., Hernández, C., Alvarez, A., Hinojosa, J., Esplugues, J.V., and Barrachina, M.D. (2013). M2 macrophages activate WNT signaling pathway in epithelial cells: relevance in ulcerative colitis. PLoS One *8*, e78128.
31. Cabello-Aguilar, S., Alame, M., Kon-Sun-Tack, F., Fau, C., Lacroix, M., and Colinge, J. (2020). SingleCellSignalR: inference of intercellular networks from single-cell transcriptomics. Nucleic Acids Res. *48*, e55.
32. Hugo, W., Zaretsky, J.M., Sun, L., Song, C., Moreno, B.H., Hu-Lieskovan, S., Berent-Maoz, B., Pang, J., Chmielowski, B., Cherry, G., et al. (2016). Genomic and transcriptomic features of response to anti-PD-1 therapy in metastatic melanoma. Cell *165*, 35–44.
33. Azarkhalili, B., Saberi, A., Chitsaz, H., and Sharifi-Zarchi, A. (2019). DeePathology: deep multi-task learning for inferring molecular pathology from cancer transcriptome. Sci. Rep. *9*, 16526–16614.
34. Cieślik, M., and Chinnaiyan, A.M. (2018). Cancer transcriptome profiling at the juncture of clinical translation. Nat. Rev. Genet. *19*, 93–109.
35. Way, G.P., Sanchez-Vega, F., La, K., Armenia, J., Chatila, W.K., Luna, A., Sander, C., Cherniack, A.D., Mina, M., Ciriello, G., et al. (2018). Machine learning detects pan-cancer ras pathway activation in the cancer genome atlas. Cell Rep. *23*, 172–180.e3.
36. Zhang, L., Li, Z., Skrzypczynska, K.M., Fang, Q., Zhang, W., O'Brien, S.A., He, Y., Wang, L., Zhang, Q., Kim, A., et al. (2020). Single-cell analyses inform mechanisms of myeloid-targeted therapies in colon cancer. Cell *181*, 442–459.e29.
37. Suthen, S., Lim, C.J., Nguyen, P.H.D., Dutertre, C.A., Lai, H.L.H., Wasser, M., Chua, C., Lim, T.K.H., Leow, W.Q., Loh, T.J., et al. (2022). Hypoxia-driven immunosuppression by Treg and type-2 conventional dendritic cells in HCC. Hepatology *76*, 1329–1344.

38. Liu, Y., Zhang, Q., Xing, B., Luo, N., Gao, R., Yu, K., Hu, X., Bu, Z., Peng, J., Ren, X., and Zhang, Z. (2022). Immune phenotypic linkage between colorectal cancer and liver metastasis. Cancer Cell *40*, 424–437.e5.

39. Wang, Y., Smith, W., Hao, D., He, B., and Kong, L. (2019). M1 and M2 macrophage polarization and potentially therapeutic naturally occurring compounds. Int. Immunopharmacol. *70*, 459–466.

40. Yunna, C., Mengru, H., Lei, W., and Weidong, C. (2020). Macrophage M1/M2 polarization. Eur. J. Pharmacol. *877*, 173090.

41. Martínez, V.G., Rubio, C., Martínez-Fernández, M., Segovia, C., López-Calderón, F., Garín, M.I., Teijeira, A., Munera-Maravilla, E., Varas, A., Sacedón, R., et al. (2017). BMP4 induces M2 macrophage polarization and favors tumor progression in bladder cancer. Clin. Cancer Res. *23*, 7388–7399.

42. Aljabery, F., Olsson, H., Gimm, O., Jahnson, S., and Shabo, I. (2018). M2-macrophage infiltration and macrophage traits of tumor cells in urinary bladder cancer. In Urologic Oncology: Seminars and Original Investigations, *4* (Elsevier), pp. 159.e119–159.e126.

43. Suárez-Sánchez, F.J., Lequerica-Fernández, P., Suárez-Canto, J., Rodrigo, J.P., Rodriguez-Santamarta, T., Domínguez-Iglesias, F., García-Pedrero, J.M., and de Vicente, J.C. (2020). Macrophages in oral carcinomas: Relationship with cancer stem cell markers and PD-L1 expression. Cancers *12*, 1764.

44. Li, X., Rawat, K., and Jakubzick, C.V. (2021). Targeting resident macrophages in cancer. Nat. Immunol. *22*, 1078–1079.

45. Viola, M.F., and Boeckxstaens, G. (2020). Intestinal resident macrophages: Multitaskers of the gut. Neuro Gastroenterol. Motil. *32*, e13843.

46. Weber, B., Saurer, L., Schenk, M., Dickgreber, N., and Mueller, C. (2011). CX3CR1 defines functionally distinct intestinal mononuclear phagocyte subsets which maintain their respective functions during homeostatic and inflammatory conditions. Eur. J. Immunol. *41*, 773–779.

47. Bain, C.C., Oliphant, C.J., Thomson, C.A., Kullberg, M.C., and Mowat, A.M. (2018). Proinflammatory Role of Monocyte-Derived CX3CR1(int) Macrophages in Helicobacter hepaticus-Induced Colitis. Infect. Immun. *86*, e00579-17.

48. Qualls, J.E., Kaplan, A.M., van Rooijen, N., and Cohen, D.A. (2006). Suppression of experimental colitis by intestinal mononuclear phagocytes. J. Leukoc. Biol. *80*, 802–815.

49. Zhu, Y., Herndon, J.M., Sojka, D.K., Kim, K.-W., Knolhoff, B.L., Zuo, C., Cullinan, D.R., Luo, J., Bearden, A.R., Lavine, K.J., et al. (2017). Tissue-resident macrophages in pancreatic ductal adenocarcinoma originate from embryonic hematopoiesis and promote tumor progression. Immunity *47*. 597–338.e326.

50. Nalio Ramos, R., Missolo-Koussou, Y., Gerber-Ferder, Y., Bromley, C.P., Bugatti, M., Núñez, N.G., Tosello Boari, J., Richer, W., Menger, L., Denizeau, J., et al. (2022). Tissue-resident FOLR2(+) macrophages associate with CD8(+) T cell infiltration in human breast cancer. Cell *185*, 1189–1207.e25.

51. Schürch, C.M., Bhate, S.S., Barlow, G.L., Phillips, D.J., Noti, L., Zlobec, I., Chu, P., Black, S., Demeter, J., McIlwain, D.R., et al. (2020). Coordinated cellular neighborhoods orchestrate antitumoral immunity at the colorectal cancer invasive front. Cell *182*, 1341–1359.e19.

52. Leader, A.M., Grout, J.A., Maier, B.B., Nabet, B.Y., Park, M.D., Tabachnikova, A., Chang, C., Walker, L., Lansky, A., Le Berichel, J., et al. (2021). Single-cell analysis of human non-small cell lung cancer lesions refines tumor classification and patient stratification. Cancer Cell *39*, 1594–1609.e12.

53. D'Amore, B., Smolinski-Zhao, S., Daye, D., and Uppot, R.N. (2021). Role of Machine Learning and Artificial Intelligence in Interventional Oncology. Curr. Oncol. Rep. *23*, 70–78.

54. Kann, B.H., Hosny, A., and Aerts, H.J. (2021). Artificial intelligence for clinical oncology. Cancer Cell *39*, 916–927.

55. Elkhader, J., and Elemento, O. (2021). Artificial Intelligence in Oncology: From Bench to Clinic (Elsevier).

56. Tong, Z., Liu, Y., Ma, H., Zhang, J., Lin, B., Bao, X., Xu, X., Gu, C., Zheng, Y., Liu, L., et al. (2020). Development, validation and comparison of artificial neural network models and logistic regression models predicting survival of unresectable pancreatic cancer. Front. Bioeng. Biotechnol. *8*, 196.

57. Cannarile, M.A., Weisser, M., Jacob, W., Jegg, A.-M., Ries, C.H., and Rüttinger, D. (2017). Colony-stimulating factor 1 receptor (CSF1R) inhibitors in cancer therapy. J. Immunother. Cancer *5*, 53.

58. Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., 3rd, Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive Integration of Single-Cell Data. Cell *177*, 1888–1902.e21.

59. Leek, J.T., Johnson, W.E., Parker, H.S., Jaffe, A.E., and Storey, J.D. (2012). The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics *28*, 882–883.

60. Wilkerson, M.D., and Hayes, D.N. (2010). ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. Bioinformatics *26*, 1572–1573.

61. Eide, P.W., Bruun, J., Lothe, R.A., and Sveen, A. (2017). CMScaller: an R package for consensus molecular subtyping of colorectal cancer pre-clinical models. Sci. Rep. *7*, 16618.

62. Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N.J., Livak, K.J., Mikkelsen, T.S., and Rinn, J.L. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat. Biotechnol. *32*, 381–386.

63. Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinf. *9*, 559.

64. Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., and Zhan, L. (2021). clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. Innovation *2*, 100141.

65. Hänzelmann, S., Castelo, R., and Guinney, J. (2013). GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinf. *14*, 7.

66. Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., III, Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., and Zager, M. (2021). Integrated analysis of multimodal single-cell data. Cell *184*, 3573–3587.

67. Qiu, X., Hill, A., Packer, J., Lin, D., Ma, Y.-A., and Trapnell, C. (2017). Single-cell mRNA quantification and differential analysis with Census. Nat. Methods *14*, 309–315.

68. Garcia-Alonso, L., Holland, C.H., Ibrahim, M.M., Turei, D., and Saez-Rodriguez, J. (2019). Benchmark and integration of resources for the estimation of human transcription factor activities. Genome Res. *29*, 1363–1375.

69. Wickham, H. (2011). ggplot2. Wiley Interdisciplinary Reviews: Comput. Stat. *3*, 180–185.

70. Taube, J.M., Akturk, G., Angelo, M., Engle, E.L., Gnjatic, S., Greenbaum, S., Greenwald, N.F., Hedvat, C.V., Hollmann, T.J., Juco, J., et al. (2020). The Society for Immunotherapy of Cancer statement on best practices for multiplex immunohistochemistry (IHC) and immunofluorescence (IF) staining and validation. J. Immunother. Cancer *8*, e000155.

71. Zhu, Y., Weiss, T., Zhang, Q., Sun, R., Wang, B., Yi, X., Wu, Z., Gao, H., Cai, X., Ruan, G., et al. (2019). High-throughput proteomic analysis of FFPE tissue samples facilitates tumor stratification. Mol. Oncol. *13*, 2305–2328.

72. Gao, H., Zhang, F., Liang, S., Zhang, Q., Lyu, M., Qian, L., Liu, W., Ge, W., Chen, C., Yi, X., et al. (2020). Accelerated Lysis and Proteolytic Digestion of Biopsy-Level Fresh-Frozen and FFPE Tissue Samples Using Pressure Cycling Technology. J. Proteome Res. *19*, 1982–1990.

73. Li, J., Van Vranken, J.G., Pontano Vaites, L., Schweppe, D.K., Huttlin, E.L., Etienne, C., Nandhikonda, P., Viner, R., Robitaille, A.M., Thompson, A.H., et al. (2020). TMTpro reagents: a set of isobaric labeling mass tags enables

simultaneous proteome-wide measurements across 16 samples. Nat. Methods *17*, 399–404.

74. Shen, B., Yi, X., Sun, Y., Bi, X., Du, J., Zhang, C., Quan, S., Zhang, F., Sun, R., Qian, L., et al. (2020). Proteomic and Metabolomic Characterization of COVID-19 Patient Sera. Cell *182*, 59–72.e15.

75. Nie, X., Qian, L., Sun, R., Huang, B., Dong, X., Xiao, Q., Zhang, Q., Lu, T., Yue, L., Chen, S., et al. (2021). Multi-organ proteomic landscape of COVID-19 autopsies. Cell *184*, 775–791.e14.

76. Braun, D.A., Street, K., Burke, K.P., Cookmeyer, D.L., Denize, T., Pedersen, C.B., Gohil, S.H., Schindler, N., Pomerance, L., Hirsch, L., et al. (2021). Progressive immune dysfunction with advancing disease stage in renal cell carcinoma. Cancer Cell *39*, 632–648.e8.

77. Bao, X., Shi, R., Zhao, T., and Wang, Y. (2020). Mast cell-based molecular subtypes and signature associated with clinical outcome in early-stage lung adenocarcinoma. Mol. Oncol. *14*, 917–932.

78. Engebretsen, S., and Bohlin, J. (2019). Statistical predictions with glmnet. Clin. Epigenetics *11*, 123.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit monoclonal Anti-F4/80	Abcam	Cat#ab300421; RRID:AB_2936298
Rabbit monoclonal Anti-Cathepsin D	Abcam	Cat#ab75852; RRID:AB_1523267
CD68 (D4B9C) XP® Rabbit mAb	Cell Signaling Technology	Cat#76437; RRID:AB_2799882
Rabbit monoclonal Anti-CD4	Abcam	Cat#ab183685; RRID:AB_2686917
*In*VivoMAb anti-mouse CSF1R (CD115)	Bioxcell	Cat#BE0213; RRID:AB_2687699
CD4 (EP204) Rabbit mAb	Cell Signaling Technology	Cat#48274; RRID:AB_3076699
CD8α (D8A8Y) Rabbit mAb	Cell Signaling Technology	Cat#85336s; RRID:AB_2800052
Rabbit monoclonal Anti-CD8a	Abcam	Cat#ab217344; RRID:AB_2890649)
Mouse monoclonal Anti-FOLR2	Abcam	Cat#ab103988; RRID:AB_10711133
Rabbit polyclonal Anti-FOLR2	Abcam	Cat#ab228643
Rabbit monoclonal Anti-PTGS2	Abcam	Cat#ab179800; RRID:AB_2894871
Rabbit monoclonal Anti-PD-1	Abcam	Cat# ab214421; RRID:AB_294180
Mouse monoclonal Anti-PD-1	Abcam	Cat#ab52587; RRID:AB_881954
Rabbit polyclonal Anti-DKK3	Invitrogen	Cat#PA5-102626; RRID:AB_2852023
Rabbit monoclonal Anti-FOXP3	Invitrogen	Cat#700914; RRID:AB_2532349
BD Pharmingen^{TM} Purified Mouse Anti-Ki-67	BD Biosciences	Cat#550609; RRID:AB_393778
Chemicals, peptides, and recombinant proteins		
Dextran sulfate (sodium salt)	Cayman Chemical	Cat#23250
RPMI 1640	Gibco	Cat#11875–093
Fluorouracil	MCE	Cat#HY-90006
Protease inhibitor	Solarbio	Cat#P6730
Dispase II	Sigma−Aldrich	Cat#42613-33-2
Type VIII Collagenase	Sigma−Aldrich	Cat#C2139
DNase I	NEB	Cat#M0303S
Fetal bovine serum	Gibco	Cat#16000–044
PBS	Solarbio	Cat#P1020
Nylon cell strainer	Falcon	Cat#352340
Red blood cell lysis buffer	Invitrogen	Cat#00-4333-57
Bovine serum albumin	Sigma−Aldrich	Cat#B2064
Dead Cell Removal Kit	Miltenyi Biotec	Cat#130-090-101
TBST	Solarbio	Cat#T1082
Opal Polymer HRP Ms+Rb	AKOYA Biosciences	Cat#NEL820001KT
Blocking solution	Proteintech	Cat#B900780
Diphtheria toxin	MCE	Cat#HY-108851
4% Paraformaldehyde	Solarbio	Cat#P1110
Modified Hematoxylin-Eosin (HE) Stain Kit	Solarbio	Cat#G1121
0.1% ammonia water	Solarbio	Cat#G1823
Parafin	Solarbio	Cat#YA0011
10% neutral buffered formalin (NBF)	Solarbio	Cat#G2161
Xylene	Sinopharm Chemical ReagentCo., Ltd	Cat#10023418
Alcohol	Sinopharm Chemical ReagentCo., Ltd	Cat#10009218
Epitope retrieval agent	Solarbio	Cat#C1032

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Anti-mouse/rabbit universal immunohistochemical detection kit	Proteintech	Cat#pk10006
Biological samples		
B6/JGpt-Apc$^{em1Cin(Min)/}$Gpt mice	Gempharmatech Co., Ltd	N/A
Critical commercial assays		
Single-cell sequencing	OE Biotech Co., Ltd	N/A
Deposited data		
Single-cell sequencing data	This paper	HRA003569
TCGA-COAD cohort RNA sequencing data	TCGA database	N/A
TCGA-READ cohort RNA sequencing data	TCGA database	N/A
GEO cohorts	GEO	GSE56699, GSE14333, GSE39582, GSE17536, GSE17537, GSE33113, GSE37892, GSE146771, GSE72970, GSE19860
Software and algorithms		
R software (version 4.0.4)	R Core	https://www.r-project.org/
Seurat (version 3.1.1)[58]	Stuart T. et al., 2018[58]	https://satijalab.org/seurat/
Cell Ranger software (version 3.1.0)	10× genomics	https://support.10xgenomics.com/single-cell-gene-expression/software/overview/welcome
sva R package (version 3.44.0)[59]	Leek, J.T. et al.[59]	https://bioconductor.org/packages/release/bioc/html/sva.html
ConsensusClusterPlus (version 3.16)[60]	Matt, W., Peter, W.[60]	https://git.bioconductor.org/packages/ConsensusClusterPlus
CMScaller R package[61]	Peter, W.E. et al.[61]	https://github.com/peterawe/CMScaller
SingleCellSignalR R package[31]	Simon, C.A. et al.[31]	https://git.bioconductor.org/packages/SingleCellSignalR
Monocle2 package (v2.8.0)[62]	Trapnell, C. et al.[62]	
WGCNA R package[63]	Peter, L. et al.[63]	https://github.com/cran/WGCNA
ClusterProfiler R package (version 4.5.2)[64]	Yu, G. et al.[64]	https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html
GSVA R package (version 1.45.5)[65]	Hänzelmann, S. et al.[65]	https://bioconductor.org/packages/release/bioc/html/GSVA.html
Image-Pro Plus software (version 6.0)	Media Cybernetics	https://www.mediacy.com/78-products/image-pro-plus
ImageJ software (version 4.0)	NIH	https://imagej.nih.gov/ij/
Countess	Thermo	N/A

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Peng Zhao (zhaop@zju.edu.cn).

Materials availability

This study did not generate new unique reagents.

Data and code availability

The RNA sequencing data were deposited at Gene Expression Omnibus (GEO) (GSE56699, GSE14333, GSE39582, GSE17536, GSE17537, GSE33113, GSE37892, GSE146771, GSE72970, and GSE19860) and The Cancer Genome Atlas (TCGA), which are publicly available. All raw data generated by this study have been deposited in the Chinese national genomics data center (https://ngdc.cncb.ac.cn), under accession number HRA003569. The software and algorithms for data analyses used in this study are published and referenced throughout the STAR Methods section. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Animal objects

B6/JGpt-*Apc*$^{em1Cin(Min)}$/Gpt mice were purchased from the Gempharmatech Co., Ltd (China), and FOLR2-DTR mice were a kind gift from Prof. Jianpeng Sheng. All mice were housed in the SPF facility of the First Affiliated Hospital, Zhejiang University School of Medicine, with approval from the Institutional Animal Care and Use Committee (IACUC2022-71) and Zhejiang Center of Laboratory Animals, with approval from the Institutional Animal Care and Use Committee (ZJCLA-IACUC-20020173), respectively. The FOLR2-DTR-*Apc*^{Min} mice was bred by crossing the FOLR2-DTR-eGFP mice with the B6/JGpt-*Apc*$^{em1Cin(Min)}$/Gpt mice. To establish spontaneous CRC, 3% Dextran sulfate (sodium salt) (DSS) (Cayman Chemical, #23250, USA) was given in the drinking water for 1 week followed by ordinary drinking water. Mice were treated with diphtheria toxin (MCE, #HY-108851, USA) or 50 mg/kg 5-Fluorouracil (5FU) (MCE, #HY-90006, USA) through intraperitoneal injection for 2 weeks before sacrificed, and the injections were performed every 3 days. On day 42, all mice were sacrificed, and tumors were collected for H&E, IHC, and mIHC staining.

Human subjects

The bulk transcriptome data and clinical information were obtained from the Gene Expression Omnibus (GEO) repository (https://www.ncbi.nlm.nih.gov/geo/) and The Cancer Genome Atlas (TCGA) (colon cancer and rectal cancer, https://xenabrowser.net/datapages/). GSE39582, GSE17536, GSE17537, GSE14333, GSE56699, GSE37892, and GSE33113 expression profile data were integrated as an integrated cohort after removing batch effects using the sva R package.[59] A total of 2081 patients were included in the survival analysis. 1405 patients were used as the integrative training cohorts and 682 patients were included in the four independent validation cohorts. The detailed clinical information is provided in Table S5. The transcriptome and the chemotherapy response data from GSE72970 were used to build the FOLR2$^+$ resident-phenotype macrophage (RM) model, and data from GSE19860 were used to test the efficiency.

The scRNA-seq data for our in-house cohort and the GSE146771 cohort from the GEO repository were integrated. This study was approved by the Ethics Committee of the First Affiliated Hospital, Zhejiang University School of Medicine and the Ethics Committee of the Affiliated Hospital of Southwest Medical University (IIT20220758A) and the Ethics Committee of the Affiliated Hospital of Southwest Medical University (KY2023312). Tumor tissues were obtained from biopsies of 8 patients diagnosed with CRC. All samples for scRNA-seq were verified by pathological examinations as adenocarcinoma. Among the eight samples, three samples came from right-side (ascending colon and transverse colon), three samples came from left-side (sigmoid colon and descending colon) and two samples came from rectum. Each sample was cut into 5 mm^3 size with a sterile scalpel to keep the consistency between each sample and followed by dissociation. The detailed clinical information for scRNA-seq patient is provided in Table S6. The inclusion criteria for both the training cohort and validation cohort were as follows: patients who received first-line therapy, with the first-line chemotherapy regimens being CAPOX (Capecitabine + Oxaliplatin) or FOLFOX (Leucovorin Calcium + Fluorouracil + Oxaliplatin). The age range of the patients was from 18 to 75 years, and they had an ECOG performance status ranging from 0 to 2, along with adequate renal, hepatic, and bone marrow function. The major exclusion criteria included active inflammation, prior or concurrent malignant disease within the last 5 years, previous use of study drugs, and the presence of poorly controlled hypertension, diabetes, serious cardiovascular disease, or other chronic diseases, among others. Tumor response was regularly assessed every 2 months, following the Response Evaluation Criteria in Solid Tumors (RECIST) version 1.1. A total of 226 patients were included, with 181 in the training cohort and 45 in the testing cohort. More detailed clinical information can be found in Tables S7 and S8. The tumor tissues contained both the stromal area and intraepithelial layer for the next step analysis.

METHOD DETAILS

scRNA-seq processing

RPMI 1640 (Gibco, #11875–093, USA) with 1 mM protease inhibitor (Solarbio, #P6730, China) was used to transport CRC tissues. Tissues were digested with a dissociation enzyme cocktail prepared by dissolving 2 mg/mL Dispase II (Sigma−Aldrich, #42613-33-2, USA), 1 mg/mL Type VIII Collagenase (Sigma−Aldrich, #C2139, USA), and 1 unit/mL DNase I (NEB, #M0303S, USA) in PBS with 5% fetal bovine serum (FBS; Gibco, #16000–044, USA) for 40 min at 37°C. The cells were dissociated and collected every 20 min and then filtered using a 40 μm nylon cell strainer (Falcon, #352340, USA). Red blood cell lysis buffer (Invitrogen, #00-4333-57, USA) with 1 unit/mL DNase I was used to remove red blood cells. Finally, the cells were washed in PBS (Solarbio, #P1020, China) with 0.04% bovine serum albumin (BSA; Sigma−Aldrich, #B2064, USA). The concentration of the single-cell suspension was computed with Countess (Thermo) and adjusted to 1000 cells/μL. After removing the dead cells by Dead Cell Removal Kit (Miltenyi Biotec, #130-090-101, Germany), the cells were loaded according to the Chromium single-cell 3′ kit standard protocol to capture 5,000–10,000 cells/chip position (V2 chemistry). Library construction and all the other processes were performed according to the standard manufacturer’s protocol.

Illumina HiSeq X Ten was used to obtain single-cell libraries using 150 nt paired-end sequencing. The Cell Ranger software pipeline (version 3.1.0) provided by 10×Genomics was used to demultiplex cellular barcodes, map reads to the genome and transcriptome using the STAR aligner, and down-sample reads as required to generate normalized aggregate data across samples, producing a matrix of gene counts versus cells. We processed the unique molecular identifier (UMI) count matrix using the R package Seurat (version 3.1.1). To remove low-quality cells and likely multiplet captures, which is a major concern in microdroplet-based experiments, a set of criteria were conducted: (1) removing cells with less than 500 UMIs; (2) removing high mitochondrial RNA UMIs (more than four times of the median number of mitochondrial UMIs across cells); (3) filtering genes expressed in less than five cells. Also, the doublets were excluded by the parameter: nFeature_RNA <6000 of subset function in Seurat package. Library size normalization was performed with NormalizeData function in Seurat to obtain the normalized count. Specifically, the global-scaling normalization method ‘‘LogNormalize’’ normalized the gene expression measurements for each cell by the total expression, multiplied by a scaling factor (10,000 by default), and the results were log transformed. Subclusters were identified by principal component analysis (PCA) and t-distributed stochastic neighbor embedding (t-SNE) using the Seurat R package and the Seurat FindVariableGenes function.[66] The featured gene signature was constructed with the Seurat R package. Single-cell trajectories in CRC were constructed using the Monocle2 package (v2.8.0) with the features of each populations as the input.[67] The transcription factor (TF) analysis was performed with dorothea package using the default parameters.[68]

Functional enrichment analysis

The GSVA R package was applied for single-sample GSEA (ssGSEA) for each sample.[65] The gene sets obtained from scRNA-seq and from Broad Institute (http://www.gsea-msigdb.org/gsea/msigdb/index.jsp, C7 represents cell states and perturbations within the immune system) were used for the enrichment analysis. The results were plotted with the ggplot2 R package and pheatmap R package.[69]

Cell−cell interaction (CCI) analysis using SingleCellSignalR

To systematically analyze CCIs, we adopted the SingleCellSignalR package to explore the ligand and target gene pairs.[31] Gene expression data of interacting cells were input into SingleCellSignalR and combined with a prior model that integrated existing knowledge on ligand-to-target signaling paths. Then, the ligand−receptor interactions that drive gene expression changes in the cells of interest were predicted.

Weighted gene co-expression network analysis (WGCNA)

A gene co-expression network was built by the WGCNA model.[63] Raising the co-expression similarity to a power β defined a weighted network adjacency. By evaluating the correlations between the FOLR2$^+$ macrophage abundance of patients from the integrated cohort and the module memberships, it was possible to identify highly correlated modules. The hub gene in the yellow module was selected and subjected to further analysis. Gene ontology (GO) analysis was performed by the clusterProfiler R package,[64] and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis was performed by Proteomaps, a bionic visualization method of all pathways (https://bionic-vis.biologie.uni-greifswald.de/).

Multiplex immunohistochemistry (mIHC) staining of tissue sections

Tumor tissues (4 μm) from CRC patients were fixed with 4% paraformaldehyde (Solarbio, #P1110, China), and then embedded in paraffin (Solarbio, #YA0011, China). The specificity and optimal dilution of each antibody were individually determined using slides from CRC tissues before being used in combination, following the SITC multiplex immunohistochemistry guidance.[70] To overcome the issue of signal-to-noise, a tyramide signal amplification (TSA) approach was applied. In summary, several key steps should be followed before performing mIHC. The first step is to select the appropriate monoclonal antibody that matches the target protein intended for labeling. Next, mIHC staining was performed. In brief, the prepared tissue sections (4 μm) were baked in an oven at 65°C for 1 h to improve sample adhesion to the slide. Then, they were dewaxed with fresh xylene twice for 15 min each and rehydrated with graded alcohol (100%, 95%, 85%, 75%) for 5 min each, respectively. After rehydration, the sections were fixed in 10% neutral buffered formalin (NBF) (Solarbio, #G2161, China) for 20 min at room temperature. Subsequently, the sections were transferred to an appropriate antigen retrieval (AR) buffer and placed in a microwave for 1 min at 100% power, followed by an additional 15 min at 20% power. Then, the slides were blocked after cooling to room temperature and subsequently incubated with a primary antibody at room temperature for 10 min. To remove any excess antibody, the slides were washed three times with TBST (Solarbio, #T1082, China). Following this, the slides were incubated with Opal Polymer HRP Ms+Rb (AKOYA Biosciences, #NEL820001KT, USA) at room temperature for 10 min. To remove any remaining wash buffer, the slides were rinsed three times with TBST before incubation with Opal Signal Generation. The process of microwave treatment, blocking, primary antibody incubation, introduction of Opal Polymer HRP, and signal amplification were repeated before each subsequent antibody incubation. Primary antibodies, including CD68 (1:100, #76437, Cell Signaling Technology), F4/80 (1:200, #ab300421, Abcam), FOLR2 (1:300, #ab103988, Abcam), FOLR2 (1:300, #ab228643, Abcam), PTGS2 (1:100, #ab179800, Abcam), CD4 (1:200, #48274, Cell Signaling Technology), CD4 (1:100, #ab183685, Abcam), CD8α (1:200, #85336s, Cell Signaling Technology), CD8a (1:100, #ab217344, Abcam), PD-1 (1:200, #ab214421, Abcam), PD-1 (1:100, #ab52587, Abcam), DKK3 (1:100, #PA5-102626, Invitrogen), FOXP3 (1:100, #700914, Invitrogen), CTSD (1:100, #ab75852, Abcam) were labeled. Next, the slides were incubated with DAPI working solution in the dark for 5 min at room temperature. Afterward, the slides were washed with distilled water and TBST before mounting. Finally, a confocal microscope (Nikon, Japan) or a Vectra Polaris Quantitative Pathology Imaging Systems was used to capture images of those tissue samples, and the acquired images were analyzed using ImageJ (version 4.0).

AI-guided chemotherapy model construction

The FOLR2$^+$ macrophage gene signature was applied to build the neural network model for predicting the chemotherapy response for CRC. Briefly, the transcriptome data and the chemotherapy response data from GSE72970 were used as the input and output for training the FOLR2-RM model. GSE19860 was applied as the testing cohort. And the traditional neural network was applied with 25 layers for construction.

The mIHC staining slide and paired HE staining slide were prepared for model construction. HE staining was scanned by 3D Histech and used for classifying tumor areas. The classifiers tab function of HALO digital pathology system was utilized to identify tumor areas. Subsequently, the tumor areas in the mIHC slides were aligned with the tumor areas in the corresponding HE staining slides. Only tumor areas were considered for the subsequent analysis in the CCIM pipeline (detailed pipeline shown in Figure S8). In the image preprocessing stage, in order to preserve the original information of the image as much as possible, only local pixel resampling and color normalization are adopted. Since the sensitivity of color features in the task of determining chemotherapeutic response from multi-color fluorescence images, techniques for color enhancement other than color normalization are not used during preprocessing or even during training. For CCIM-net, 8 models (ResNet18, ResNet34, ResNet50, ResNet101, InceptionV3, Xception, DenseNet121, ResNext50) are pre-trained on ImageNet, which makes the model have better initialization parameters for deep feature extraction. To help the model more effectively adapt to the initial learning rate, we use one round of warm-up prior to the training phase. At 30, 50, and 80 epochs, the learning rate then declines using a multi-step method to 0.8 times its initial value. The learning rate and batch size hyperparameters for these models were set at 0.001 and 4, respectively. After training each model for 100 epochs, save the model parameters for the best-performing epoch. The ResNet several models introduced the conception of residual network. For a block, let the function that can be fitted be $\mathrm{F(x)}$ and its expected latent mapping be $\mathrm{H(x)}$. ResNet network model learns the residual $\mathrm{H(x)} - \mathrm{x}$, namely $\mathrm{F(x)} := \mathrm{H(x)} - \mathrm{x}$, so that the rise path becomes $\mathrm{F(x)+x}$, and $\mathrm{F(x)+x}$ is used to fit H(x). In ResNet, the block formed by $\mathrm{F(x)+x}$ is called Residual block. Multiple similar Residual blocks in series formed the ResNet. The ResNet18, ResNet34, ResNet50 and ResNet101 are the different versions of ResNet. The main difference of each version came from the composition of residual block and the number of convolution layers. The ResNet18 and ResNet34 were constructed with BasicBlock while ResNet50 and ResNet101 were constructed with Bottleneck. The ResNext model can improve the accuracy in the case of not improving the parameter complexity and reducing the number of super parameters. For instance, the algorithm of traditional fully connected layer is:

$$\sum_{i=1}^{D} \omega_i x_i$$

where $x = x_1, x_2, x_3, \ldots, x_D$.

x_D is a D-channel input vector to the neuron and ω_i is a filter's weight for the i-th channel. The ResNext replaced the $\omega_i x_i$ to a more common function as:

$$F(x) = \sum_{i=1}^{C} T_i(x)$$

where $T_i(x)$ can be an arbitrary function. C is cardinality (the size of the set of transformations). T_i have the same topological pattern. The ResNext50 was selected for building the CCIM-net for the highest accuracy in the testing cohort.

These models all adopt the cross-entropy loss function, which is frequently used to measure the difference between two probability distributions and aims to describe how difficult it is to express a probability distribution p by a probability distribution q. Assuming that p and q are two probability distributions of the data x, the cross-entropy loss of N samples of p represented by q can be calculated as follows:

$$L = -\sum_{i=1}^{N} p(x)\log q(x) + (1 - p(x))(\log(1 - q(x)))$$

According to the formula, the closer the probability distributions p and q are, the smaller the cross-entropy loss. In the training process, by constructing the loss constraint between the prediction result and the ground truth and multiple rounds of optimization iterations, the cross-entropy loss will converge, and the model has the ability to characterize the chemotherapy response results corresponding to the multi-color fluorescence images. Finally, the probability distribution output by the model on the test samples is close to the probability distribution of the training samples, which makes the prediction results more accurate.

Haematoxylin and eosin (HE) staining

The tumor tissue sections (4μm) were deparaffinized using xylene (Sinopharm Chemical Reagent Co., Ltd, #10023418, China), followed by rehydration with graded alcohol (Sinopharm Chemical Reagent Co., Ltd, #10009218, China). To stain the nuclei, the sections were subjected to hematoxylin staining at room temperature for 30 min, using hematoxylin (Solarbio, #G1121, China). After staining, the sections were washed with PBS to remove any excess hematoxylin and prevent over-staining. Next, 0.1% ammonia water (Solarbio, #G1823, China) was applied to the sections to change the color of the hematoxylin-stained nuclei from reddish to a distinct blue-purple hue. To prepare the sections for cytoplasm staining, they were rinsed with 75% alcohol at room temperature for 2 min. Cytoplasm staining was performed using eosin (Solarbio, #G1121, China) at room temperature for 1 h. After eosin staining, the sections were directly rinsed with graded alcohol to remove excess dye and prepare them for further analysis. Finally, the sections were treated with xylene, which acted as an anhydrous alcohol, before being mounted on slides. The sections were examined using a light microscope (Leica), and the images were analyzed with Image-Pro Plus software (version 6.0).

Immunohistochemistry (IHC) staining

The sections were deparaffinized using xylene, a widely-used solvent known for its ability to effectively remove wax from tissue sections. Subsequently, the sections were rehydrated using graded alcohol to ensure optimal preparation for subsequent procedures. Microwave heating was employed to enhance the detection of specific proteins and retrieve antigen epitopes. A specific epitope retrieval agent (Solarbio, #C1032, China) was utilized to achieve this goal. By subjecting the sections to microwave heating alongside the epitope retrieval agent, the antigens were effectively unmasked, resulting in improved efficiency of antibody binding. To prevent non-specific binding, the sections were then treated with a blocking solution and allowed to incubate for 1 h at room temperature. Following this, the primary antibody specific to Ki67 (1:200, #550609, BD Biosciences) was introduced and allowed to incubate overnight at 4°C. The next day, the sections were thoroughly washed three times with PBS to eliminate any excess primary antibody and unbound molecules. For immunostaining, a versatile anti-mouse/rabbit universal immunohistochemical detection kit (Proteintech, #pk10006, China) was employed. The staining procedure was meticulously carried out according to the manufacturer's instructions, ensuring superior detection and visualization of the Ki67 protein. Subsequently, the mounted sections were meticulously examined using a Leica microscope specifically calibrated for light microscopy. Finally, the obtained images were subjected to detailed analysis using Image-Pro Plus software (version 6.0).

Tandem mass tag (TMT)-based proteomic analysis

The tissues were dewaxed, rehydrated and then acidic hydrolysis with formic acid (FA) was performed. Proteins were denatured with 6 M urea (Sigma–Aldrich, Germany) and 2 M thiourea (Sigma–Aldrich, Germany) before being digested into peptides with trypsin (1:20; Hualishi, Beijing, China) and Lys-C (1:80; Hualishi, Beijing, China) using pressure-cycling technology (PCT).[71,72] TMTpro 16 plex (Thermo Fisher Scientific, San Jose, USA) was used to label peptides.[73] In the TMT126 channel, each batch contained 15 experimental samples and one pooled sample for normalization. The fractions (60 per batch) were separated using offline high-pH reversed-phase fractionation with a Thermo Dionex Ultimate 3000 RSLC nano System and then merged to produce a total of 30 fractions per batch. The fractionated samples were subsequently separated using a Thermo Dionex Ultimate 3000 RSLC nano System before being examined with an HF mass spectrometer in data-dependent acquisition (DDA) mode (Thermo Fisher Scientific, San Jose, USA). Using Proteome Discoverer (version 2.4, Thermo Fisher Scientific, Waltham, MA), all reviewed human entries from UniProt (downloaded on 14 April 2020, containing 20,365 proteins) were searched. The detailed parameters were previously described without modification.[74,75]

Construction of the CCIM scoring system

A CCIM scoring system was constructed with modifications from previous studies.[76,77] A CCIM signature was built with the featured genes in FOLR2$^+$ macrophages, exhausted CD4$^+$ T cells, Treg cells, exhausted CD8$^+$ T cells, tolerance NK cells and the genes from CCI ligand−receptor pairs. The least absolute shrinkage and selection operator (LASSO) was applied to construct the CCIM model.[78] The CCIM scoring system was built by including individual normalized gene expression values weighted by their LASSO Cox coefficients as follows: $\sum_i Coefficient(mRNA_i) \times Normalized\ Expression(mRNA_i)$.

QUANTIFICATION AND STATISTICAL ANALYSIS

All experimental data were analyzed using R software (R version 4.0.4). Measurement data are expressed as the mean ± standard error. A p value <0.05 was considered to indicate significance. Univariate and multivariate Cox analysis was performed to determine whether a variable significantly affected disease-free survival (DFS).

Cell Reports Medicine, Volume *5*

Supplemental information

A multiomics analysis-assisted deep learning model identifies a macrophage-oriented module as a potential therapeutic target in colorectal cancer

Xuanwen Bao, Qiong Li, Dong Chen, Xiaomeng Dai, Chuan Liu, Weihong Tian, Hangyu Zhang, Yuzhi Jin, Yin Wang, Jinlin Cheng, Chunyu Lai, Chanqi Ye, Shan Xin, Xin Li, Ge Su, Yongfeng Ding, Yangyang Xiong, Jindong Xie, Vincent Tano, Yanfang Wang, Wenguang Fu, Shuiguang Deng, Weijia Fang, Jianpeng Sheng, Jian Ruan, and Peng Zhao

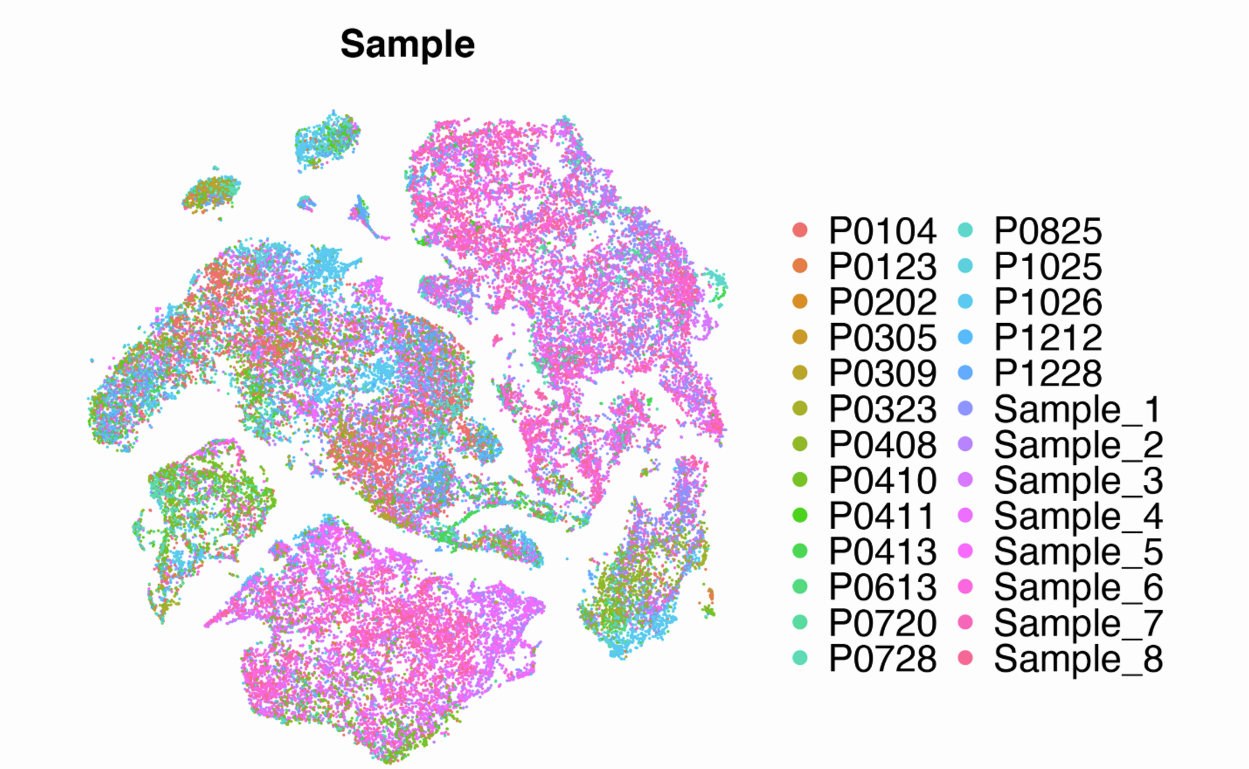


Fig. S1: The t-SNE plot coloured by CRC patients, related to Figure 1. A total of 24 patients were included for analysis. *t-SNE: t-distributed stochastic neighbour embedding; CRC: colorectal cancer.*

Figure S2

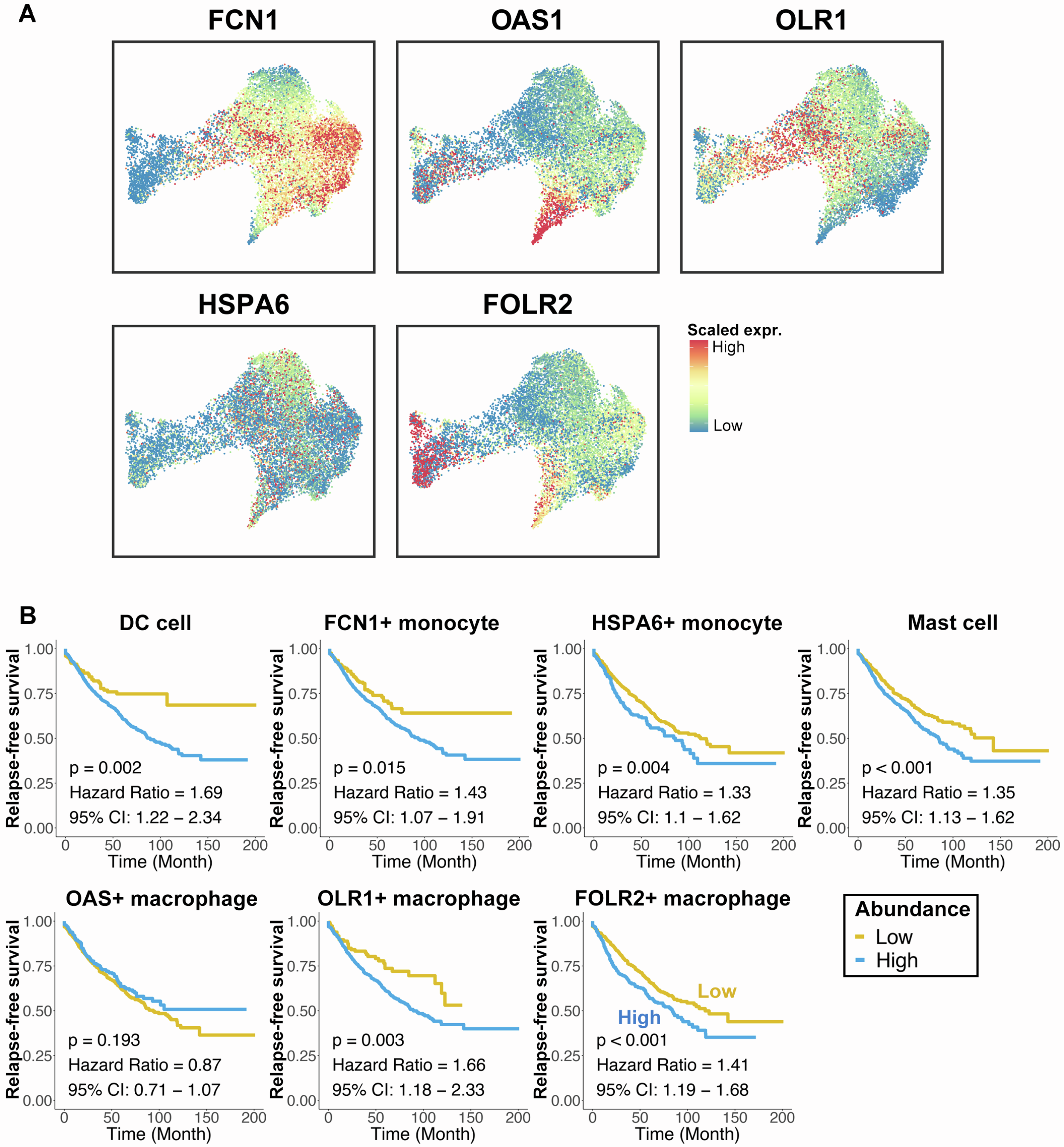


Fig. S2: The monocyte/macrophage subpopulations and their prognostic values, related to Figure 2. A.) The UMAP plot showing the expression level of the selected features in monocyte/macrophage subpopulations. B.) The association of DFS with the enrichment of each myeloid cell type. Gene signatures of myeloid cells were defined as the feature genes of each cell type and obtained from the scRNA-seq data (Pts n=1405). The cell abundance was calculated according to the gene signature for each cell type based on bulk transcriptome data using the GSVA method. *UMAP: Uniform Manifold Approximation and Projection; DFS: disease-free survival; scRNA-Seq: single-cell RNA sequencing; GSVA: gene set variation analysis, Pts: patients.*

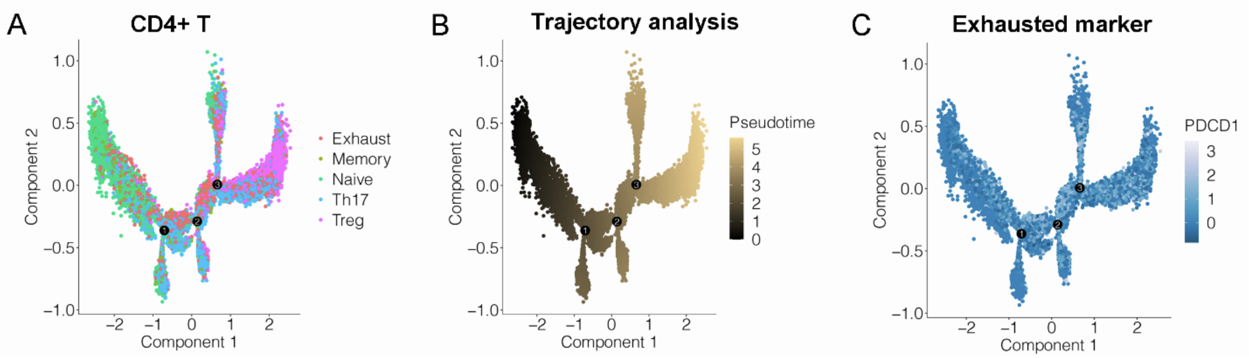


Fig. S3: The CD4+ T-cell subpopulations in CRC, related to Figure 4. A.) Trajectory inference for 5 subclusters of CD4$^+$ T cells. B.) Trajectory inference with the pseudotime score for 5 subclusters of CD4$^+$ T cells. The colour key from black to gold indicates the pseudotime score from low to high. C.) Trajectory inference with PD-1 expression for 5 subclusters of CD4$^+$ T cells. The colour key from blue to white indicates PD-1 expression from low to high. *CRC: colorectal cancer.*

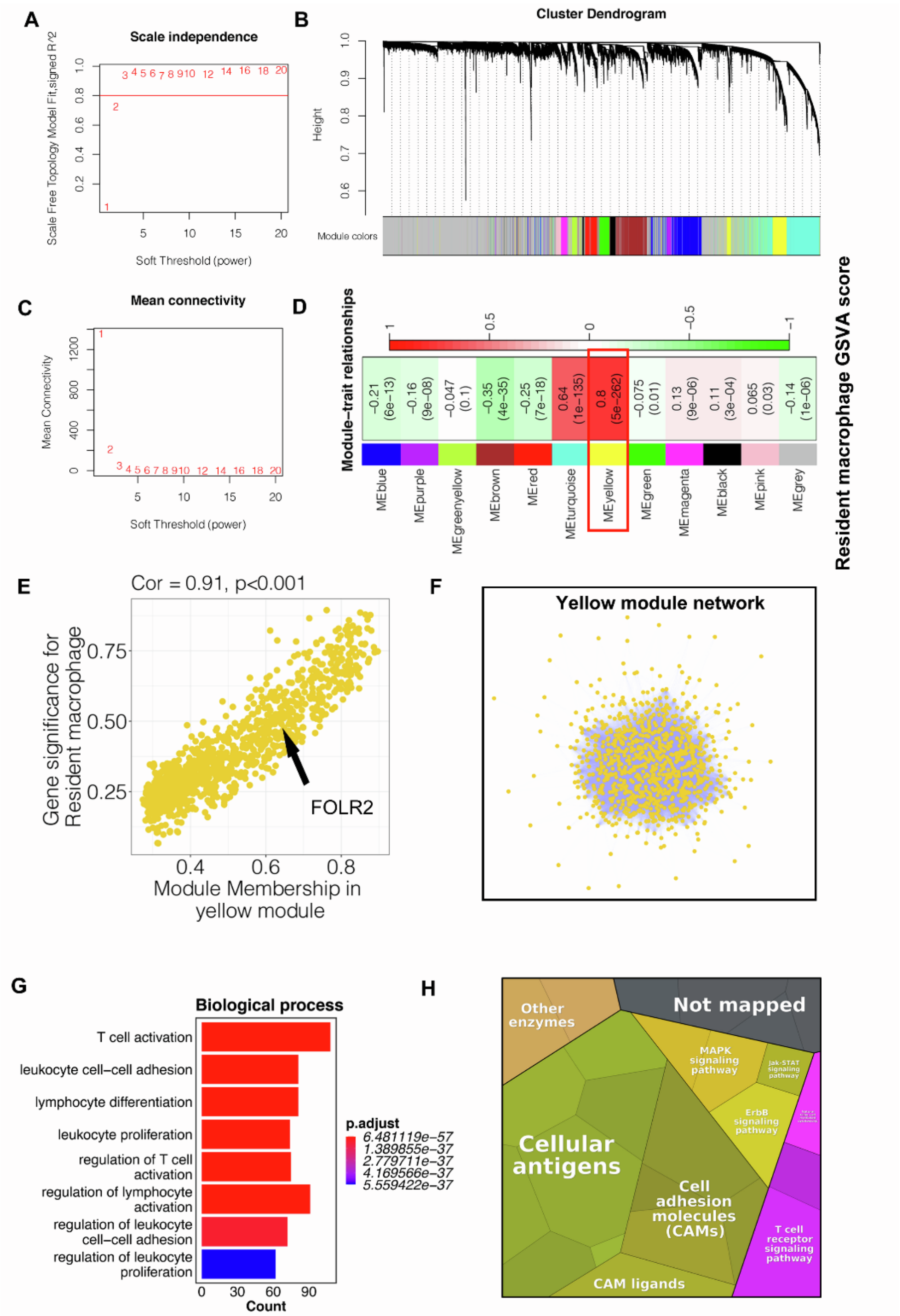


Fig. S4: Optimal soft threshold selection and sample clustering in WGCNA, related to Figure 5. A.) A power of β = 3 was chosen as the optimal soft threshold to ensure scale-free coexpression. B.) A power of β = 3 represents low mean connectivity in the WGCNA model. C.) The clustering of gene modules in WGCNA. D.) The association between identified modules and FOLR2$^+$ macrophage abundance calculated with the FOLR2$^+$ macrophage gene signature on the bulk transcriptome. The colour key from green to red indicates the relative correlation level from low to high. E.) The correlation between gene significance for FOLR2$^+$ macrophages and module membership in the yellow module. The black arrow indicates that FOLR2 is involved in the yellow module. Network clustering identified 11 non-grey (meaningful) gene modules. F.) The gene network in the yellow module. G.) Gene ontology analysis based on the genes within the yellow module (FOLR2$^+$ macrophage-related module) indicating the enriched biological process. H.) Proteomaps pathway analysis of genes within the yellow module (FOLR2$^+$ macrophage-related module). Each small polygon corresponds to a single KEGG pathway, and the size correlates with the number of genes involved in the pathway. *WGCNA: weighted gene coexpression network analysis, KEGG: Kyoto Encyclopedia of Genes and Genomes.*

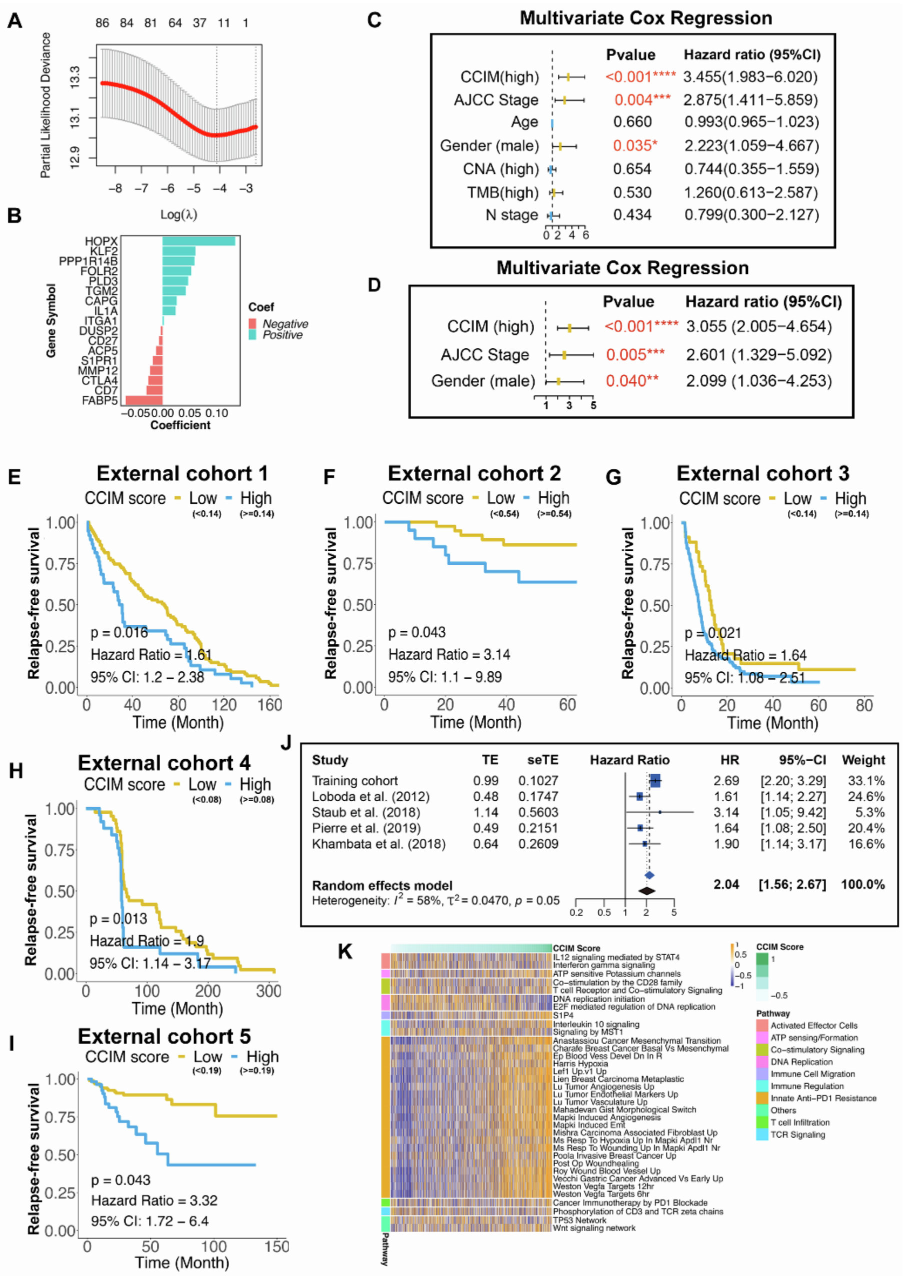


Fig. S5: Construction of the CCIM scoring system, related to Figure 5. A.) The LASSO Cox regression model was used to identify the most robust markers. B.) Distribution of LASSO coefficients of the gene signature. C.) Multivariate survival analysis of the CCIM score and other clinicopathological features in the TCGA-CRC cohort. D.) Multivariate survival analysis of the CCIM score and other significant clinicopathological features in the TCGA-CRC cohort. E.) The CCIM score stratified patients with poor prognosis in independent CRC cohort #1 (GSE28722) (Pts n = 125). F.) The CCIM score stratified patients with poor prognosis in independent CRC cohort #2 (GSE12945) (Pts n = 62). G.) The CCIM score stratified patients with a poor prognosis in independent CRC cohort #3 (GSE72970) (Pts n = 124). H.) The CCIM score stratified patients with a poor prognosis in independent CRC cohort #4 (GSE5851) (Pts n = 80). For Figure S5E-S5I best-cutoff method was applied for generating the CCIM high and low group. I.) The CCIM score stratified patients with a poor prognosis in independent CRC cohort #5 (TCGA-CRC) (Pts n = 274). J.) Meta-analysis for the training and testing cohorts. Immune-related pathway alterations in patients with distinct CCIM scores. K.) The pathway alterations were assessed with the GSVA method using their gene sets. *CCIM: colorectal cancer immunosuppressive module; LASSO: least absolute shrinkage and selection operator, CRC: colorectal cancer; pts: patients.*

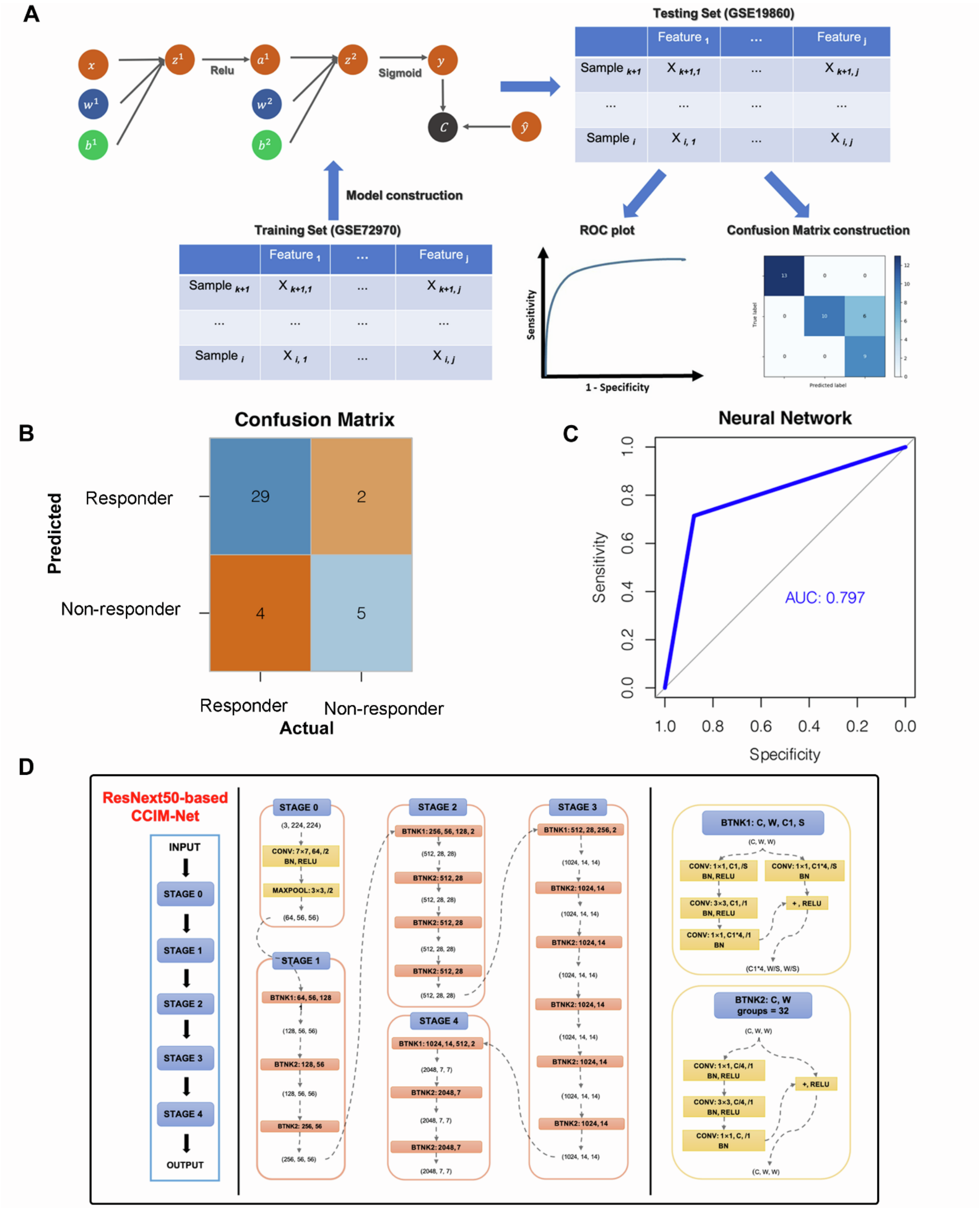


Fig. S6: A FOLR2$^+$ macrophage signature-based neural network model predicts the chemotherapy response, related to Figure 6. A.) Schematic illustration of FOLR2$^+$ macrophage signature prediction model construction and validation. B.) The confusion matrix for predicting the chemotherapy response by the FOLR2-RM model in an external cohort. The column indicates the predicted response, while the row represents the actual response. C.) The ROC analysis indicates the high sensitivity and specificity of the FOLR2-RM model. D.) The construction process of CCIM-Net by Resnext50. *ROC: receiver operating characteristic; FOLR2-RM model: FOLR2$^+$ macrophage-based model.*

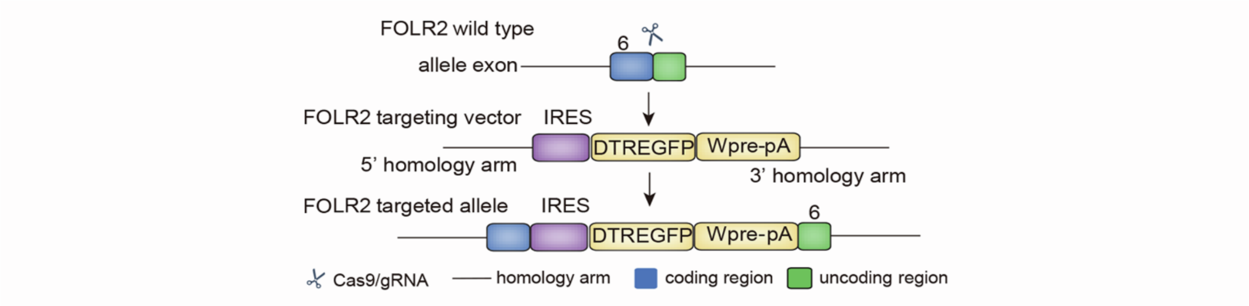


Fig. S7: The genomic schematic of editing the FOLR2-DTR-eGFP mice and gating strategy, related to Figure 7. The IRES-DTR-EGFP fusion gene was inserted into the 3'-UTR region of the FOLR2 gene locus.

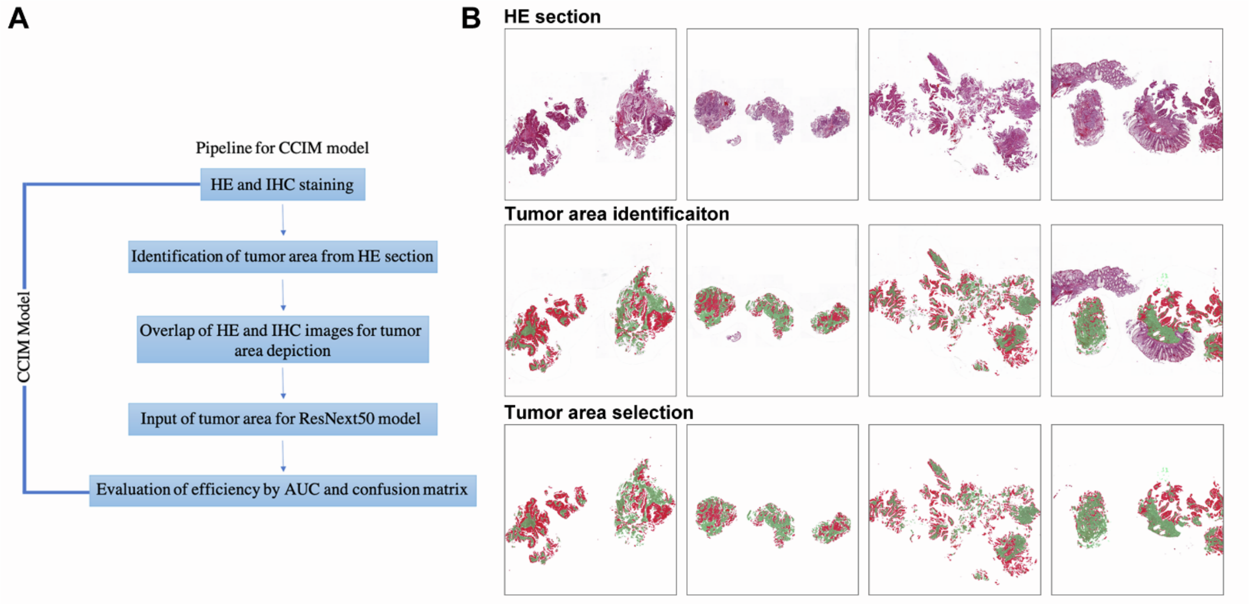


Fig. S8: The pipeline for the CCIM analysis, related to Figure 6. A.) The workflow for the deep learning pipeline in our manuscript. B.) The HE staining showing the identified tumor tissues (red indicates tumor cells and green indicates stroma cells) and para-tumor tissues on slides.

Table S1. Feature genes of myeloid cells. Related to Figure 2.

p_val	avg_log2FC	p_val_adj	cluster	gene
0	2.703467872	0	FOLR2$^+$ macrophage	APOE
0	2.569652735	0	FOLR2$^+$ macrophage	C1QC
0	2.540618403	0	FOLR2$^+$ macrophage	C1QB
0	2.520762997	0	FOLR2$^+$ macrophage	APOC1
0	2.478060246	0	FOLR2$^+$ macrophage	C1QA
0	2.156237286	0	FOLR2$^+$ macrophage	GPNMB
0	2.114256761	0	FOLR2$^+$ macrophage	RNASE1
0	1.806018055	0	FOLR2$^+$ macrophage	LGMN
0	1.702743808	0	FOLR2$^+$ macrophage	TREM2
0	1.63946639	0	FOLR2$^+$ macrophage	MSR1
0	1.790280153	0	OAS$^+$ macrophage	IFI44L
0	1.38366812	0	OAS$^+$ macrophage	OAS1
0	1.367907875	0	OAS$^+$ macrophage	OAS3
0	1.200912655	0	OAS$^+$ macrophage	CXCL10
1.74E-304	1.251750284	3.47E-301	OAS$^+$ macrophage	OASL
1.14E-280	1.740158953	2.28E-277	OAS$^+$ macrophage	TNFSF10
5.00E-272	2.076296858	9.99E-269	OAS$^+$ macrophage	MX1
1.48E-258	1.910414263	2.96E-255	OAS$^+$ macrophage	LY6E
1.89E-228	1.618872411	3.77E-225	OAS$^+$ macrophage	MT2A
1.87E-212	2.16275357	3.73E-209	OAS$^+$ macrophage	IFIT1
0	4.326570465	0	Mast cell	TPSAB1
0	4.151640883	0	Mast cell	TPSB2
0	3.719794919	0	Mast cell	CPA3
0	3.528531605	0	Mast cell	CLU
0	3.179552215	0	Mast cell	HPGDS
0	3.08481101	0	Mast cell	MS4A2
0	2.956581136	0	Mast cell	GATA2
0	2.915095278	0	Mast cell	LTC4S
3.13E-298	2.22089356	6.27E-295	Mast cell	CD9
5.56E-288	2.600934127	1.11E-284	Mast cell	VWA5A
0	2.314578538	0	FCN1$^+$ monocytes	S100A12
0	1.675024369	0	FCN1$^+$ monocytes	FCN1
0	1.443611669	0	FCN1$^+$ monocytes	RBP7
0	1.407530861	0	FCN1$^+$ monocytes	VCAN
0	1.224845612	0	FCN1$^+$ monocytes	CSTA
0	1.201167533	0	FCN1$^+$ monocytes	S100A8
0	1.109381689	0	FCN1$^+$ monocytes	CLEC12A
0	1.064708839	0	FCN1$^+$ monocytes	CFP
0	1.047715341	0	FCN1$^+$ monocytes	S100A9

0	1.01710355	0	FCN1$^+$ monocytes	PLBD1
0	1.806616701	0	OLR1$^+$ macrophage	CCL20
0	1.371324968	0	OLR1$^+$ macrophage	CXCL3
0	1.328000606	0	OLR1$^+$ macrophage	PLIN2
0	1.313447691	0	OLR1$^+$ macrophage	G0S2
0	1.229211049	0	OLR1$^+$ macrophage	NLRP3
0	1.212277595	0	OLR1$^+$ macrophage	CXCL8
0	1.168625057	0	OLR1$^+$ macrophage	UPP1
0	1.150691165	0	OLR1$^+$ macrophage	EREG
1.02E-292	1.432863049	2.04E-289	OLR1$^+$ macrophage	OLR1
2.09E-275	1.019056865	4.17E-272	OLR1$^+$ macrophage	TIMP1
0	1.631095239	0	DC cell	HLA-DQA1
0	1.390455704	0	DC cell	HLA-DQB1
0	1.364771189	0	DC cell	HLA-DMB
0	1.082507304	0	DC cell	HLA-DPB1
0	1.073994585	0	DC cell	HLA-DPA1
2.36E-275	1.452689755	4.72E-272	DC cell	PPA1
8.42E-202	1.153697115	1.68E-198	DC cell	GPR183
3.01E-191	1.060540832	6.01E-188	DC cell	UCP2
1.86E-188	1.774782864	3.72E-185	DC cell	CLEC10A
2.28E-170	1.211102005	4.55E-167	DC cell	CPVL
2.46E-124	0.801642856	4.93E-121	HSPA6$^+$ monocytes	PI3
5.59E-69	0.751508145	1.12E-65	HSPA6$^+$ monocytes	SLPI
1.04E-52	0.508773665	2.08E-49	HSPA6$^+$ monocytes	LRRC25
1.15E-37	0.499143034	2.30E-34	HSPA6$^+$ monocytes	FOS
7.02E-37	0.767185495	1.40E-33	HSPA6$^+$ monocytes	LRG1
8.01E-33	0.457828519	1.60E-29	HSPA6$^+$ monocytes	S100A6
4.74E-32	0.633038698	9.49E-29	HSPA6$^+$ monocytes	ZFAND2A
3.00E-27	0.479852005	6.00E-24	HSPA6$^+$ monocytes	CXCR4
2.10E-25	0.874316719	4.21E-22	HSPA6$^+$ monocytes	HSPA6
7.76E-24	0.598128716	1.55E-20	HSPA6$^+$ monocytes	HCAR3

Table S2. Feature genes of CD4$^+$ T cells. Related to Figure 4.

p_val	avg_log2FC	p_val_adj	cluster	gene
0	2.10186081	0	Treg	IL2RA
0	1.861650068	0	Treg	FOXP3
0	1.836449779	0	Treg	TNFRSF18
0	1.744619545	0	Treg	TNFRSF4
0	1.591211126	0	Treg	BATF
0	1.573961766	0	Treg	LAIR2
0	1.529423238	0	Treg	TNFRSF9
0	1.454852486	0	Treg	DUSP4
0	1.38027878	0	Treg	TIGIT
0	1.210566934	0	Treg	CTLA4
0	1.525099519	0	Naive	CCR7
7.99E-284	1.271369831	1.60E-280	Naive	MAL
1.40E-252	0.496372502	2.81E-249	Naive	PABPC1
6.67E-208	0.920624495	1.33E-204	Naive	PLAC8
3.22E-189	0.508841364	6.43E-186	Naive	LDHB
4.45E-178	0.702815017	8.90E-175	Naive	YPEL5
1.74E-168	0.912903959	3.49E-165	Naive	AREG
7.86E-151	0.624424222	1.57E-147	Naive	GPR183
2.12E-143	0.721272861	4.23E-140	Naive	SLC2A3
2.71E-139	0.620338115	5.42E-136	Naive	FOS
0	1.3892901	0	Th17	CCL5
0	1.188210898	0	Th17	ANXA1
0	0.931498989	0	Th17	GPR171
1.79E-292	0.92752317	3.58E-289	Th17	KLRB1
2.55E-285	0.72817835	5.10E-282	Th17	IL7R
1.47E-284	0.940234127	2.94E-281	Th17	FKBP11
1.66E-283	0.953685169	3.33E-280	Th17	ID2
6.20E-280	1.011205571	1.24E-276	Th17	ANKRD28
4.59E-244	0.954412521	9.18E-241	Th17	GPR65
1.32E-234	0.970994223	2.63E-231	Th17	IL17A
1.45E-149	1.288450419	2.90E-146	Exhaust	PDCD1
4.28E-146	0.937759753	8.55E-143	Exhaust	TIGIT
9.08E-146	1.444081009	1.82E-142	Exhaust	TOX2
9.98E-127	1.282741513	2.00E-123	Exhaust	ICA1
5.59E-115	0.767821251	1.12E-111	Exhaust	LIMS1
3.93E-107	0.888092037	7.86E-104	Exhaust	TBC1D4
8.71E-105	0.872393829	1.74E-101	Exhaust	SH2D1A
6.01E-100	0.763778656	1.20E-96	Exhaust	MAF
2.05E-90	0.622782937	4.09E-87	Exhaust	ITM2A

5.32E-75	1.568188242	1.06E-71	Exhaust	CXCL13
4.84E-249	2.754926942	9.68E-246	Memory	TMEM204
5.75E-20	0.639291027	1.15E-16	Memory	GIMAP7
4.56E-17	0.49324668	9.11E-14	Memory	RPS4Y1
1.80E-16	0.59686813	3.61E-13	Memory	GIMAP4
1.82E-14	0.392548485	3.63E-11	Memory	IL7R
8.93E-14	0.691505443	1.79E-10	Memory	PLAC8
7.00E-13	0.27137293	1.40E-09	Memory	LDHB
1.75E-12	0.403172704	3.50E-09	Memory	ANXA1
2.72E-08	0.410434881	5.43E-05	Memory	TRAT1
1.42E-07	0.639850827	0.000284346	Memory	CCR7

Table S3. Feature genes of CD8$^+$ T cells. Related to Figure 4.

p_val	avg_log2FC	p_val_adj	cluster	gene
0	2.200701212	0	Effector memory	GZMK
0	1.30252866	0	Effector memory	SH2D1A
0	1.21816221	0	Effector memory	ITGB2
0	1.023573045	0	Effector memory	CD27
0	0.830763101	0	Effector memory	CD74
8.73E-271	0.867991915	1.75E-267	Effector memory	HLA-DPB1
7.68E-255	1.167004015	1.54E-251	Effector memory	GZMH
8.49E-215	0.782787791	1.70E-211	Effector memory	GPR183
2.64E-200	0.678554339	5.27E-197	Effector memory	CCR7
5.64E-177	0.864636785	1.13E-173	Effector memory	HLA-DRA
0	1.894683286	0	Exhaust	TMIGD2
0	1.85965092	0	Exhaust	CD160
0	1.459115224	0	Exhaust	LDLRAD4
0	1.317366481	0	Exhaust	HOPX
0	1.169102265	0	Exhaust	KLRD1
0	0.937885175	0	Exhaust	CD7
1.54E-291	1.66042112	3.09E-288	Exhaust	SPRY1
6.94E-291	1.146572869	1.39E-287	Exhaust	GPR65
9.10E-285	1.224418415	1.82E-281	Exhaust	ITGA1
3.62E-283	0.908541583	7.24E-280	Exhaust	CD63
3.02E-196	1.276163725	6.03E-193	Exhaust NKT	HAVCR2
2.89E-175	1.846789084	5.78E-172	Exhaust NKT	CXCL13
3.98E-171	1.839221638	7.96E-168	Exhaust NKT	GNLY
4.34E-170	1.093208479	8.67E-167	Exhaust NKT	ENTPD1
5.70E-156	1.101150394	1.14E-152	Exhaust NKT	PHLDA1
4.11E-154	1.284014249	8.23E-151	Exhaust NKT	GZMB
5.07E-136	0.995602299	1.01E-132	Exhaust NKT	DUSP4
5.73E-125	0.821779459	1.15E-121	Exhaust NKT	CD70
2.60E-119	0.766485572	5.19E-116	Exhaust NKT	ACP5
2.37E-89	0.776973157	4.74E-86	Exhaust NKT	CD9
0	2.405764276	0	Tissue resident	ZNF683
7.47E-49	0.509462029	1.49E-45	Tissue resident	S100A4
8.58E-39	0.54810102	1.72E-35	Tissue resident	GZMB
6.50E-33	0.583926417	1.30E-29	Tissue resident	GZMH
1.34E-31	0.530854143	2.68E-28	Tissue resident	LGALS1
2.51E-27	0.368744099	5.02E-24	Tissue resident	CAPG
3.64E-25	0.487994878	7.27E-22	Tissue resident	HLA-DRA
5.12E-25	0.346868476	1.02E-21	Tissue resident	S100A6
5.16E-24	0.416649564	1.03E-20	Tissue resident	CLU

1.57E-23	0.375491203	3.15E-20	Tissue resident	MT2A
2.62E-165	2.310225713	5.24E-162	Tolerant	DKK3
3.50E-65	1.196683211	7.00E-62	Tolerant	GZMK
3.77E-30	0.758749457	7.53E-27	Tolerant	SH2D1A
2.53E-26	0.759367969	5.07E-23	Tolerant	CD27
9.11E-24	0.510185222	1.82E-20	Tolerant	CST7
3.34E-22	0.571940512	6.69E-19	Tolerant	GPR183
6.39E-19	0.58504908	1.28E-15	Tolerant	ITGB2
5.54E-18	0.529356191	1.11E-14	Tolerant	TRAT1
5.87E-17	0.414195428	1.17E-13	Tolerant	DUSP2
1.14E-15	0.66396579	2.27E-12	Tolerant	CCR7

Table S4. Cell-cell interaction pairs. Related to Figure 5.

Ligand	Receptor	Interaction.type	LRscore	Ligand_cell	Receptor_cell
CD86	CTLA4	paracrine	0.958	FOLR2$^+$ macrophage	Treg
CXCL16	CXCR6	paracrine	0.946	FOLR2$^+$ macrophage	Treg
MRC1	PTPRC	paracrine	0.933	FOLR2$^+$ macrophage	Treg
CD80	CTLA4	paracrine	0.933	FOLR2$^+$ macrophage	Treg
SPP1	ITGA4	paracrine	0.915	FOLR2$^+$ macrophage	Treg
CD14	ITGA4	paracrine	0.913	FOLR2$^+$ macrophage	Treg
ICAM1	IL2RG	paracrine	0.900	FOLR2$^+$ macrophage	Treg
LGALS1	PTPRC	paracrine	0.899	FOLR2$^+$ macrophage	Treg
IL18	IL18R1	paracrine	0.894	FOLR2$^+$ macrophage	Treg
CCL20	CCR6	paracrine	0.893	FOLR2$^+$ macrophage	Treg
F13A1	ITGA4	paracrine	0.891	FOLR2$^+$ macrophage	Treg
FN1	ITGA4	paracrine	0.890	FOLR2$^+$ macrophage	Treg
IL1RN	IL1R1	paracrine	0.881	FOLR2$^+$ macrophage	Treg
IL1B	IL1R1	paracrine	0.879	FOLR2$^+$ macrophage	Treg
VCAN	SELL	paracrine	0.851	FOLR2$^+$ macrophage	Treg
TGM2	ITGA4	paracrine	0.841	FOLR2$^+$ macrophage	Treg
CALM2	SELL	paracrine	0.832	FOLR2$^+$ macrophage	Treg
VCAN	ITGA4	paracrine	0.819	FOLR2$^+$ macrophage	Treg
THBS1	ITGA4	paracrine	0.789	FOLR2$^+$ macrophage	Treg
SPP1	S1PR1	paracrine	0.750	FOLR2$^+$ macrophage	Treg
C3	IFITM1	paracrine	0.737	FOLR2$^+$ macrophage	Treg
IL1A	IL1R1	paracrine	0.697	FOLR2$^+$ macrophage	Treg
MRC1	PTPRC	paracrine	0.935	FOLR2$^+$ macrophage	Exhaust CD8$^+$ T
CXCL16	CXCR6	paracrine	0.922	FOLR2$^+$ macrophage	Exhaust CD8$^+$ T
LGALS1	CD69	paracrine	0.919	FOLR2$^+$ macrophage	Exhaust CD8$^+$ T
LGALS1	PTPRC	paracrine	0.902	FOLR2$^+$ macrophage	Exhaust CD8$^+$ T
SPP1	ITGA4	paracrine	0.902	FOLR2$^+$ macrophage	Exhaust CD8$^+$ T
CD14	ITGA4	paracrine	0.899	FOLR2$^+$ macrophage	Exhaust CD8$^+$ T
ICAM1	IL2RG	paracrine	0.883	FOLR2$^+$ macrophage	Exhaust CD8$^+$ T
IL18	CD48	paracrine	0.876	FOLR2$^+$ macrophage	Exhaust CD8$^+$ T
F13A1	ITGA4	paracrine	0.874	FOLR2$^+$ macrophage	Exhaust CD8$^+$ T
FN1	ITGA4	paracrine	0.872	FOLR2$^+$ macrophage	Exhaust CD8$^+$ T
IL18	IL18R1	paracrine	0.818	FOLR2$^+$ macrophage	Exhaust CD8$^+$ T
TGM2	ITGA4	paracrine	0.818	FOLR2$^+$ macrophage	Exhaust CD8$^+$ T
VCAN	ITGA4	paracrine	0.794	FOLR2$^+$ macrophage	Exhaust CD8$^+$ T
THBS1	ITGA4	paracrine	0.760	FOLR2$^+$ macrophage	Exhaust CD8$^+$ T
C3	IFITM1	paracrine	0.724	FOLR2$^+$ macrophage	Exhaust CD8$^+$ T
CD86	CTLA4	paracrine	0.940	FOLR2$^+$ macrophage	Exhaust CD4$^+$ T
MRC1	PTPRC	paracrine	0.937	FOLR2$^+$ macrophage	Exhaust CD4$^+$ T

SPP1	ITGA4	paracrine	0.920	FOLR2$^+$ macrophage	Exhaust CD4$^+$ T
CD14	ITGA4	paracrine	0.918	FOLR2$^+$ macrophage	Exhaust CD4$^+$ T
CD80	CTLA4	paracrine	0.906	FOLR2$^+$ macrophage	Exhaust CD4$^+$ T
LGALS1	PTPRC	paracrine	0.905	FOLR2$^+$ macrophage	Exhaust CD4$^+$ T
CXCL16	CXCR6	paracrine	0.902	FOLR2$^+$ macrophage	Exhaust CD4$^+$ T
F13A1	ITGA4	paracrine	0.897	FOLR2$^+$ macrophage	Exhaust CD4$^+$ T
FN1	ITGA4	paracrine	0.896	FOLR2$^+$ macrophage	Exhaust CD4$^+$ T
LGALS1	CD69	paracrine	0.886	FOLR2$^+$ macrophage	Exhaust CD4$^+$ T
ICAM1	IL2RG	paracrine	0.885	FOLR2$^+$ macrophage	Exhaust CD4$^+$ T
TGM2	ITGA4	paracrine	0.850	FOLR2$^+$ macrophage	Exhaust CD4$^+$ T
VCAN	ITGA4	paracrine	0.829	FOLR2$^+$ macrophage	Exhaust CD4$^+$ T
C3	IFITM1	paracrine	0.826	FOLR2$^+$ macrophage	Exhaust CD4$^+$ T
CCL20	CCR6	paracrine	0.824	FOLR2$^+$ macrophage	Exhaust CD4$^+$ T
IL18	CD48	paracrine	0.823	FOLR2$^+$ macrophage	Exhaust CD4$^+$ T
THBS1	ITGA4	paracrine	0.799	FOLR2$^+$ macrophage	Exhaust CD4$^+$ T
SPP1	S1PR1	paracrine	0.774	FOLR2$^+$ macrophage	Exhaust CD4$^+$ T
VCAN	SELL	paracrine	0.586	FOLR2$^+$ macrophage	Exhaust CD4$^+$ T
CALM2	SELL	paracrine	0.550	FOLR2$^+$ macrophage	Exhaust CD4$^+$ T
MRC1	PTPRC	paracrine	0.940	FOLR2$^+$ macrophage	Tolerant CD8$^+$ T
LGALS1	PTPRC	paracrine	0.910	FOLR2$^+$ macrophage	Tolerant CD8$^+$ T
CXCL16	CXCR6	paracrine	0.908	FOLR2$^+$ macrophage	Tolerant CD8$^+$ T
SPP1	ITGA4	paracrine	0.902	FOLR2$^+$ macrophage	Tolerant CD8$^+$ T
CD14	ITGA4	paracrine	0.899	FOLR2$^+$ macrophage	Tolerant CD8$^+$ T
LGALS1	CD69	paracrine	0.897	FOLR2$^+$ macrophage	Tolerant CD8$^+$ T
IL18	CD48	paracrine	0.897	FOLR2$^+$ macrophage	Tolerant CD8$^+$ T
SPP1	S1PR1	paracrine	0.896	FOLR2$^+$ macrophage	Tolerant CD8$^+$ T
F13A1	ITGA4	paracrine	0.874	FOLR2$^+$ macrophage	Tolerant CD8$^+$ T
FN1	ITGA4	paracrine	0.872	FOLR2$^+$ macrophage	Tolerant CD8$^+$ T
ICAM1	IL2RG	paracrine	0.852	FOLR2$^+$ macrophage	Tolerant CD8$^+$ T
TGM2	ITGA4	paracrine	0.818	FOLR2$^+$ macrophage	Tolerant CD8$^+$ T
VCAN	ITGA4	paracrine	0.793	FOLR2$^+$ macrophage	Tolerant CD8$^+$ T
THBS1	ITGA4	paracrine	0.760	FOLR2$^+$ macrophage	Tolerant CD8$^+$ T
VCAN	SELL	paracrine	0.743	FOLR2$^+$ macrophage	Tolerant CD8$^+$ T
CALM2	SELL	paracrine	0.714	FOLR2$^+$ macrophage	Tolerant CD8$^+$ T
CLEC2B	KLRF1	paracrine	0.709	FOLR2$^+$ macrophage	Tolerant CD8$^+$ T

Table S5. Clinical Characteristics of 1405 CRC Patients Involved in This Study. Related to Figure 5.

Characteristics	Number of patients (n =1405)
Cohort	
GSE14333	290 (20.6%)
GSE17536	177 (12.6%)
GSE17537	55 (3.9%)
GSE33113	96 (6.9%)
GSE37892	130 (9.3%)
GSE39582	585 (41.6%)
GSE56699	72 (5.1%)
Gender	
Male	719 (51.2%)
Female	608 (43.3%)
NA	78 (5.5%)
Median age (years)	66.67 (22-97)
Location	
Left	233 (16.6%)
Right	82 (5.8%)
NA	1090 (77.6%)
Satge	
I	66 (4.7%)
II	512 (36.4%)
III	343 (24.4%)
IV	116 (8.3%)
NA	368 (26.2%)
Grade	
I	17 (1.2%)
II	153 (10.9%)
III	34 (2.4%)
NA	1201 (85.5%)
T stage	
Tis	3
T0	1
T1	12 (0.8)
T2	49 (3.5%)
T3	379 (27%)
T4	119 (8.5%)
NA	842 (60%)
N stage	
N0	314 (22.3%)
N1	137 (9.8%)
N2	100 (7.1%)
N3	6 (0.4%)
NA	848 (60.4%)
M stage	
M0	499 (35.5%)
M1	61 (4.3%)
NA	845 (60.2%)
MMR status	
dMMR	77 (5.5%)
pMMR	459 (32.7%)
NA	869 (61.8%)
TP53 status	
Mutation	190 (13.5%)
Wild	161 (11.5%)
NA	1054 (75%)

KRAS status	
Mutation	217 (15.5%)
Wild	328 (23.3%)
NA	860 (61.2%)
BRAF status	
Mutation	51 (3.6%)
Wild	461 (32.8%)
NA	893 (63.6%)
Treatment	
Naive	497 (35.4%)
Chemotherapy	358 (25.5%)
NA	550 (39.1%)

Table S6. Clinical Characteristics of 26 CRC Patients Involved in This Study Related to Figure 1

Characteristics	**Number of patients (n = 26)**
Gender	
Male	11 (42.3%)
Female	15 (57.7%)
Median age (years)	69 (40-89)
Maximum tumor size	
Small (<38 cm²)	19 (73.1%)
Medium (38-69 cm²)	4 (15.4%)
Large (>69 cm²)	4 (15.4%)
Stage	
I	1 (3.81%)
II	13 (50%)
III	11 (42.3%)
IV	1 (3.8%)
Grade	
Low	3 (11.5%)
Low or Moderate	4 (15.4%)
Moderate	15 (57.7%)
High	4 (15.4%)

Data are median (IQR) or n (%).

Table S7. Clinical Characteristics of 181 CRC Patients Involved in This Study. Related to Figure 6.

Characteristics	Number of patients (n = 181)
Gender	
Male	95 (52.5%)
Female	86 (47.5%)
Median age (years)	62 (21-82)
Location	
L-CRC	116 (64.1%)
R-CRC	65 (35.9%)
KRAS Mutation	
Wild	119 (65.7%)
Mutation	62 (34.3%)
Response	
Responder (CR+PR)	118 (65.2%)
Non-responder (SD+PD)	63 (34.8%)

Data are median (IQR) or n (%).
R-CRC = right-sided colorectal cancer; L-CRC, left-sided colorectal cancer.
CR = complete response; PR = partial response; PD = progressive disease; SD = stable disease.

Table S8. Clinical Characteristics of 45 CRC Patients Involved in This Study. Related to Figure 6.

Characteristics	Number of patients (n = 45)
Gender	
Male	27 (60%)
Female	18 (40%)
Median age (years)	66 (25-78)
Location	
LCRC	33 (73.3%)
RCRC	12 (26.7%)
KRAS Mutation	
Wild	28 (62.2%)
Mutation	17 (37.8%)
Response	
Responder (CR+PR)	34 (75.6%)
Non-responder (SD+PD)	11 (24.4%)

Data are median (IQR) or n (%).
R-CRC = right-sided colorectal cancer; L-CRC, left-sided colorectal cancer.
CR = complete response; PR = partial response; PD = progressive disease; SD = stable disease.