Cancer Immunology, Immunotherapy (2025) 74:255
https://doi.org/10.1007/s00262-025-04119-8

RESEARCH

Diversity of mast cell subpopulations in the tumor microenvironment of colorectal cancer and their prognostic implications

Tianyu Qiao[1,3] · Chao Ding[2,3] · Songtao Yu[1] · Wenyang Li[1] · Yonghou Zhao[3] · Guiyu Wang[1]

Received: 30 April 2025 / Accepted: 17 June 2025

Abstract

Background Colorectal cancer (CRC) is one of the most common and deadly malignancies worldwide, with a particularly low 5-year survival rate in advanced patients. Immune cells in the tumor microenvironment, especially mast cells, play crucial roles in tumor initiation and progression. However, the dual role of mast cells in CRC remains poorly understood.

Methods In this study, we used single-cell RNA sequencing (scRNA-seq), bulk RNA sequencing, and bioinformatics analyses to explore the heterogeneity of mast cell subpopulations in the CRC tumor microenvironment and their relationship with prognosis. We analyzed gene expression signatures associated with mast cell subpopulations derived from single-cell data of 40 CRC tumor samples and combined bulk RNA-seq data from HMU, GEO, and TCGA cohorts for prognostic prediction. Non-negative matrix factorization was used for clustering of mast cell subpopulations, followed by analysis of their specific gene markers, transcription factor activity, and biological pathways. Survival analysis and ROC curves were performed to assess their prognostic significance.

Results Mast cells in the CRC tumor microenvironment were classified into three distinct subpopulations, each with unique gene markers and functional pathways. Mast cell subpopulations 1 and 3 were highly associated with pro-tumor pathways, while mast cell subpopulation 2 primarily exhibited anti-tumor immune regulatory characteristics. High expression of mast cell subpopulations 1 and 3 was associated with poor survival prognosis, while high expression of subpopulation 2 was linked to a better survival outcome. Key marker genes such as DNAJB1, SEMA7A, and XCR1 were identified as potential prognostic factors, with high expression of DNAJB1 and SEMA7A being significantly associated with poor prognosis, while high expression of XCR1 was linked to a favorable prognosis.

Conclusion This study reveals the functional heterogeneity of mast cell subpopulations in the CRC tumor microenvironment and their differential roles in tumor progression. Identification of mast cell subpopulation-specific marker genes provides new molecular targets for clinical diagnosis, prognostic prediction, and personalized immunotherapy in CRC.

Keywords Colorectal cancer · Tumor microenvironment · Mast cells · Single-cell RNA sequencing · Prognosis · Immunotherapy · Marker genes

Introduction

Colorectal cancer (CRC) is one of the most common and deadly malignancies worldwide [1]. According to the World Health Organization (WHO), CRC has become the third most common cancer globally and the second leading cause of cancer-related death [2]. Despite recent advances in early screening, surgical treatment, chemotherapy, targeted therapy, and immunotherapy, the 5-year survival rate for CRC remains low, especially in advanced-stage patients [3]. Therefore, treatment of CRC continues to pose a significant challenge, necessitating deeper research into its pathogenesis and the development of more effective

Tianyu Qiao and Chao Ding have contributed equally to this work.

✉ Yonghou Zhao
18845055802@163.com

✉ Guiyu Wang
guiywang@hrbmu.edu.cn

[1] Department of Colorectal Surgery, The Second Affiliated Hospital of Harbin Medical University, Harbin, China

[2] Department of General Surgery, The Second Affiliated Hospital of Harbin Medical University, Harbin, China

[3] Heilongjiang Mental Hospital, Harbin, China

Published online: 30 June 2025

therapies. The occurrence and progression of CRC involve complex interactions of multiple factors, including genetic susceptibility, environmental influences, and changes in the tumor microenvironment. The tumor microenvironment (TME) refers to the complex environment surrounding tumor cells, which includes non-tumor cells, blood vessels, immune cells, extracellular matrix, and signaling molecules [4]. The tumor microenvironment plays a critical role in cancer initiation, progression, metastasis, and therapeutic response. Immune cells, especially mast cells (MCs), play a significant role in the CRC microenvironment.

Mast cells are immune cells derived from the bone marrow, widely distributed in various tissues, particularly in the skin, lungs, intestines, and urinary tract [5, 6]. Mast cells play a key role in immune responses, mainly by releasing various bioactive substances stored in their granules, such as histamine, cytokines, chemokines, proteases, and lipid mediators, which regulate immune responses, inflammation, and tissue repair [7]. The role of mast cells in the tumor immune microenvironment has been a focus of research. Mast cells are not only involved in allergic reactions, antimicrobial immunity, and tissue repair but are also believed to be closely related to tumor initiation and progression[8]. Mast cells can regulate the immune response in the tumor microenvironment by secreting various mediators, thereby influencing tumor growth, metastasis, and response to treatment [9].

The role of mast cells in cancer is complex and bidirectional. Studies have shown that mast cell abundance is closely associated with cancer prognosis [10]. Some studies suggest that mast cell infiltration in tumors correlates with higher malignancy, increased metastasis, and drug resistance. In such cases, mast cells may promote tumor progression by secreting pro-tumor factors [11, 12]. In contrast, other studies have reported that mast cells can suppress tumor development. They may do so by enhancing anti-tumor immune responses and inducing apoptosis in tumor cells. In the cancer immune microenvironment, mast cells serve as key regulatory immune cells. They interact with tumor-associated immune cells—such as T cells, B cells, and macrophages—by releasing bioactive substances [13]. Through the secretion of cytokines and chemokines, mast cells can also recruit other immune cells to the tumor site, influencing immune evasion mechanisms.

The occurrence of CRC is closely related to the interactions of various cells in the tumor microenvironment. As one of the key immune cells in the tumor microenvironment, the role of mast cells in CRC is not yet fully understood. The dual role of mast cells in CRC makes them a promising research subject. Future studies may reveal their important roles in immune evasion, tumor metastasis, and immunotherapy. Therefore, further investigation into the functional mechanisms of mast cells in CRC will provide new insights and potential targets for tumor immunotherapy strategies.

Materials and methods

Patients

This study included 40 formalin-fixed paraffin-embedded CRC specimens from patients treated at the Second Affiliated Hospital of Harbin Medical University between 2010 and 2020 (HMU Cohort). The inclusion criteria were: (1) Patients aged 18 to 75; (2) Diagnosis of adenocarcinoma confirmed by postoperative pathology; (3) Underwent R0 surgical resection; (4) At least 5 years of follow-up data available. Exclusion criteria were: (1) patients who had received neoadjuvant therapy before surgery and (2) patients with incomplete clinicopathological data or missing postoperative follow-up and treatment information. The study was approved by the ethics committee of the Second Affiliated Hospital of Harbin Medical University (Approval number: YISKY2024-269).

RNA isolation and sequencing

The RNA quality and quantity were assessed through several steps: (1) Initial contamination and degradation checks were performed using 1% agarose gel electrophoresis; (2) RNA purity and concentration were measured with a NanoPhotometer® spectrophotometer; and (3) RNA integrity was further analyzed using the RNA Nano 6000 Assay Kit on the Bioanalyzer 2100 system [14]. These RNA samples were then used for library preparation. First, mRNA was captured using mRNA Capture Beads with Oligo(dT), followed by purification with Binding and Washing Buffers. The mRNA was then randomly fragmented to 100–200 nt using Fragmentation Buffer and reverse-transcribed into cDNA. The resulting cDNA was purified using DNA Clean Beads and prepared for adaptor ligation. After ligation with adaptors containing UMIs, cDNA underwent one-step PCR amplification. The PCR products were then purified again using DNA Clean Beads and eluted in nuclease-free water. The final library was pooled and sequenced on the Illumina Novaseq 6000 platform, generating 6G of raw paired-end data (150 nt reads)[15].

For initial processing, Perl scripts were used to process raw FASTQ data. Low-quality reads, those containing adapters, or poly-N sequences were removed, resulting in clean reads. All subsequent analyses were performed using these high-quality clean reads. The reference genome (Homo sapiens GRCh38.103) index was created using STAR (version 2.7.11b), and the clean reads were aligned to this

reference genome with STAR. Gene expression counts were obtained using RSEM (version 1.3.3) [16].

Acquisition and preprocessing of single-cell data

The CRC single-cell RNA sequencing (scRNA-seq) data were obtained from a dataset published by Pelk et al., comprising 236,950 cells from 40 tumor tissue samples (Table 1) [17]. The scRNA-seq data were processed using the R package Seurat (version 5.2.1).

First, we converted the raw gene expression matrices into individual Seurat objects. Three quality control steps were applied: (1) genes expressed in fewer than five cells were removed; (2) cells with fewer than 100 detected genes were excluded; and (3) cells with more than 5% mitochondrial gene content were filtered out. Each sample was normalized using the NormalizeData() function with the default "LogNormalize" method, and the top 2,000 highly variable genes were identified using FindVariableFeatures() (selection.method = "vst").

To correct for batch effects and integrate the 40 samples, we used Seurat's integration workflow. Specifically, we applied FindIntegrationAnchors() with dims = 1:20 to identify shared features across samples, followed by IntegrateData() using the same dimensions. The integrated dataset was then scaled using ScaleData() and dimensionality reduction was performed using RunPCA(). The top 20 principal components were selected for downstream clustering analysis using FindNeighbors() and FindClusters().

Differentially expressed genes for each cluster were identified using FindAllMarkers() with the following thresholds: FDR < 0.05 and |log2(fold change)| > 0.25.

For cell-type annotation, major immune and stromal cell types were identified based on canonical marker genes, including CD3D, CD8A, CD4, CD56, FOXP3 (T cells and NK cells); CD79A, MS4A1 (B cells); CD14, CD68 (myeloid cells); COL1A2, COL3A1 (fibroblasts); VWF, PECAM1 (endothelial cells); EPCAM (epithelial cells); and TPSAB1, CPA3 (mast cells).

Table 1 Datasets related colorectal cancer

Number	Samples	Platform	Type
GSE225857	40	GPL24676	scRNA-seq
GSE110223	13	GPL96	Array
GSE12945	62	GPL96	Array
GSE24514	34	GPL96	Array
GSE46862	69	GPL6244	Array
GSE143985	91	GPL570	Array
GSE161158	250	GPL570	Array
GSE39582	566	GPL570	Array
GSE110224	17	GPL570	Array
GSE13067	74	GPL570	Array
GSE13294	155	GPL570	Array
GSE18088	53	GPL570	Array
GSE18105	77	GPL570	Array
GSE31595	37	GPL570	Array
GSE33113	90	GPL570	Array
GSE35452	46	GPL570	Array
GSE35896	62	GPL570	Array
GSE37892	130	GPL570	Array
GSE38832	122	GPL570	Array
GSE39084	70	GPL570	Array
GSE45404	42	GPL570	Array
GSE60697	20	GPL570	Array
GSE64857	81	GPL570	Array
GSE75316	59	GPL570	Array
GSE81980	150	GPL570	Array
GSE9348	70	GPL570	Array
GSE81558	23	GPL15207	Array
GSE103479	156	GPL23985	Array
TCGA Cohort	571	Illumina	Bulk RNA-seq
HMU Cohort	80	Illumina	Bulk RNA-seq

Acquisition and preprocessing of bulk RNA-Seq data

We performed a comprehensive search for CRC transcriptome data in public databases, including GEO (Gene Expression Omnibus, GEO Cohort) and TCGA (The Cancer Genome Atlas, TCGA Cohort). Inclusion criteria were as follows: (1) patients had not received chemotherapy or radiotherapy prior to surgery; (2) samples were primary CRC tumor tissues; (3) datasets contained at least 1,000 genes; and (4) microarray datasets were in CEL format. A total of 27 microarray datasets were collected from GEO, including 2,619 tumor samples (Table 1). The R package affy (version 1.78.0) was used to process and normalize the CEL files of these microarray datasets, and probe IDs were converted to gene symbols.

Cell–cell interaction analysis

We used the Python package CellPhoneDB (version 2.0) [18] to construct a cell–cell interaction network in the CRC tumor immune microenvironment, following the software's default settings. The potential interaction strength between different cell subpopulations was predicted based on the expression levels of ligand-receptor pairs. The interactions were subsequently trimmed based on significance ($P < 0.05$). Specifically, interactions were defined as input or output if the cell expressed a receptor or a ligand, respectively. Biologically relevant ligand-receptor pairs were analyzed between different cell subpopulations. The relative expression levels (Z-scores) and adjusted P-values

of ligand-receptor pairs were visualized using dot plots or interaction heatmaps.

Prediction of transcription factor activity using SCENIC

We used the SCENIC Python workflow (version 0.12.1, https://github.com/aertslab/pySCENIC) to analyze gene regulatory networks (GRNs) and transcription factor (TF) activity, using default parameters [19]. The input data were a normalized expression matrix of the target cells. The analysis used the RcisTarget and GRNboost transcription factor binding site databases (Homo sapiens), which are available from official resources (https://pyscenic.readthedocs.io/en/latest). The SCENIC workflow generated regulon-specificity scores and identified active transcription factors in a binary matrix. Additionally, we used the Wilcoxon test on the AUC matrix to identify differentially active transcription factors, with thresholds set at FDR < 0.05 and Fold change > 2. The results were visualized using heatmaps.

CytoTRACE analysis

The CytoTRACE (version 0.3.3) algorithm, developed by Gulati et al., is an advanced tool for analyzing scRNA-Seq data, with the core function of capturing, refining, and quantifying gene expression levels that are highly correlated with single-cell gene counts. After CytoTRACE calculations, each cell was assigned a score describing its stemness state within the specific dataset. As a reliable computational method, CytoTRACE can accurately predict cell differentiation states and has been validated in large-scale datasets, outperforming traditional stemness evaluation algorithms [20]. In this study, we used the R package CytoTRACE to compute CytoTRACE scores for malignant cells, with scores ranging from 0 to 1. Higher scores indicate stronger stemness (lower differentiation), while lower scores suggest weaker stemness (higher differentiation).

Pseudotime trajectory analysis

We used Monocle2 (version 2.20.0) [21] for pseudotime analysis to determine the differentiation trajectory of cells. After reading the UMI matrix from the Seurat object, we created the object using the newCellDataSet function. For trajectory analysis, we selected genes with a mean expression greater than 0.1, followed by dimensionality reduction using the DDRTree method and sorting cells with the orderCells function.

Gene set enrichment analysis

To investigate the heterogeneous expression features of different cell subtypes, we used the R package GSVA (version 1.44.3) to perform Gene Set Variation Analysis (GSVA). Pathway enrichment analysis used gene sets from the Molecular Signatures Database (MSigDB, https://www.gsea-msigdb.org/gsea/msigdb/index.jsp), specifically the Hallmark gene sets.

Statistical analysis

We used the R package survminer (version 0.4.9) to calculate the optimal cutoff value using the surv_cutpoint function and grouped patients accordingly. Kaplan–Meier survival curves were plotted for different subtypes of patients using the survminer and survival (version 3.3–1) packages, and the log-rank test was performed to assess the significance of differences. All statistical analyses were conducted using the R programming language (version 4.2.0), with $P < 0.05$ considered statistically significant.

Results

Single-cell atlas of the tumor microenvironment in colorectal cancer

We obtained 10 × scRNA-seq data from a previous study [17], including 40 colon tumor samples. To eliminate batch effects between samples, the single-cell dataset was integrated, ensuring minimal differences in the major cell-type characteristics across different patients. After quality control and filtering, a total of 236,950 immune cells were retained for unsupervised clustering analysis, resulting in 33 clusters (Fig. 1A). Based on typical marker genes, we successfully identified 11 major cell subtypes, including B cells, CD4 + T cells, NK cells, exhausted T cells, regulatory T cells, cytotoxic T cells, germinal neutrophils, myeloid cells, fibroblasts, epithelial cells, and endothelial cells (Fig. 1B-C). Additionally, we displayed the marker genes of six major cell types (Fig. 1D-I), including myeloid cells (CD68), B cells (CD79A), T cells (CD3E), fibroblasts (COL1A2), endothelial cells (VWF), and epithelial cells (EPCAM).

Myeloid cell subpopulation characteristics

Through unsupervised clustering analysis, 19,722 myeloid cells were divided into 22 clusters (Fig. 2A), which were further classified into 11 major subtypes (Fig. 2B-C). These subtypes included monocytes, M0 macrophages, M1 macrophages, M2 macrophages, tumor-associated

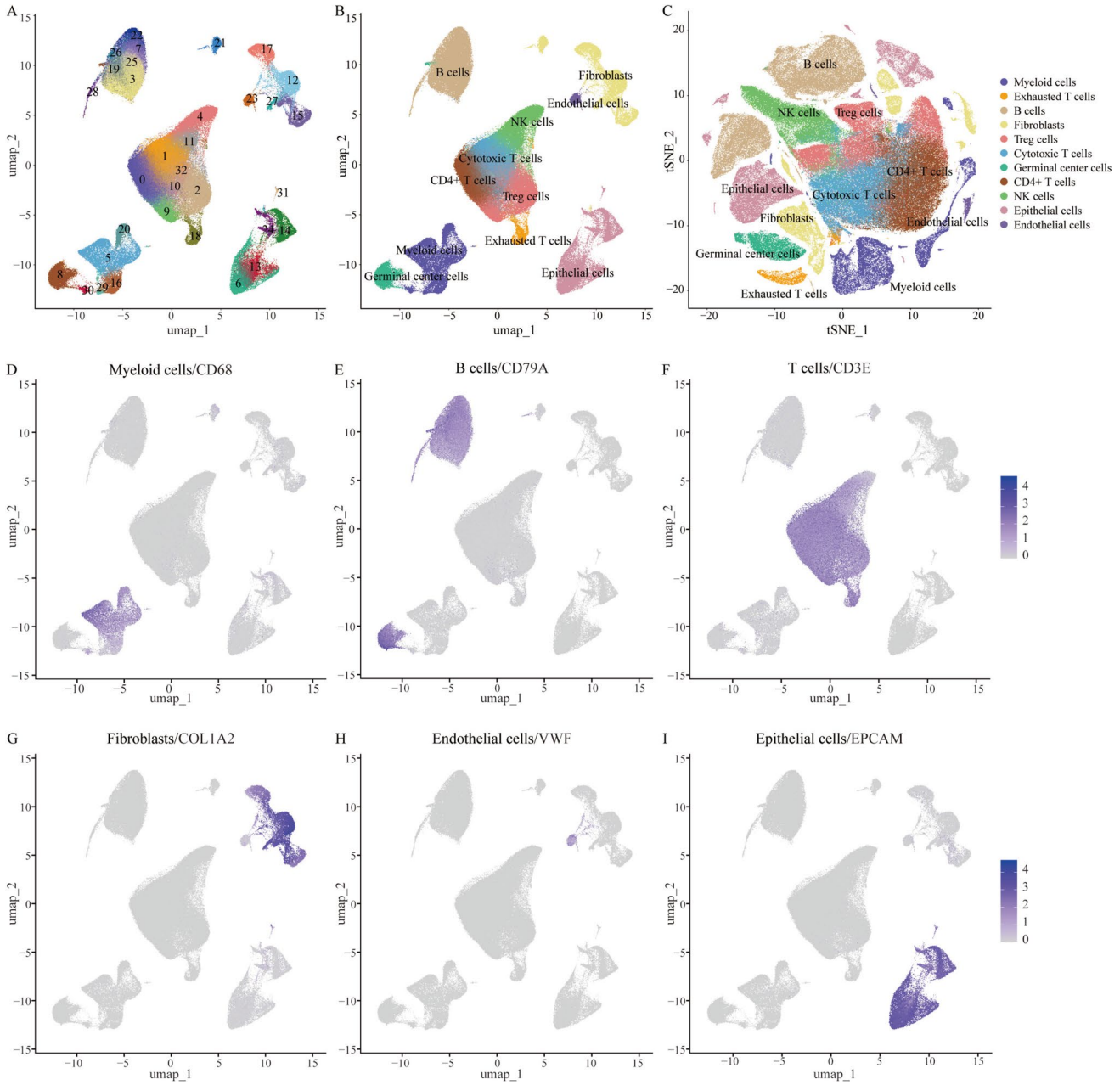


Fig. 1 Single-cell sequencing reveals the cellular heterogeneity landscape in colorectal cancer. **A** UMAP dimensionality reduction clustering plot of all cells, showing 33 distinct cell clusters, each marked with a different color. **B** UMAP visualization of major cell types. **C** t-SNE dimensionality reduction showing the distribution of cell types. **D–I** Differential expression plots of marker genes for myeloid cells, B cells, T cells, fibroblasts, endothelial cells, and epithelial cells

macrophages (TAM), classical dendritic cells 1 (cDC1), classical dendritic cells 2 (cDC2), activated dendritic cells (aDC), plasmacytoid dendritic cells (pDC), central granulocytes, and mast cells. The analysis revealed significant molecular differences between the subpopulations (Fig. 2D). For instance, mast cells specifically overexpress XCR1, CLEC9A, and CLNK, whereas TAMs highly express SPP1, ERRFI1, and RNASE1. When comparing characteristics of different subpopulations, we found significant differences in cell subpopulation features from various tissues and patients. Despite the lower proportion of mast cells among all cell types, their gene expression levels were among the highest, suggesting high activity in these subpopulations. In contrast, monocytes dominate the tumor microenvironment but exhibit lower overall gene expression, indicating weaker activity.

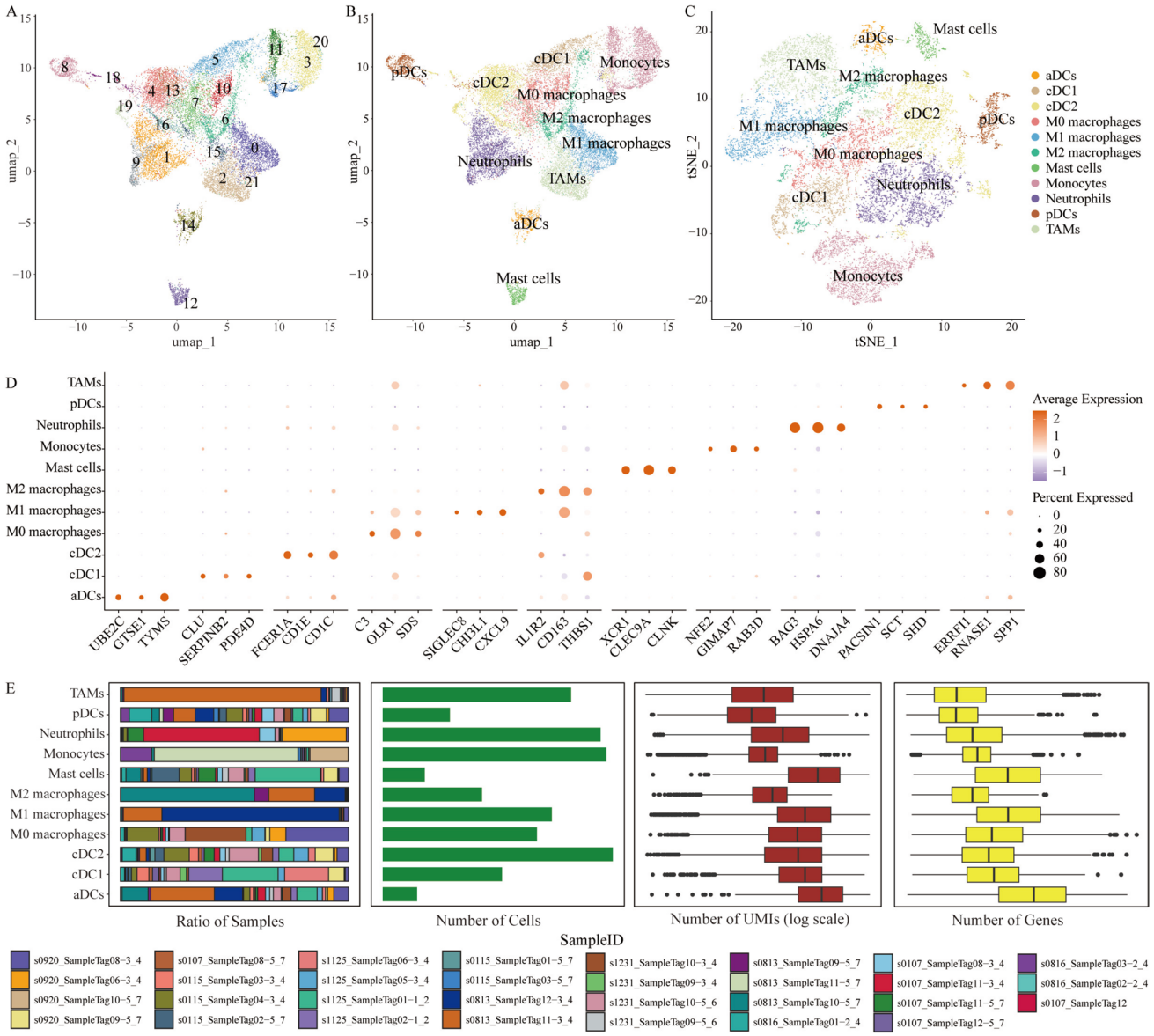


Fig. 2 Heterogeneity of myeloid cell subpopulations. **A** UMAP distribution plot of myeloid cells, showing the distribution of unannotated cell clusters, with different clusters marked by distinct colors and numbers. **B** UMAP plot after annotating myeloid cell subpopulations, displaying the distribution of different subpopulations, including dendritic cells (DC), monocytes (Mon), macrophages (Mac), and others. **C** tSNE plot after annotating myeloid cell subpopulations, displaying the distribution of different subpopulations, including dendritic cells (DC), monocytes (Mon), macrophages (Mac), and others. **D** Bubble plot showing the differential expression of marker genes for each cell subpopulation. E. Basic characteristics of myeloid cell subpopulations, from left to right: distribution of subpopulation proportions across different samples (stacked bar chart); cell count statistics for each subpopulation (bar chart); distribution of UMI counts for each subpopulation (box plot, logarithmic scale); distribution of gene counts for each subpopulation (box plot)

Heterogeneity of mast cells in colorectal cancer

Given the high activity of mast cells in the tumor microenvironment, we explored the activity differences of mast cells from a heterogeneity perspective in CRC. First, we performed non-negative matrix factorization on 535 mast cells and found that $k=3$ was the optimal number of clusters (Fig. 3A), resulting in three mast cell subpopulations: Mast Cell 1, Mast Cell 2, and Mast Cell 3. To identify specific marker genes for each mast cell subpopulation, we performed differential expression analysis and selected specific marker genes based on thresholds (avg_log2FC > 0.5 & pct.1 > 0.4 & pct.2 < 0.6). The differential expression results revealed that Mast Cell 1 had five specific marker

genes, Mast Cell 2 had 51, and Mast Cell 3 had 25 (Fig. 3B). We also displayed the top 10 differentially expressed genes for each subpopulation (Fig. 3C), such as DNAJB1 in Mast Cell 1, XCR1 in Mast Cell 2, and SEMA7A in Mast Cell 3.

Next, we performed GO and KEGG pathway enrichment analysis on the specific marker genes of each mast cell subpopulation, revealing significant functional heterogeneity across the subpopulations. Mast Cell 1 and Mast Cell 3 were mainly enriched in cancer-related pathways (Fig. 3D-E), such as "serine phosphorylation of STAT protein," PPAR signaling, and AMPK signaling in Mast Cell 1, and "Transcriptional misregulation in cancer," FoxO signaling, and MAPK signaling in Mast Cell 3. In contrast, Mast Cell 2 was mainly enriched in pathways related to inflammatory responses, such as "regulation of inflammatory response" and "positive regulation of cytokine production."

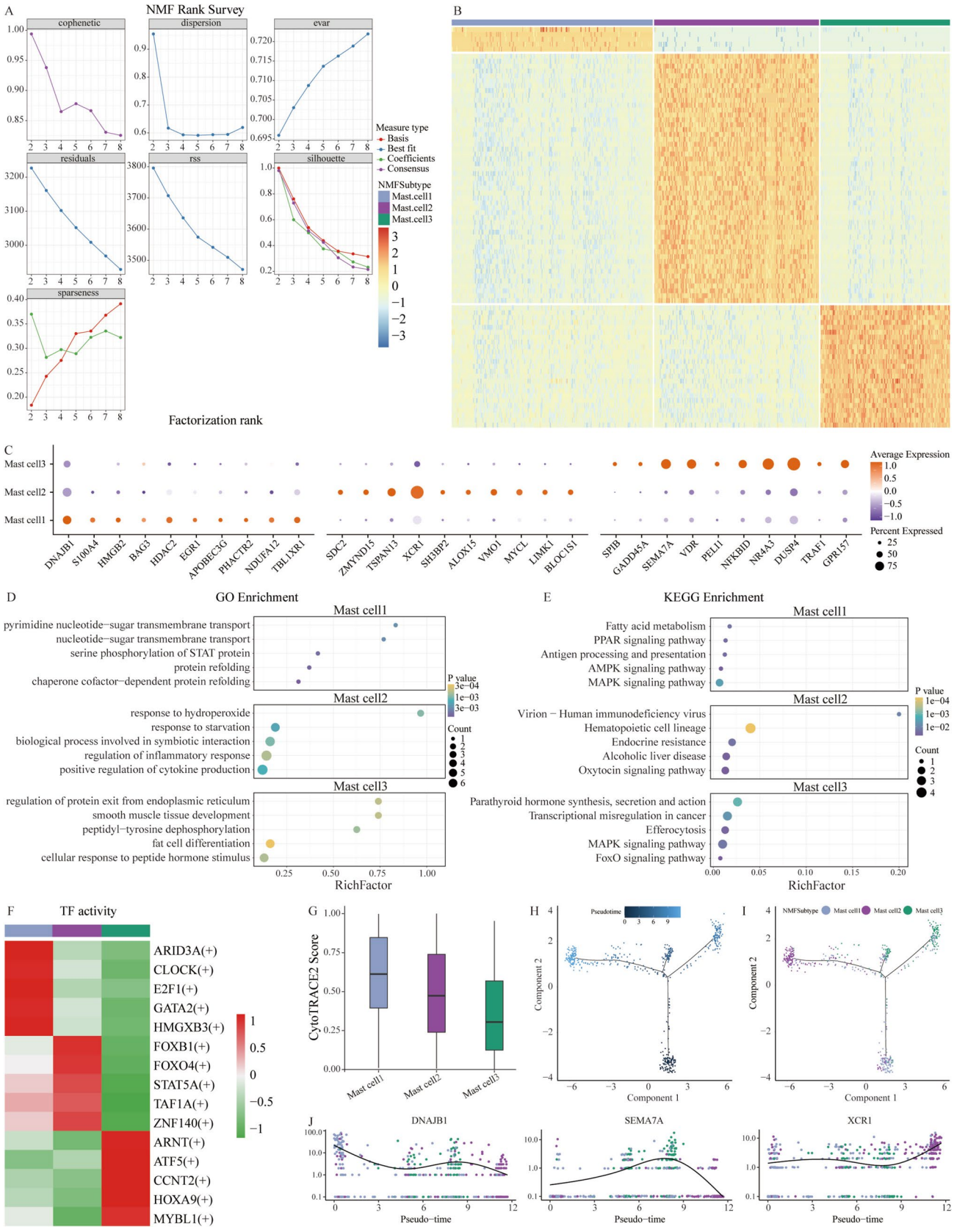
Further SCENIC transcriptional network analysis revealed highly specific transcription factor activity in each mast cell subpopulation (Fig. 3F). In Mast Cell 1, transcription factors like CLOCK and GATA2 had high regulatory specificity, suggesting their potential role in promoting tumor proliferation. In Mast Cell 2, transcription factors such as FOXB1 and TAF1A exhibited high regulatory specificity; previous studies have shown that knocking out FOXB1 enhances cisplatin sensitivity and inhibits cell proliferation [22]. In Mast Cell 3, transcription factors like ATF5 and HOXA9 had high regulatory specificity; HOXA9 is considered a prognostic factor for various cancers [23].

Additionally, CytoTrace analysis was used to evaluate the differentiation potential of different mast cell subpopulations, with results showing a decreasing trend in CytoTrace scores across the three subpopulations, indicating that Mast Cell 1 may differentiate into Mast Cell 2 and Mast Cell 3 (Fig. 3G). Pseudotime trajectory analysis further confirmed the developmental relationship between the three subpopulations (Fig. 3H-I). The expression levels of specific marker genes for the three subpopulations showed significant changes with pseudotime (Fig. 3J).

Through CellPhoneDB analysis, we explored the cell–cell interactions between mast cell subpopulations (Mast Cell 1, Mast Cell 2, and Mast Cell 3) within the CRC tumor microenvironment (Supplementary Fig. 1 A-D). While the overall interaction patterns between cells show relatively small differences (Supplementary Fig. 1 E–G), subtle variations in ligand-receptor interactions were observed among the different mast cell subpopulations. Mast Cell 1 and Mast Cell 3 mainly interact with TAM, Exhausted T cells, Treg cells, fibroblasts, and M2 macrophages, with stronger outgoing interactions. The ligand-receptor pairs identified in these interactions, such as CCL2–CCR2 and CXCL8–CXCR2, suggest pro-inflammatory and immune-suppressive roles, which could contribute to tumor immune evasion and progression (Supplementary Fig. 1 E, G). Mast Cell 2, in contrast, shows stronger interactions with M1 macrophages, cytotoxic T cells, and NK cells, particularly through CXCL10–CXCR3 and TNFSF14–LTBR. These immune-activating ligand-receptor pairs indicate that Mast Cell 2 may enhance anti-tumor immune responses by recruiting cytotoxic T cells and activating NK cells, thus inhibiting tumor progression (Supplementary Figs. 1 F).

Prognostic value of mast cell subpopulation characteristics in colorectal cancer

To further investigate the prognostic significance of the specific marker genes of the three mast cell subpopulations, we collected three bulk RNA-seq datasets: the HMU cohort, the GEO cohort, and the TCGA cohort. We constructed features based on the specific marker genes for each subpopulation and calculated standardized enrichment scores for each patient using single-sample gene set enrichment analysis (ssGSEA). Subsequently, we calculated the optimal cutoff value for each patient's standardized enrichment score and divided patients into high and low scoring groups. The clustering analysis revealed that the standardized enrichment scores of the Mast Cell 2 were mainly concentrated in normal tissue samples, while those of Mast Cell 1 and Mast Cell 3 were mainly concentrated in tumor tissue samples (Fig. 4A-C). Figure 4D-F shows the distribution of standardized enrichment scores, overall survival time, survival status, and expression patterns of Mast Cell 1, Mast Cell 2, and Mast Cell 3 subpopulations in the HMU cohort. The results showed that patients with low scores in Mast Cell 1 and 3 had lower mortality rates, while patients with high scores in Mast Cell 2 had lower mortality rates, suggesting that Mast Cell 1 and 3 characteristics may be detrimental to CRC prognosis, while Mast Cell 2 characteristics may be beneficial. Similar trends were observed in the GEO and TCGA cohorts (Supplementary Fig. 2A-F). Kaplan–Meier survival curve analysis further showed that patients with high scores in Mast Cell 1 and 3 had shorter overall survival, suggesting that their higher standardized enrichment scores may act as potential promotors of CRC (Fig. 4G, I). In contrast, patients with high scores in Mast Cell 2 had longer overall survival, indicating that their higher standardized scores may serve as potential inhibitors of CRC (Fig. 4H, Supplementary Fig. 2G-L). To assess the accuracy of the mast cell subpopulation characteristics in predicting patient prognosis, we performed ROC curve analysis, and the results showed good predictive power for all three mast cell subpopulation characteristics, particularly in the HMU cohort (Fig. 4J-L).


A
NMF Rank Survey
cophenetic
dispersion
evar
residuals
rss
silhouette
sparseness
Factorization rank
Measure type
Basis
Best fit
Coefficients
Consensus
NMFSubtype
Mast.cell1
Mast.cell2
Mast.cell3
B
C
Mast cell3
Mast cell2
Mast cell1
DNAJB1 S100A4 HMGB2 BAG3 HDAC2 EGR1 APOBEC3G PHACTR2 NDUFA12 TBL1XR1
SDC2 ZMYND15 TSPAN13 XCR1 SH3BP2 ALOX15 VMO1 MYCL LIMK1 BLOC1S1
SPIB GADD45A SEMA7A VDR PELI1 NFKBID NR4A3 DUSP4 TRAF1 GPR157
Average Expression
Percent Expressed
D
GO Enrichment
Mast cell1
pyrimidine nucleotide−sugar transmembrane transport
nucleotide−sugar transmembrane transport
serine phosphorylation of STAT protein
protein refolding
chaperone cofactor−dependent protein refolding
Mast cell2
response to hydroperoxide
response to starvation
biological process involved in symbiotic interaction
regulation of inflammatory response
positive regulation of cytokine production
Mast cell3
regulation of protein exit from endoplasmic reticulum
smooth muscle tissue development
peptidyl−tyrosine dephosphorylation
fat cell differentiation
cellular response to peptide hormone stimulus
RichFactor
P value
Count
E
KEGG Enrichment
Mast cell1
Fatty acid metabolism
PPAR signaling pathway
Antigen processing and presentation
AMPK signaling pathway
MAPK signaling pathway
Mast cell2
Virion − Human immunodeficiency virus
Hematopoietic cell lineage
Endocrine resistance
Alcoholic liver disease
Oxytocin signaling pathway
Mast cell3
Parathyroid hormone synthesis, secretion and action
Transcriptional misregulation in cancer
Efferocytosis
MAPK signaling pathway
FoxO signaling pathway
RichFactor
P value
Count
F
TF activity
ARID3A(+)
CLOCK(+)
E2F1(+)
GATA2(+)
HMGXB3(+)
FOXB1(+)
FOXO4(+)
STAT5A(+)
TAF1A(+)
ZNF140(+)
ARNT(+)
ATF5(+)
CCNT2(+)
HOXA9(+)
MYBL1(+)
G
CytoTRACE2 Score
Mast cell1
Mast cell2
Mast cell3
H
Pseudotime
Component 1
Component 2
I
NMFSubtype
Mast cell1
Mast cell2
Mast cell3
Component 1
Component 2
J
DNAJB1
SEMA7A
XCR1
Pseudo−time

◂**Fig. 3** Clustering analysis of mast cell subpopulations. **A** The relationship between the coefficients in NMF and the number of mast cell clusters (*k*). **B** Heatmap showing the differential expression of significant genes across the 3 mast cell subpopulations. **C** Bubble plot illustrating the differential expression of marker genes in the 3 mast cell subpopulations. **D** Bubble plot showing the GO terms enriched in the 3 mast cell subpopulations. **E** Bubble plot displaying the KEGG pathways enriched in the 3 mast cell subpopulations. **F** Heatmap of the transcriptional regulatory networks specific to the 3 mast cell subpopulations, revealed by SCENIC analysis, showing the activity of key transcription factors in different mast cell subpopulations. **G** CytoTRACE2 analysis revealing the differentiation potential differences among the 3 mast cell subpopulations. **H** Monocle2 pseudotime analysis showing the developmental trajectories of the 3 mast cell subpopulations. The color gradient represents the progression of pseudotime. **I** The cell subpopulation types are marked on the pseudotime trajectory. **J** The expression of the three mast cell subpopulation marker genes changes as pseudotime progresses

Gene expression of mast cell subpopulation features in colorectal cancer

To further validate the accuracy of the mast cell subpopulation characteristics, we compared the expression levels of the specific marker genes of the three mast cell subpopulations. The results showed that the specific marker genes DNAJB1 and SEMA7A of Mast Cell 1 and 3 were highly expressed in tumor tissues, while the specific marker gene XCR1 of Mast Cell 2 was lowly expressed in tumor tissues (Fig. 5A-C). We also verified the protein expression levels of these three genes. In the immunohistochemical analysis from the HPA database, DNAJB1 and SEMA7A exhibited higher protein expression levels in tumor tissues, while XCR1 had higher protein expression levels in normal tissues (Fig. 5D-F).

Based on the expression levels of these genes in each patient, we calculated the optimal cutoff values and divided patients into high and low expression groups. Kaplan–Meier survival curve analysis showed that patients with high expression of DNAJB1 and SEMA7A had shorter overall survival, suggesting that these two genes may be potential risk factors for CRC (Fig. 5G-I). In contrast, patients with high expression of XCR1 had longer overall survival, suggesting that XCR1 may be a potential protective factor for CRC (Fig. 5G-I, Supplementary Fig. 3A-F).

Discussion

This study integrates single-cell transcriptomics and bulk transcriptomics data to explore the heterogeneity of mast cell subpopulations in the CRC microenvironment and their impact on disease prognosis, revealing the complex role of mast cells in the CRC tumor microenvironment.

Mast cells, as important immune cells in the tumor microenvironment, have long been a subject of debate regarding their role in CRC. Previous studies have suggested that mast cells may have a dual role in the tumor microenvironment: on one hand, they may promote cancer by releasing histamine, proteases, chemokines, and other pro-tumor mediators, facilitating angiogenesis, tumor invasion, and metastasis; on the other hand, some studies have found that mast cells can exert anti-tumor effects by inducing endoplasmic reticulum stress in tumor cells or directly promoting tumor cell apoptosis [24]. For example, recent studies have shown that cystatin C secreted by mast cells significantly inhibits CRC progression by inducing endoplasmic reticulum stress, further supporting the potential anti-tumor role of mast cells [25]. In this study, we also found that mast cell subpopulations exhibit significant functional heterogeneity, with Mast Cell 1 and Mast Cell 3 being highly associated with pro-cancer pathways (such as STAT, MAPK, and FoxO signaling), while Mast Cell 2 is mainly involved in immune inflammation regulation and anti-tumor immune responses. This suggests that mast cells may simultaneously play both pro-tumor and anti-tumor roles in CRC, consistent with previous findings, highlighting the complexity and heterogeneity of mast cells in CRC.

Compared to previous single-cell studies, this study further investigates the transcriptomic heterogeneity of mast cells. Xie et al.'s single-cell study found that mast cells in CRC exhibit distinct activation characteristics, such as high expression of TPSAB1, CPA3, and KIT, indicating that mast cells in the CRC microenvironment are activated and associated with good prognosis [24]. Consistent with this, our study also observed an activated state of gene expression in mast cell subpopulations, but further revealed functional differences and developmental trajectory changes between the subpopulations, with Mast Cell 1 in an undifferentiated state gradually differentiating into Mast Cell 2 and Mast Cell 3. Furthermore, we systematically analyzed the transcriptional regulatory networks specific to mast cell subpopulations using SCENIC analysis, identifying key transcription factors such as CLOCK, GATA2, FOXB1, and HOXA9 with differential activity in different subpopulations, providing insights into the functional diversity of mast cells and offering potential therapeutic targets for precision immunotherapy in CRC.

Previous studies mostly used immunohistochemistry to assess mast cell density in CRC and its relationship with prognosis [26]. However, the results have often been controversial, with some studies showing that high mast cell infiltration correlates with better prognosis, while others suggest that it is closely associated with poor prognosis. These discrepancies may be due to methodological differences, heterogeneity of study cohorts, and different localization of mast cell subpopulations within the tumor tissue (differences between the tumor interior and the periphery). Our study overcame the limitations of tissue

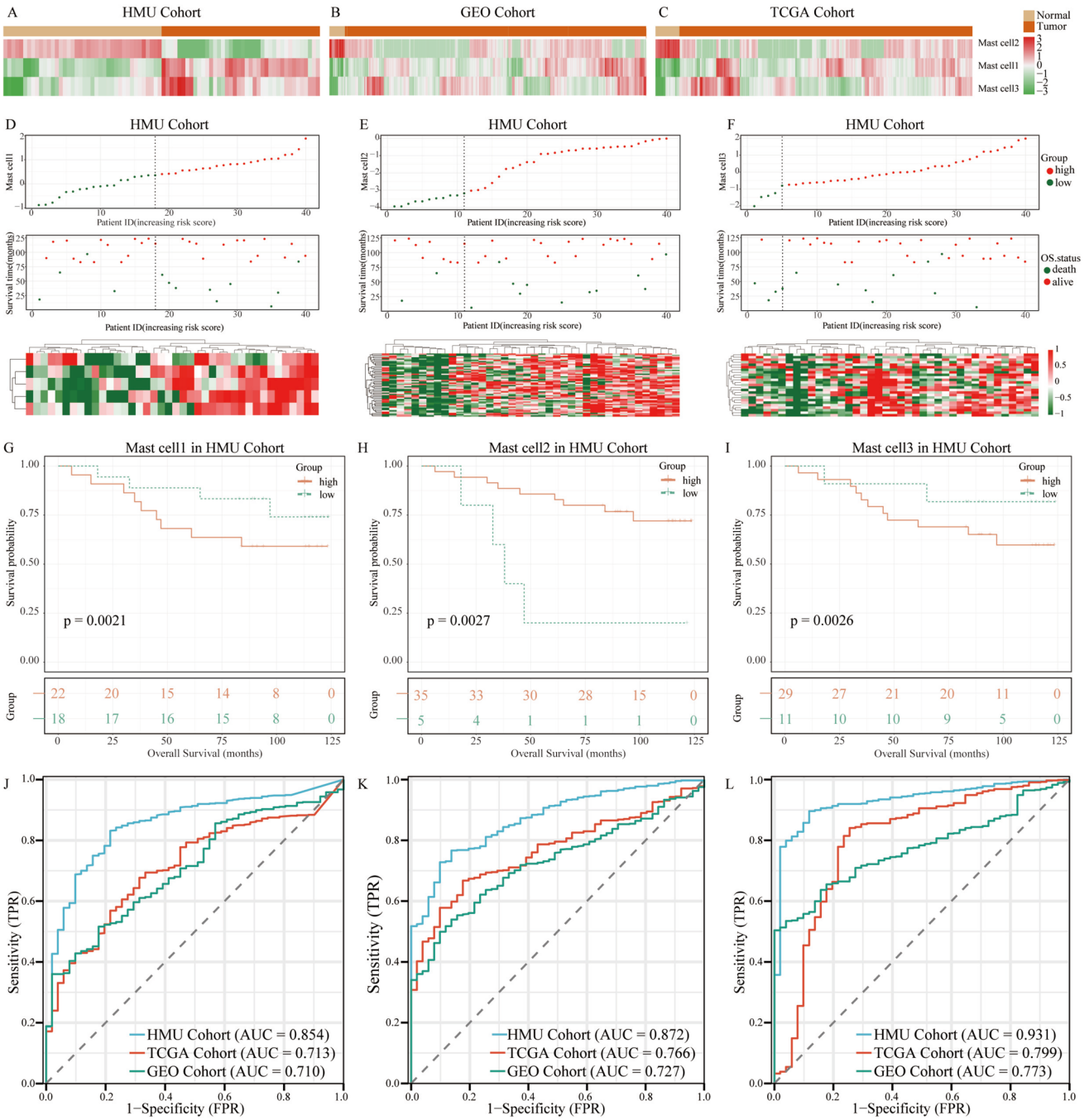


Fig. 4 Clinical significance of the 3 mast cell subpopulations. **A** Differences in the standardized enrichment scores of the 3 mast cell subpopulation signatures across different samples in the HMU cohort. **B** Differences in the standardized enrichment scores of the 3 mast cell subpopulation signatures across different samples in the GEO cohort. **C** Differences in the standardized enrichment scores of the 3 mast cell subpopulation signatures across different samples in the TCGA cohort. **D–F**. Risk factor-related plots for the standardized enrichment scores of the 3 mast cell subpopulation signatures in the HMU cohort: Top plot: Risk scores for each patient are arranged from low to high, with a vertical dashed line indicating the optimal cutoff value. This value divides patients into low-risk (green) and high-risk (red) groups. Middle plot: The relationship between risk scores and survival time, with green dots representing deceased patients and red dots representing survivors. Bottom plot: Expression patterns of the markers in patients. **G–I** Survival difference analysis based on the standardized enrichment scores of the 3 mast cell subpopulation signatures in the HMU cohort. **J–L** ROC curve analysis of the standardized enrichment scores of the 3 mast cell subpopulation signatures in the HMU, GEO, and TCGA cohorts

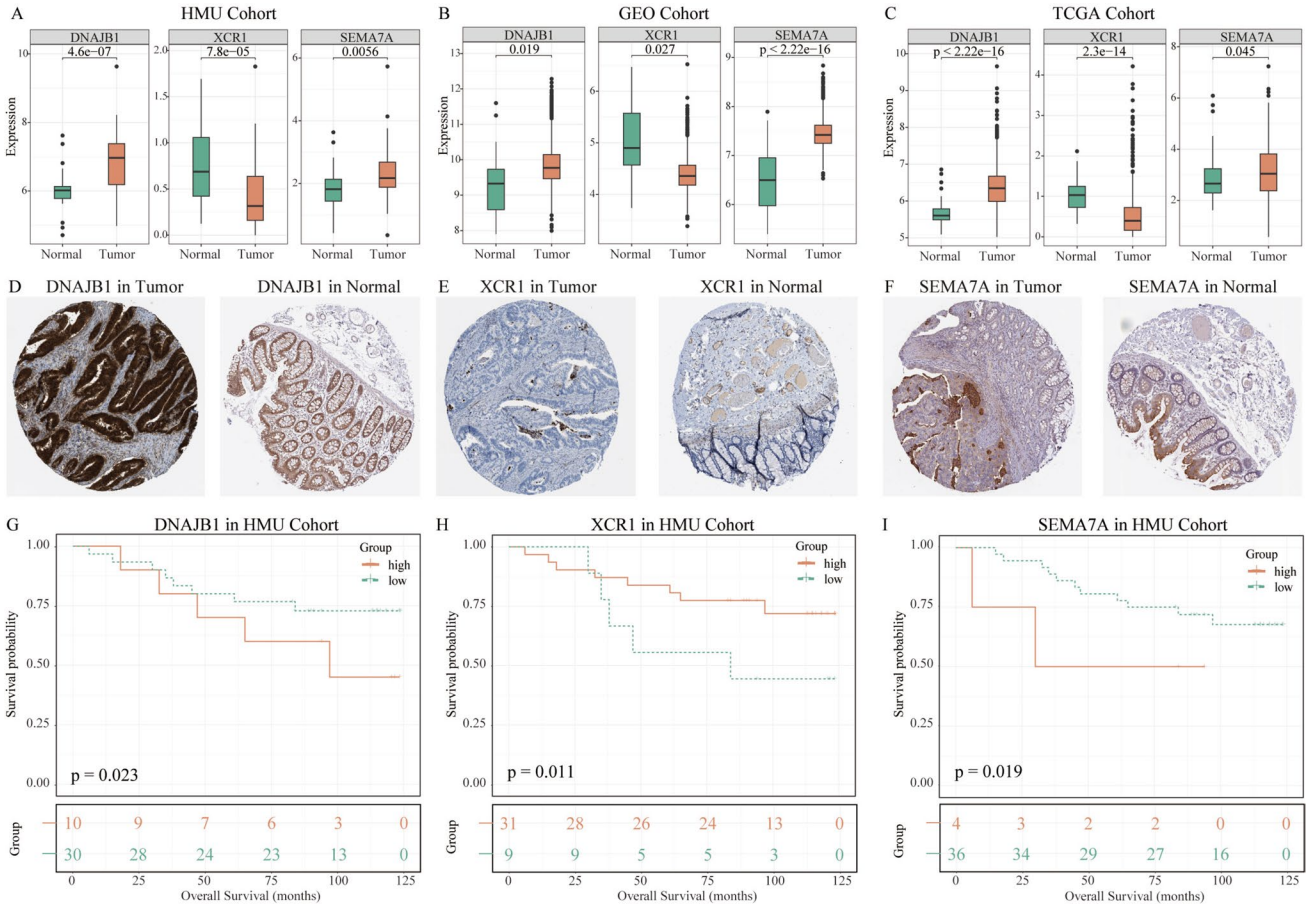


Fig. 5 Expression of marker genes for the 3 mast cell subpopulations and prognosis analysis. **A** Expression differences of the marker genes DNAJB1, XCR1, and SEMA7A in tumor and normal tissues in the HMU cohort. **B** Expression differences of the marker genes DNAJB1, XCR1, and SEMA7A in tumor and normal tissues in the GEO cohort. **C** Expression differences of the marker genes DNAJB1, XCR1, and SEMA7A in tumor and normal tissues in the TCGA cohort. **D–F**. Immunohistochemical detection of protein levels for DNAJB1, XCR1, and SEMA7A. G-I. Survival analysis based on the expression levels of DNAJB1, XCR1, and SEMA7A in the HMU cohort

localization using single-cell and bulk transcriptomics analysis, further clarifying the relationship between the specific gene expression profiles of mast cell subpopulations and patient prognosis. We have also clarified the differential impact of different mast cell subpopulations on prognosis, offering a new perspective to resolve the controversies in previous research.

In addition, this study is the first to combine CytoTrace and pseudotime analysis methods to explore the developmental relationships between mast cell subpopulations in the CRC microenvironment. This suggests that the function of mast cells may undergo dynamic changes depending on their differentiation status. This represents a significant supplement to previous research, which has mainly focused on the overall function of mast cells, overlooking the potential dynamic evolution process between subpopulations.

However, this study also has some limitations. First, the sample size in our cohort is relatively small and comes from a single medical institution, which may limit the generalizability of our findings. Larger, multi-center independent cohorts are needed for further validation. Second, while we identified mast cell subpopulation–specific marker genes from scRNA-seq data using stringent criteria, we acknowledge that these genes may not be exclusively expressed by mast cells. This is particularly relevant when applying gene signatures to bulk RNA-seq data, where the expression may reflect not only mast cells but also other cell types within the tumor microenvironment. Therefore, the prognostic associations observed should be interpreted as correlative rather than causal. Further functional and spatial validation, such as multiplex immunostaining or spatial transcriptomics, is needed to confirm the cell-type specificity and biological relevance of these gene signatures. Third, we did not perform multivariate Cox regression analysis to adjust for clinical confounders such as tumor stage, age, or treatment history. This was due to incomplete or inconsistent clinical annotation across several public

datasets, particularly from GEO. We acknowledge this as a limitation, and future studies incorporating well-annotated prospective clinical cohorts will be essential to evaluate the independent prognostic value of mast cell–associated signatures.

Conclusion

This study systematically revealed the heterogeneity of mast cell subpopulations in the CRC tumor microenvironment and their relationship with prognosis through single-cell transcriptomics, bulk transcriptomics, and bioinformatics analysis. Mast Cell 1 exhibits distinct pro-tumor characteristics, Mast Cell 3 is highly associated with cancer-related pathways, while Mast Cell 2 primarily displays anti-tumor immune regulatory properties. High expression of characteristics in Mast Cell 1 and 3 significantly predicts poor overall survival, whereas high expression of Mast Cell 2 is associated with better survival. The differential expression of mast cell subpopulation-specific marker genes (DNAJB1, SEMA7A, and XCR1) further confirms that high expression of DNAJB1 and SEMA7A and low expression of XCR1 are significantly correlated with poor prognosis. These findings provide new candidate molecules for clinical diagnosis, prognostic prediction, and the development of personalized immunotherapy targets in CRC.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00262-025-04119-8.

Acknowledgements The results here are in whole or part based upon data generated by the TCGA Research Network: https://www.cancer.gov/tcga.

Author contributions All authors reviewed the manuscript and agreed to submission. Tianyu Qiao designed the project. Chao Ding performed administrative, technical, and material support. Tianyu Qiao performed statistical analysis. Tianyu Qiao wrote the manuscript. Guiyu Wang revillsed the paper.

Funding This work was supported by the Noncommunicable Chronic Diseases-National Science and Technology Major Project (2024ZD0520300, 2024ZD0520305), Key Project of Regional Joint Fund of National Natural Science Foundation of China (U23A20482), National Nature Science Foundation of China (62276084).

Availability of data and material The datasets supporting the findings of this study are available in the from GEO and TCGA. Further inquiries can be directed to the corresponding author.

Declarations

Ethics approval and consent to participate Not applicable.

Consent for publication Not applicable.

Competing interests The authors declare no competing interests.

References

1. Siegel RL, Miller KD, Fuchs HE, Jemal A (2022) Cancer statistics, 2022. CA Cancer J Clin 72(1):7–33
2. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, Jemal A (2024) Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 74(3):229–263
3. Miller KD, Nogueira L, Devasia T, Mariotto AB, Yabroff KR, Jemal A, Kramer J, Siegel RL (2022) Cancer treatment and survivorship statistics, 2022. CA Cancer J Clin 72(5):409–436
4. Ferrone C, Dranoff G (2010) Dual roles for immunity in gastrointestinal cancers. J Clin Oncol 28(26):4045–4051
5. Paivandy A, Pejler G (2021) Novel strategies to target mast cells in disease. J Innate Immun 13(3):131–147
6. Kalesnikoff J, Galli SJ (2008) New developments in mast cell biology. Nat Immunol 9(11):1215–1223
7. Poh AR, Love CG, Masson F, Preaudet A, Tsui C, Whitehead L, Monard S, Khakham Y, Burstroem L, Lessene G *et al*: Inhibition of Hematopoietic Cell Kinase Activity Suppresses Myeloid Cell-Mediated Colon Cancer Progression. *Cancer Cell* 2017, 31(4):563–575 e565.
8. Reissfelder C, Stamova S, Gossmann C, Braun M, Bonertz A, Walliczek U, Grimm M, Rahbari NN, Koch M, Saadati M et al (2015) Tumor-specific cytotoxic T lymphocyte activity determines colorectal cancer patient prognosis. J Clin Invest 125(2):739–751
9. Bruni D, Angell HK, Galon J (2020) The immune contexture and Immunoscore in cancer prognosis and therapeutic efficacy. Nat Rev Cancer 20(11):662–680
10. Marichal T, Tsai M, Galli SJ (2013) Mast cells: potential positive and negative roles in tumor biology. Cancer Immunol Res 1(5):269–279
11. Englund A, Molin D, Enblad G, Karlen J, Glimelius I, Ljungman G, Amini RM (2016) The role of tumour-infiltrating eosinophils, mast cells and macrophages in classical and nodular lymphocyte predominant Hodgkin lymphoma in children. Eur J Haematol 97(5):430–438
12. Andersen MD, Kamper P, Nielsen PS, Bendix K, Riber-Hansen R, Steiniche T, Hamilton-Dutoit S, Clausen M, d'Amore F (2016) Tumour-associated mast cells in classical Hodgkin's lymphoma: correlation with histological subtype, other tumour-infiltrating inflammatory cell subsets and outcome. Eur J Haematol 96(3):252–259
13. Antsiferova M, Martin C, Huber M, Feyerabend TB, Forster A, Hartmann K, Rodewald HR, Hohl D, Werner S (2013) Mast cells are dispensable for normal and activin-promoted wound healing and skin carcinogenesis. J Immunol 191(12):6147–6155

14. Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 10(1):57–63
15. Dillies MA, Rau A, Aubert J, Hennequet-Antier C, Jeanmougin M, Servant N, Keime C, Marot G, Castel D, Estelle J et al (2013) A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. Brief Bioinform 14(6):671–683
16. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR (2013) STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29(1):15–21
17. Wang F, Long J, Li L, Wu ZX, Da TT, Wang XQ, Huang C, Jiang YH, Yao XQ, Ma HQ et al (2023) Single-cell and spatial transcriptome analysis reveals the cellular heterogeneity of liver metastatic colorectal cancer. Sci Adv 9(24):eadf5464
18. Vento-Tormo R, Efremova M, Botting RA, Turco MY, Vento-Tormo M, Meyer KB, Park JE, Stephenson E, Polanski K, Goncalves A et al (2018) Single-cell reconstruction of the early maternal-fetal interface in humans. Nature 563(7731):347–353
19. Aibar S, Gonzalez-Blas CB, Moerman T, Huynh-Thu VA, Imrichova H, Hulselmans G, Rambow F, Marine JC, Geurts P, Aerts J et al (2017) SCENIC: single-cell regulatory network inference and clustering. Nat Methods 14(11):1083–1086
20. Gulati GS, Sikandar SS, Wesche DJ, Manjunath A, Bharadwaj A, Berger MJ, Ilagan F, Kuo AH, Hsieh RW, Cai S et al (2020) Single-cell transcriptional diversity is a hallmark of developmental potential. Science 367(6476):405–411
21. Qiu X, Hill A, Packer J, Lin D, Ma YA, Trapnell C (2017) Single-cell mRNA quantification and differential analysis with census. Nat Methods 14(3):309–315
22. Huang H, Keathley R, Kim U, Cardenas H, Xie P, Wei J, Lengyel E, Nephew KP, Zhao G, Fu Z et al (2024) Comparative transcriptomic, epigenomic and immunological analyses identify drivers of disparity in high-grade serous ovarian cancer. NPJ Genom Med 9(1):64
23. Tang L, Peng L, Tan C, Liu H, Chen P, Wang H (2022) Role of HOXA9 in solid tumors: mechanistic insights and therapeutic potential. Cancer Cell Int 22(1):349
24. Xie Z, Niu L, Zheng G, Du K, Dai S, Li R, Dan H, Duan L, Wu H, Ren G et al (2023) Single-cell analysis unveils activation of mast cells in colorectal cancer microenvironment. Cell Biosci 13(1):217
25. Song F, Zhang Y, Chen Q, Bi D, Yang M, Lu L, Li M, Zhu H, Liu Y, Wei Q et al (2023) Mast cells inhibit colorectal cancer development by inducing ER stress through secreting cystatin C. Oncogene 42(3):209–223
26. Liu X, Li X, Wei H, Liu Y, Li N (2023) Mast cells in colorectal cancer tumour progression, angiogenesis, and lymphangiogenesis. Front Immunol 14:1209056