BMC Medical Genomics

RESEARCH

Open Access

Putative function and prognostic molecular marker of mast cells in colorectal cancer

Jiani Guo[1†], Jie Chen[2†], Yiting Wang[2†], Xiaoming Bai[3], Haimei Feng[3], Siqi Sheng[2], Hongyu Wang[3], Ke Xu[2], Mengxi Huang[2,5*], Zengjie Lei[1,2,3,4,5*] and Xiaoyuan Chu[1,2,3,4,5*]

Abstract

Background The increased demand for markers for colorectal cancer (CRC) highlights the importance of investigating immune cells involved in CRC progression. This study aims to dissect the mast cells in CRC, characterize the role of mast cells in CRC development, coordinate molecular communication between mast cells and malignant cells, and construct and validate a prognostic classification model based on mast cell markers.

Methods Single-cell transcriptome data of CRC patients were extracted from GSE146771 for cell classification and annotation. The malignant cells were identified by copykat and the communication between mast cells and malignant cells was analyzed by CellChat. Least absolute shrinkage and selection operator (LASSO) regression analysis and Cox regression analysis of mast cell markers were performed in the TCGA-COAD cohort to construct a prognostic classification model. qRT-PCR was performed to detect the mRNA expression of the molecules in the classification model in P815 and MC-9 cells. The co-culture experiment of MC38 and P815 cells were performed in 12-well transwell dish. Wound healing assay and Transwell assay were performed to detect cell migration and invasion.

Results 10,186 high-quality cells in GSE146771 were annotated to 9 cell types. Six markers in mast cells (HDC, GATA2, ASAH1, BTBD19, TIMP1, FAM110A) were selected to construct a classification model. The high-risk score defined showed high infiltration of immunosuppressive cells, including endothelial cells, CAFs, Tregs and high angiogenesis and epithelial-mesenchymal transition (EMT) activities. In the model, HDC were abnormally low expressed in P815 cells, while BTBD19, FAM110A, GATA2, ASAH1 and TIMP1 showed excessive expression in P815 cells. Knockdown of GATA2 in the co-culture system of P815 and MC38 cells blocked cell migration and invasion.

Conclusion This study identified the cell types within CRC, elaborated the cellular functions of mast cells in CRC development and their molecular communication to coordinate malignant cells, and highlighted the molecular components and biological features that constitute promising prognostic classification model.

Keywords Colorectal cancer, Mast cells, Intercellular communication, Classification model, Prognosis, Tumor microenvironment, Cells co-culture, Migration, Invasion

†Jiani Guo, Jie Chen and Yiting Wang contributed equally to this work.

*Correspondence:
Mengxi Huang
huangmengxi1@163.com
Zengjie Lei
leizengjie@163.com
Xiaoyuan Chu
chuxiaoyuan000@163.com
Full list of author information is available at the end of the article

Introduction

Colorectal cancer (CRC) is the third most diagnosed malignancy and the second leading cause of cancer death worldwide (https://www.who.int/news-room/fact-sheets/detail/colorectal-cancer). According to the data of the American Cancer Society in 2023, the incidence of cancers is 8% for both males and females, and the mortality is 8% for male cancers and 9% for female cancers, respectively [1]. Most CRC cases are sporadic, which have traditionally been viewed as a malignancy of older individuals. With demographic aging, CRC will pose a rapidly increasing challenge for numerous societies [2–4]. Moreover, CRC cases and mortality in young people have been increasing over the past few decades [5–7]. To address the burden of CRC, there is an increasing need to define the molecular, biological and microenvironmental landscape of individual patients, to develop screening strategies to facilitate its detection, and to find biomarkers for more accurate diagnosis, prognosis and treatment of CRC [8, 9].

It has become clear that mast cells, which are always present in the cancer matrix, may well contribute to the inflammatory microenvironment that forms cancer behavior [10]. Mast cells are tissue-resident effector cells derived from the hematopoietic system and represent a non-negligible immune cell population in tumor microenvironment (TME) [11]. Mast cells are considered to be distinguishable participants and coordinators of both pro-tumor and anti-tumor responses. On the one hand, they can promote different processes leading to tumor progression, such as angiogenesis, lymphangiogenesis, fibrosis and metastasis; however, mast cells can additionally release mediators capable of inducing the recruitment of other immune cells to the tumor, which can perform either pro-or anti-tumor functions [12]. The exact role of mast cells in tumor initiation and growth of CRC also remains controversial. Some studies have found that the activation of immune responses in CRC patients with low mast cell density may help prolong survival, while in patients with high mast cell density, genes and chemokines related to epithelial-mesenchymal transition (EMT) are highly expressed, suggesting a poorer prognosis for CRC patients with high mast cell density [13]. Mast cells in CRC exhibit heterogeneity, including different subtypes and activation states, which may be an important factor contributing to the diverse effects of mast cells [14] and are extremely sensitive to TME cues and represent an important tool for CRC outcome [15]. Therefore, before the determination on whether mast cells can be used as prognostic/predictive markers, it is necessary to characterize their biology in greater depth and identify specific features associated with their activation, localization, and complex relationships with tumor cells.

Due to averaging measurements of cell populations, traditional bulk transcriptome studies are limited by insufficient resolution to characterize the expression of ligands and receptors in specific cell types as well as in different cell types [16, 17]. In this study, single-cell transcriptome analysis of CRC samples from the GSE146771 dataset was performed to dissect cell types in tumors and characterize mast cell density and proximity communication between mast cells and malignant cells in the context of CRC. Then, combined with bulk RNA-seq data, the molecules that can be used as prognostic markers of CRC were selected from mast cell markers to build a classification model, and the effect of selected molecules on the fate of tumor cells was verified by cell experiments in vitro.

Materials and methods

Collection of CRC transcriptome data

TCGA database was accessed through TCGA GDC APC, and the data of different attributes of colon cancer patients in TCGA-COAD cohort were captured, including transcriptome data and clinical follow-up data. Samples with poorly documented survival information were filtered out, and the remaining 483 colon cancer samples were included in the analysis. Ensembl_ID was converted to Gene symbol and expression matrix was transformed according to log2(FPKM + 1). The GSE38832 dataset was also retrieved from the GEO database, and the microarray data and annotation information of 122 CRC samples were downloaded. The probes were matched to genes according to the annotation information.

Collection and analysis of CRC single-cell transcriptome data

The dataset GSE146771 was uploaded into the GEO database to extract single-cell transcriptome data of 10 CRC patients and analyzed using Seurat [18] in R. For quality control, 10,186 high quality cells were screened out by filtering out cells with more than 5% mitochondrial genes and 1000–7000 gene counts. The data was integrated using “Harmony” package. A SeuratObject was generated for each sample, normalized and scaled using the "NormalizeData" function and the ScaleData function. Dimensionality reduction and clustering were performed using the "RunPCA", "RunUMAP", "FindNeighbors" and "FindClusters" functions in the Seurat package. Cell clusters were aligned with CellMarker [17]-derived labeling information to identify their identity. The specific marker genes of the cell population were identified by using the

"FindAllMakers" function. The ligand-receptor interaction was analyzed by CellChat and visualized as bubble plot.

Development of a risk classifier

A 7:3 ratio was followed to divide the TCGA-COAD cohort into two mutually exclusive parts, with 70% as the training set and 30% as the validation set. Univariate Cox regression analysis was performed on the data in the training set. The genes with a threshold of $p < 0.05$ were analyzed using the R package "glmnet" [19] for Lasso Cox regression analysis, which added a penalty function to continuously compress the coefficients to achieve the purpose of simplifying the model. Multivariate Cox regression analysis was used to further evaluate the CRC prognostic association in the genes screened by LASSO, and the risk coefficients of prognostic related genes were given to obtain the risk classifier. This risk classifier was introduced into the training set and validation set of TCGA-COAD, and the GSE38832 dataset to calculate the risk score, and the "survminer" package [20] was served to determine the boundary between the high-risk group and the low-risk group. Kaplan–Meier survival analysis and Receiver Operating Characteristic (ROC) analysis were used to evaluate the efficacy of the model in predicting the prognosis of CRC.

Formulation and evaluation of nomogram

Variables associated with CRC prognosis were calculated by univariate Cox analysis. Multivariable Cox regression was fitted using covariates, namely the clinical characteristics of the samples in the TCGA-COAD cohort, thereby selected input variables for the development of the nomogram. The nomogram consisted of a point line, a risk factor line, a probability line, and a total point line. The performance of the nomogram was judged using calibration curves and DCA, both generated by R package"rms" [21].

Pathway activity analysis

The hallmark gene sets were obtained from the human Molecular Signatures Database (MSigDB), and AUCell [22] was served to quantify the AUC scores of 50 gene sets in cells. Spearman correlation analysis was adopted to calculate the association between the expression of genes in the model and AUC scores, and the pathways with $p < 0.05$ were visualized. GSEA calculated normalized enrichment score (NES) by reading and analyzing the expression data and KEGG pathway data of the high-risk and low-risk groups in the TCGA-COAD cohort, and NES > 0 indicated activation, and NES < 0 indicated inhibition.

Cells maintenance and co-culture

The mouse CRC cell line MC38 (BNCC341872), mast cell line MC-9 (VGC-0634-0000) and P815 (JNO-M0232) were purchased from BNCC (Xinyang, China), Vigenbio (Zhenjiang, China), and Jennio-Bio (Guangzhou, China), respectively. Cells were cultured and maintained according to the instructions. The co-culture experiment of MC38 and P815 cells were performed in 12-well transwell dish based on a previous study with some modifications [23]. The 1×10^6 P815 cells following the transfection of GATA2-specific siRNA or the control siRNA were seeded in the upper chamber, and 1×10^6 MC38 cells were seeded in the lower insert (0.4 μm pore size; BD Biosciences, Franklin Lakes, NJ, USA) containing DMEM/F12 serum-free medium. The medium was changed every 3 days for 14 days, and the cells co-cultured for 2 weeks were taken for subsequent analysis.

qRT-PCR was conducted to detect mRNA levels

Total RNA was extracted from cells using TRIzol reagent (15596-026, Thermo Fisher Scientific, Waltham, MA, USA) and the mRNA concentration was determined using Nanodrop (ND-2000, Thermo Fisher Scientific, USA). qRT-PCR was performed using the one-step SYBR PrimeScript RT-PCR Kit (RR055A, Takara Bio, Shiga, Japan) according to the experimental protocol instructions provided by the manufacturer. The relative expression level was quantified using the $2^{-\triangle\triangle Ct}$ method with GAPDH as the housekeeping control. The primer sequences are shown in Table 1.

Table 1 Primers used in this study

Gene	Sequences
HDC	Forward 5′-ATGAGTCCTGCCTAAATGCCC-3′
	Reverse 5′-CCTCGGAGTGAGAAGTTGTCA-3′
BTBD19	Forward 5′- CTGGTCGTGCATGGGAAAG-3′
	Reverse 5′- ATCACTGTATCGCGGGTTGTT-3′
FAM110A	Forward 5′- GTCCCTGGCTACCTGCTAC-3′
	Reverse 5′- CTGTCACACAAGTCGATGAGG –3′
GATA2	Forward 5′- CAGCAAGGCTCGTTCCTGTT-3′
	Reverse 5′- GGCTTGATGAGTGGTCGGT-3′
ASAH1	Forward 5′- AACTCGATGCTAAGCAGGGTA –3′
	Reverse 5′- GCGATCATCAAGGAAGAAGGG-3′
TIMP1	Forward 5′- CTTCTGCAATTCCGACCTCGT-3′
	Reverse 5′- ACGCTGGTATAAGGTGGTCTG-3′
GAPDH	Forward 5′- CTGGGCTACACTGAGCACC-3′
	Reverse 5′- AAGTGGTCGTTGAGGGCAATG-3′

Cell invasion analysis

The invasion assay involves the addition of an extracellular matrix on top of a porous membrane that only allows chemotaxis of tumor cells with invasive properties [24]. Here, siRNA-transfected P815 cells were co-cultured with MC38 cells and were seeded in transwell chambers (8 μm pore, BD Biosciences, USA) coated with 0.2% Matrigel (BD Biosciences, USA) for invasion assay. The lower chamber was supplemented with medium containing 10% fetal bovine serum. After 24 h, the invaded cells were fixed with paraformaldehyde and stained with crystal violet and placed under a light microscope to capture images.

Wound healing assay

Wound-healing assay was performed on MC38 cells cocultured with transfected P815 cells to assess migration ability. In short, when the monolayer of MC38 cells were fully confluent, the wound area was delineated with a 200 μL aseptic pipette. After washing with PBS, cells were maintained in serum-free medium for 48 h, and the image of wound area was captured by a light microscope, and wound closure rate was accordingly calculated.

Statistical analysis

Bioinformatics data were statistically analyzed by R software. Each cytological experiment was repeated three times independently, and the resulting data were statistically analyzed using SPSS software [25]. Cluster visualizations of the UMAP plots were drawn using the ggplot2 package. Differences in quantitative data between the two groups were tested by the t-test and wilcoxon test. And $p < 0.05$ was considered to indicate statistical significance

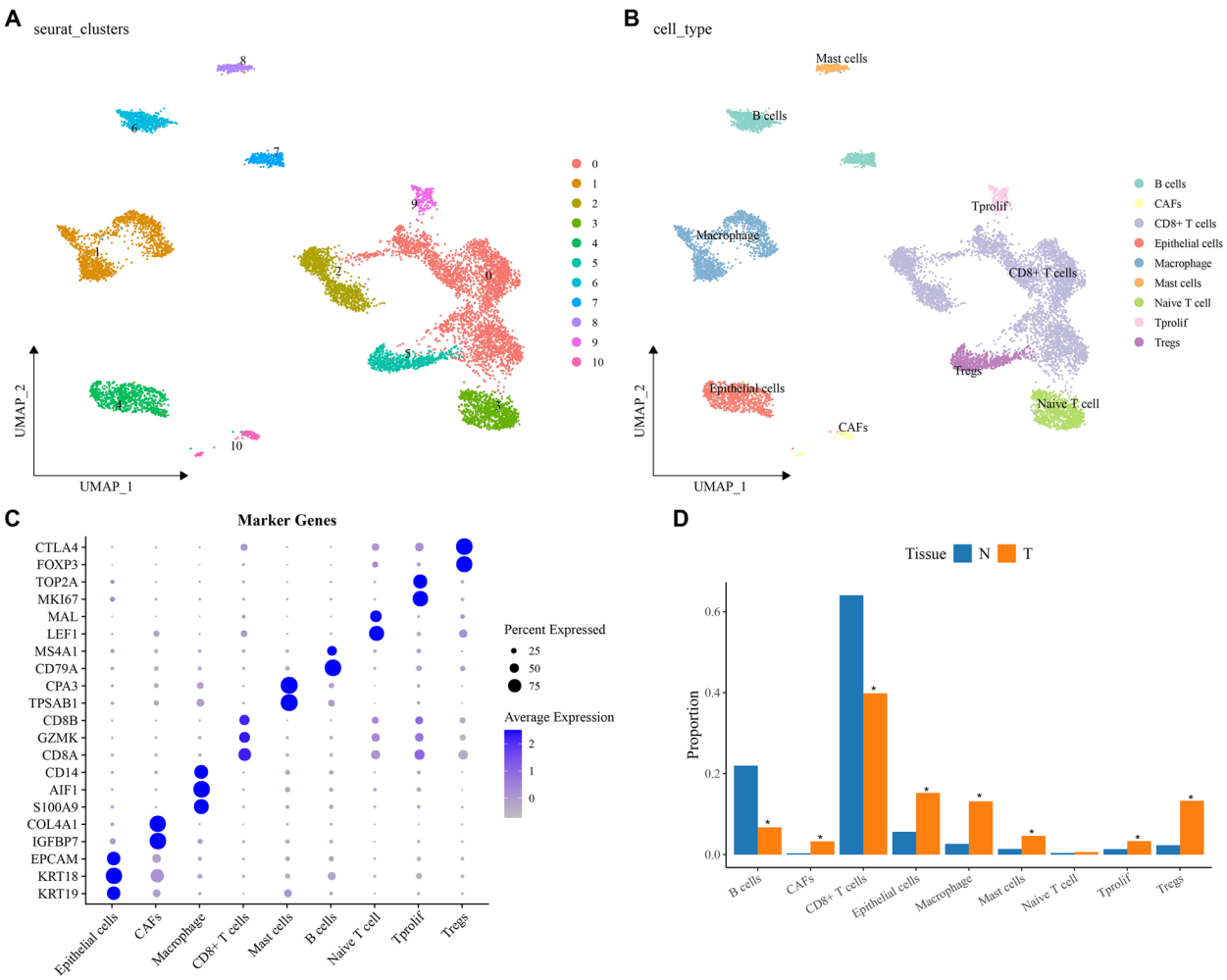

Fig. 1 Cell landscape of CRC. **A** Cluster UMAP diagram of 10,186 high-quality cells based on seurat package. **B** An UMAP diagram based on cell type annotations. **C** Mean expression levels of markers genes for each cell type displayed in the bubble plot. The depth of the blue represents the average expression level in this cell type, and the size of the dots represents the percent expressed of the gene in this cell type. **D** Proportion of different cell types in normal (blue) and CRC (orange) tumor tissues

and was represented by * in the pictures. ** was $p < 0.01$, *** was $p < 0.001$, **** was $p < 0.0001$.

Results

Cell landscape of CRC

By clustering 10,186 high-quality cells screened from GSE146771, 11 cell clusters were obtained, which were mapped to 9 cell types, namely B cells (CD79A, MS4A1), CD8 + T cells (CD8A, GZMK, CD8B), naive T cell (LEF1, MAL), regulatory T cells (Tregs) (FOXP3, CTLA4), epithelial cells (KRT19, KRT18, EPCAM), cancer-associated fibroblast (CAFs) (IGFBP7, COL4A1), macrophage (S100A9, AIF1, CD14), mast cells (TPSAB1, CPA3) and Tprolif (a type of lymphocytes with immunoregulatory functions) (MKI67, TOP2A) (Fig. 1A-C). Differentially expressed genes (DEGs) for each type of cell are shown in Supplementary Fig. 1. By comparing tumor tissue with normal tissue, it was found that the distribution of cells in CRC immune system was out of balance, Tprolif, mast cells and immunosuppressive cells CAFs, Tregs, macrophage, and epithelial cells were over-enriched, while B cells and CD8 + T-cells was insufficient (Fig. 1D). Therefore, it is speculated that CRC tumors may exhibit an immunosuppressive microenvironment.

Regulatory role of mast cells in CRC

Mast cells are also frequently observed in the TME, suggesting that they play an important role in the transition

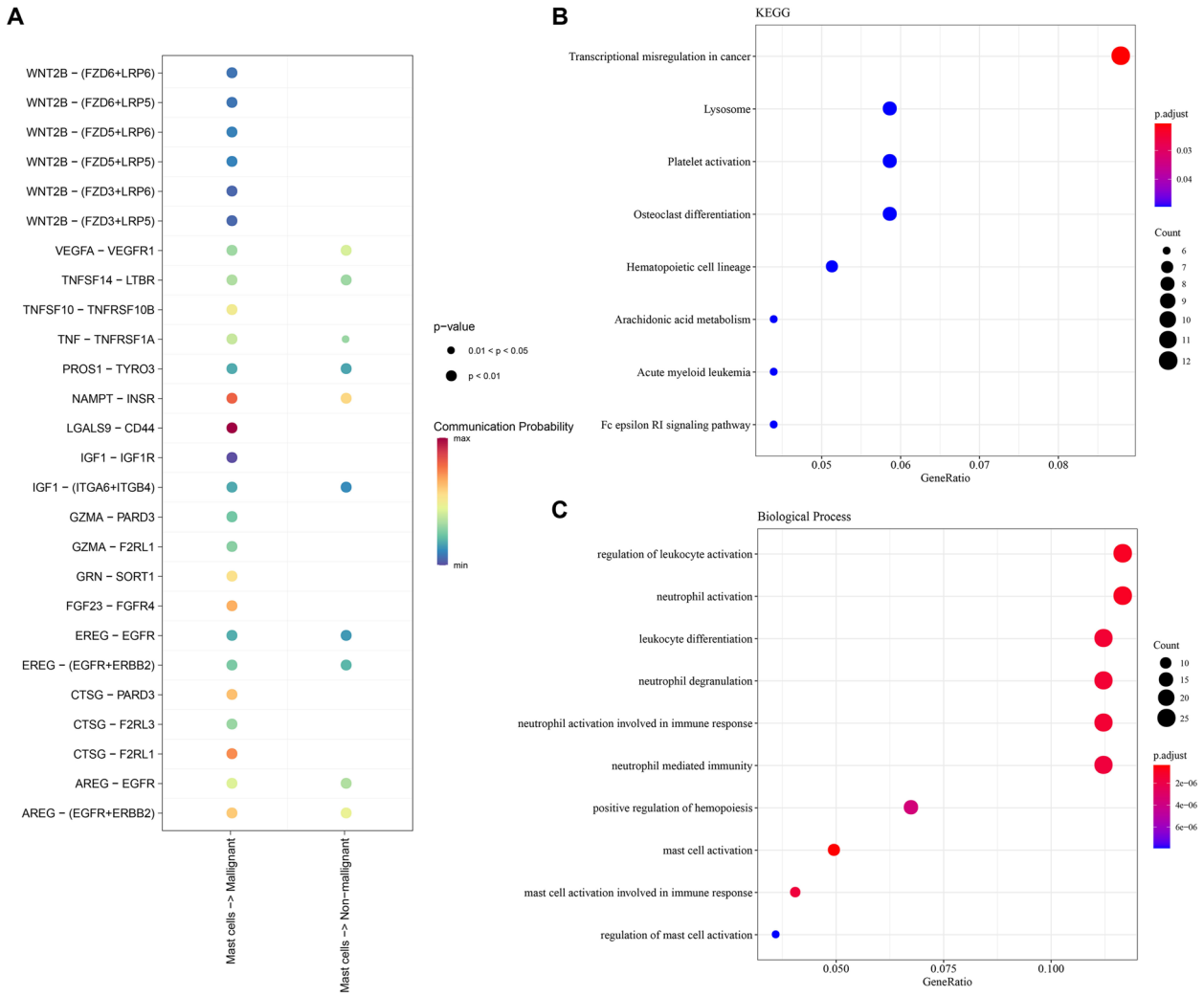

Fig. 2 Regulatory role of mast cells in CRC. **A** Ligand-receptor pairs that mediate communication between mast cells and malignant and non-malignant cells. **B** KEGG pathways annotated by mast cell specific highly expressed genes. In this figure, the size of the dots indicates the count, while the color signifies the p-value. The larger the p-value, the bluer the color; the smaller the p-value, the redder the color. And "count" represents the number of enriched genes. **C** GO biological processes annotated by mast cell specific highly expressed genes. In this figure, the size of the dots indicates the count, while the color signifies the p-value. The larger the p-value, the bluer the color; the smaller the p-value, the redder the color. And "count" represents the number of enriched genes

from chronic inflammation to cancer. It has long been recognized that patients with inflammatory bowel disease have an increased risk of colon cancer [15, 26]. Herein, the regulatory role of mast cells in CRC was explored, including the dominant ligand-receptor pair in communication with malignant cells and the regulatory pathway of mast cell marker genes. Malignant and nonmalignant cells were identified by using the copykat package. By analyzing the communication of mast cells with malignant cells and with non-malignant cells, we found that the number of ligand-receptor pairs that affect the communication of mast cells with malignant cells was higher than that of non-malignant cells. Among them, ligand-receptor pairs in WNT signal transduction, including WNT2B-(FZD6+LRP6), WNT2B-(FZD6+LRP5), WNT2B-(FZD5+LRP6),WNT2B-(FZD5+LRP5),WNT2B-(FZD3+LRP6), WNT2B-(FZD3+LRP5), immunomodulatory ligand-receptor pairs, such as TNFSF10-TNFRSF10B, LGALS9-CD44, CTSG-PARD3, CTSG-F2RL1 and GRN-SORT1, were ligand-receptor pairs particularly involved in mast cell communication with malignant cells (Fig. 2A). For the specific high expression genes of mast cell, they were significantly annotated on KEGG pathways such as transcriptional misregulation in cancer, hematopoietic cell lineage, platelet activation and immunomodulatory biological processes including activation of mast cells, leukocytes and neutrophils (Fig. 2B, C). Those analysis indicated that mast cell maybe involved in occurrence and development of CRC.

Model development and assessment

For the training and test sets divided in the TCGA-COAD cohort, there were no significant differences in clinical characteristics between the two sets, indicating the randomness and rationality of the grouping (Table 2). The most predictive molecular selection of mast cells markers was achieved by LASSO regression analysis and Cox regression analysis (Fig. 3A, B), and the classification model was formed, Risk score $= (-1.254 * HDC + 0.521 * GATA2 - 0.559 * ASAH1 + 0.707 * BTBD19 + 0.334 * TIMP1 + 0.381 * FAM110A$. The risk score of samples in each downloaded CRC sequencing set was calculated and classified according to the classification model. The prognostic classification feasibility of the model was detected in the training set, test set, unsplit TCGA-COAD dataset and GSE38832 dataset of the TCGA-COAD cohort. In terms of survival, that of the high-risk group was always significantly higher than the low-risk group at any given time point. The accuracy of the classification model for 1-year prognosis, the area under the ROC curve (AUC), was higher than 0.7 in the training set, test set, unsplit TCGA-COAD cohort and GSE38832. The 3-year AUC in the four sets were 0.71, 0.7, 0.71 and 0.64, and the 5-year AUC in the four sets were 0.72, 0.72, 0.71 and 0.64, respectively (Fig. 3C-F). The results suggest that these markers in mast cells play important roles in CRC pathology.

Table 2 Comparison of clinical features of the training and test sets in the TCGA-COAD cohort

Characteristics	Training Set(*N* = 307)	Test Set(*N* = 131)	Total(*N* = 438)	*P* value
Age				0.86
Mean ± SD	66.49 ± 12.92	66.87 ± 13.25	66.60 ± 13.00	
Median[min–max]	69.00[34.00,90.00]	68.00[31.00,90.00]	68.00[31.00,90.00]	
Gender				0.35
FEMALE	138(31.51%)	66(15.07%)	204(46.58%)	
MALE	169(38.58%)	65(14.84%)	234(53.42%)	
AJCC_stage				0.79
I	49(11.19%)	24(5.48%)	73(16.67%)	
II	114(26.03%)	53(12.10%)	167(38.13%)	
III	91(20.78%)	35(7.99%)	126(28.77%)	
IV	44(10.05%)	17(3.88%)	61(13.93%)	
NA	9(2.05%)	2(0.46%)	11(2.51%)	
Status				0.09
Alive	231(52.74%)	109(24.89%)	340(77.63%)	
Death	76(17.35%)	22(5.02%)	98(22.37%)	
OS.time				0.22
Mean ± SD	841.33 ± 708.71	992.89 ± 916.77	886.66 ± 778.81	
Median[min–max]	669.00[6.00,4126.00]	735.00[28.00,4502.00]	683.50[6.00,4502.00]	

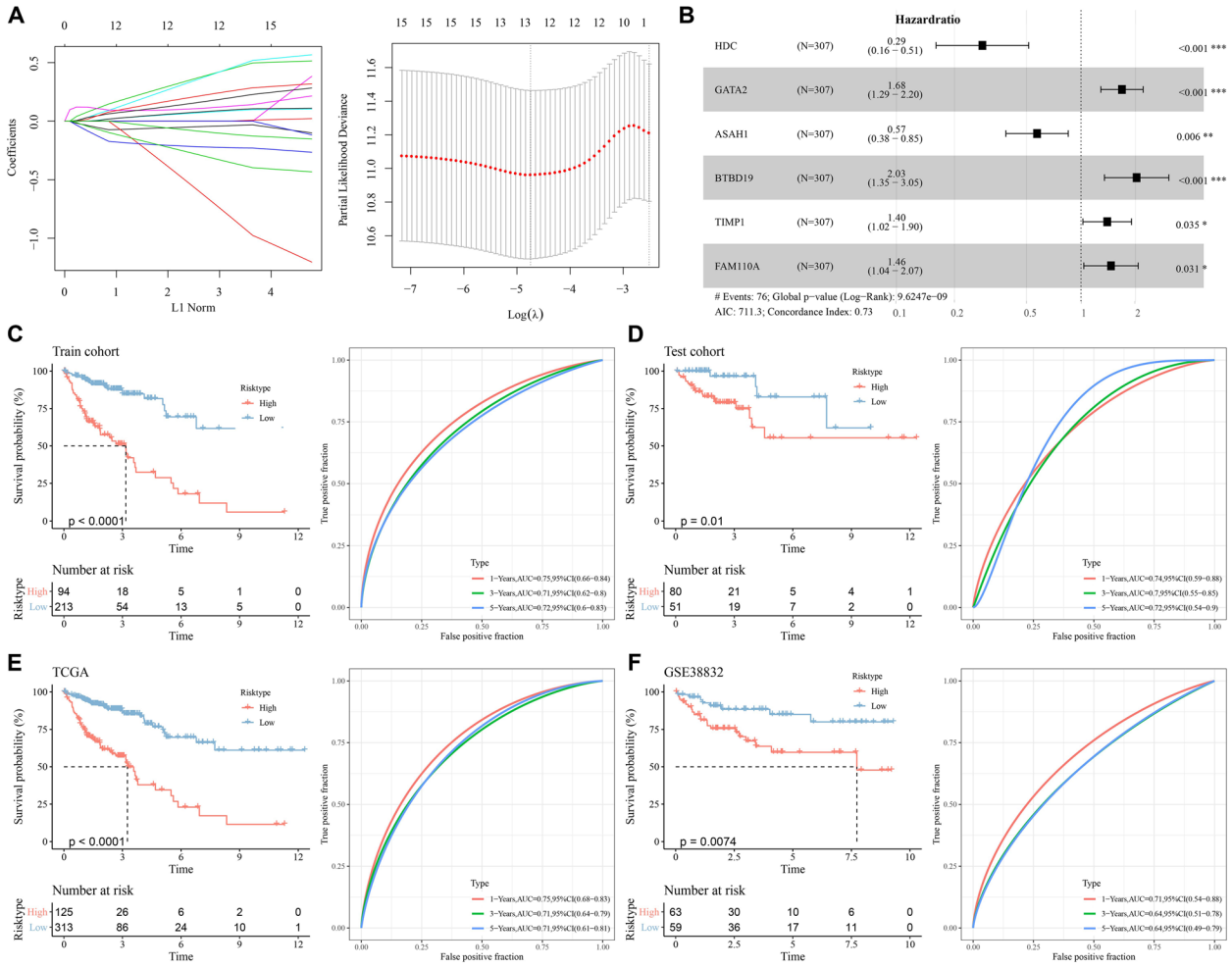

Fig. 3 Model development and assessment. **A** LASSO penalty and cross validation. **B** Multivariate Cox regression was applied to determine the effect size (hazard ratio) and statistical significance for each gene. **C**-**F** Survival curve and ROC curve generated based on prognostic classification model in training set, test set of TCGA-COAD cohort, unsplit TCGA-COAD dataset and GSE38832 dataset

Establishment of a prognostic nomogram for CRC patients

Univariate Cox regression analysis described the survival of clinical characteristic factors and risk score, age, AJCC stage and risk score were significantly relevant to the prognosis of CRC (Fig. 4A), and they still showed independent prognostic value when fitted to the multivariate Cox analysis (Fig. 4B). A nomogram integrated these independent prognostic features was plotted to predict overall survival (OS) (Fig. 4C). From the calibration curve, the 1 -, 3 -, and 5-year OS predicted by the nomogram were highly consistent with the actual OS (Fig. 4D). In addition, the decision curve analysis (DCA) curve was generated for nomogram and independent prognostic features, and the benefit of nomogram was higher than that of the three independent prognostic features (Fig. 4E). The results showed that the nomogram could systematically predict patient OS at 1, 3, and 5 years.

Immunology relevance of classification model and its molecules

In terms of immune cell infiltration, the high-risk group defined by the classification model showed a higher abundance of endothelial cells, CAFs, Tregs, and M0 macrophages infiltration and a lower abundance of NK cells, myeloid dendritic cells, than the low-risk group (Fig. 5A, $p < 0.05$). For molecules in the classification model, HDC, GATA2, ASAH1, BTBD19, and TIMP1 all showed significant positive correlations with stromal score, immune score, and ESTIMATE score, and a significant negative correlation with tumor purity (Fig. 5B, $p < 0.001$). In the high-risk group, the immune scores of T cell resting memory CD4 + T cell, activated memory CD4 + T cell, M1 macrophage infiltration were all significantly lower than those in the low-risk group (Fig. 5C, $p < 0.05$). Mast cells-related molecules in the classification model include HDC, GATA2 and

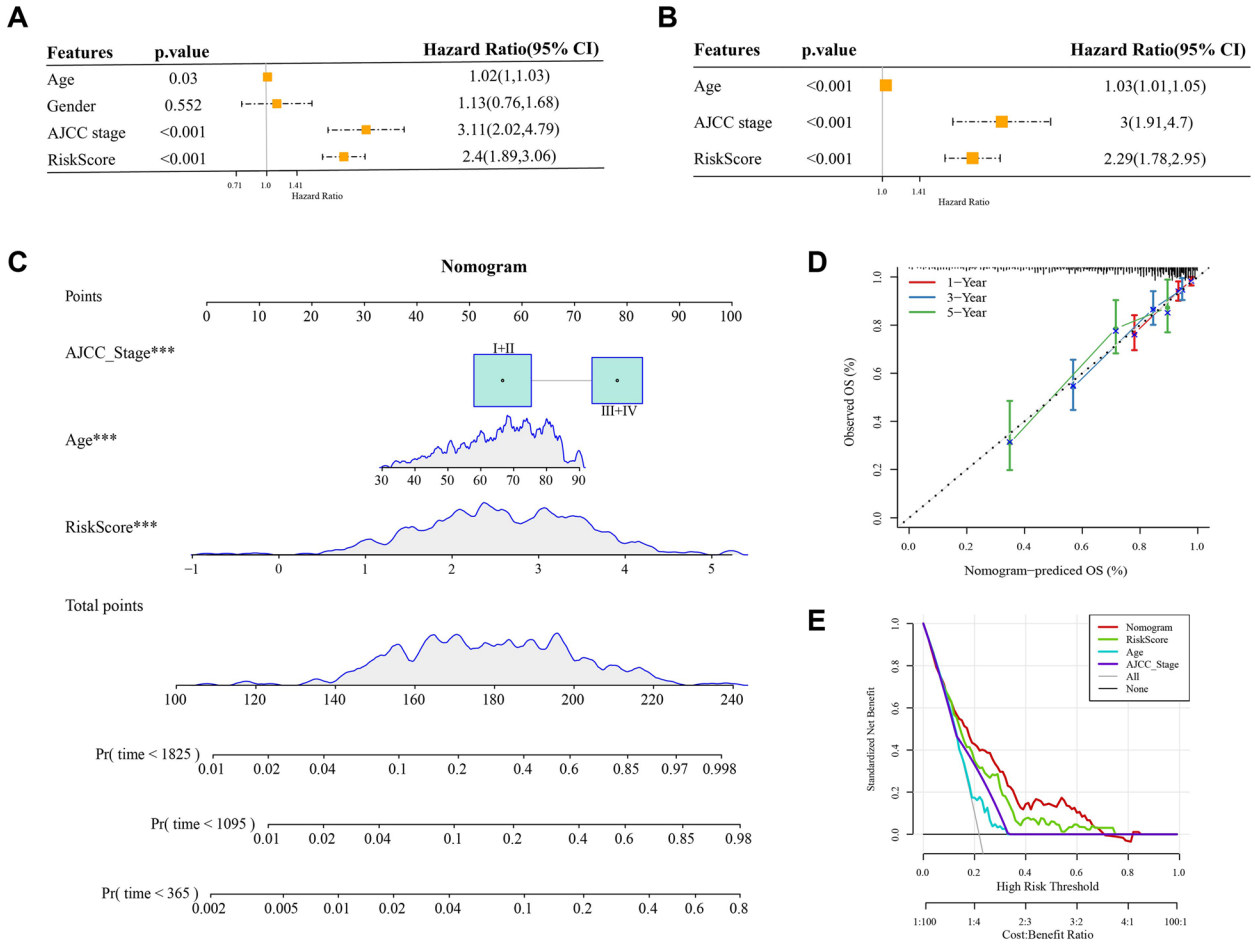

Fig. 4 The nomogram integrated independent prognostic factors of CRC to optimize the model. **A** Correlation of risk score and clinical characteristics of samples in the TCGA-COAD cohort with OS. **B** Multivariable Cox regression using covariate fit. The error bars represent the 95% confidence interval. **C** The nomogram integrating independent prognostic features. **D** Calibration curve to assess the accuracy of the nomogram. **E** Yield assessment of nomogram and independent prognostic features

BTBD19 (Fig. 5D, $p < 0.001$). It was suggested by these findings that risk score was linked with immunities since it was closely related to several critical innate immunity-related components (T cell, NK cells, and macrophages).

Classification model and signal regulation of molecules in the model

The relevance between the classification model and the molecules in it and the signaling pathways were analyzed. Both HDC and GATA2 showed significant positive correlations with immune signaling mechanisms, such as IL2 STAT signaling, IL6 JA-STAT3 signaling, and inflammatory response. HDC was also positively correlated with EMT and angiogenesis, which are essential mechanisms for cancer metastasis (Fig. 6A, B). Both ASH1 and FAM110A were significantly positively linked with complement and oxidative phosphorylation (Fig. 6C, E). TIMP also showed significant positive correlations with angiogenesis, complement, EMT, and inflammatory response (Fig. 6D). The dominant signaling pathways also showed differences between the high-risk and low-risk groups. Cancer-promoting pathways such as EMT, hypoxia and angiogenesis were significantly enriched in the high-risk group, while metabolic and immunomodulatory pathways were significantly enriched in the low-risk group (Fig. 6F). We further analyzed the correlation between genes in model and inflammatory factors such as interleukin (IL1B, IL2, IL6, IL10), chemokines, interferon-associated cytokines (IFNG), tumor necrosis factor (TNFSF10), and growth factor (TGFB1), and found

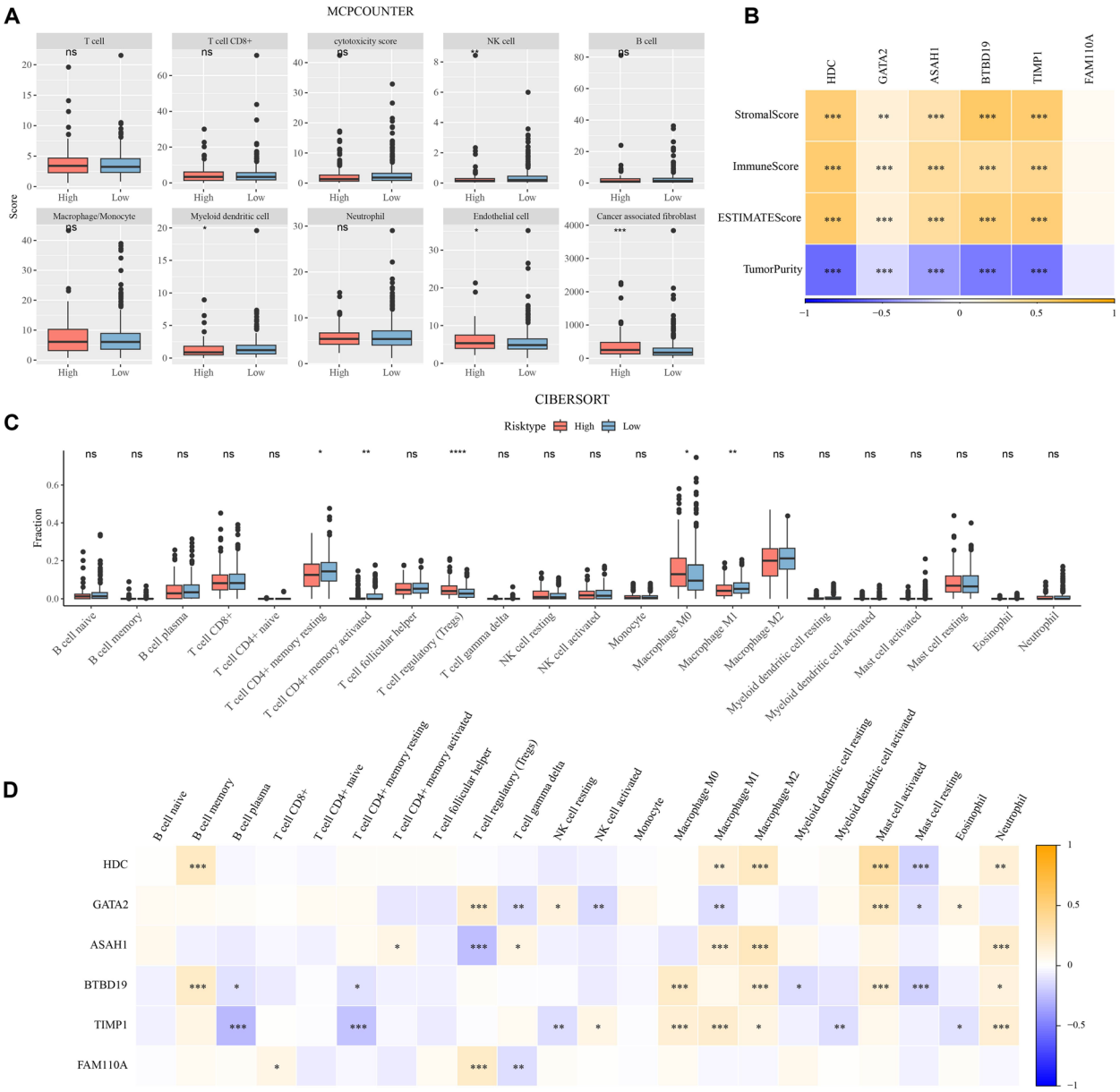

Fig. 5 Immunology relevance of classification model and its molecules. **A** The infiltration of immune cells in the high-risk and low-risk groups defined by the classification model. **B** Correlation of molecules in the classification model with stromal score, immune score, ESTIMATE score and tumor purity, respectively. **C** Differences in leukocyte infiltration between high-risk and low-risk samples. **D** Correlation matrix between genes in the classification model and leukocyte infiltration scores

that they showed a strong positive correlation with most inflammatory factors (Fig. 6G). Close associations between genes of risk score and cancer related pathways were therefore validated.

Expression and pathological function of the molecules in the classification model in CRC cells

The mRNA expression of the molecules in the classification model was examined in MC-9 and P815 cells by performing qRT-PCR. HDC in the classification model were abnormally low expressed in P815 cells, while GATA2, BTBD19, FAM110A, ASAH1 and TIMP1 showed excessive expression in P815 cells (Fig. 7A-F). When GATA2 expression was knocked down after the indirect co-culture system of MC38 and P815 was constructed, then the cell migration and invasion were measured, and the cells had a reduced spread range, as indicated by reduced would closure rate and invasion rate

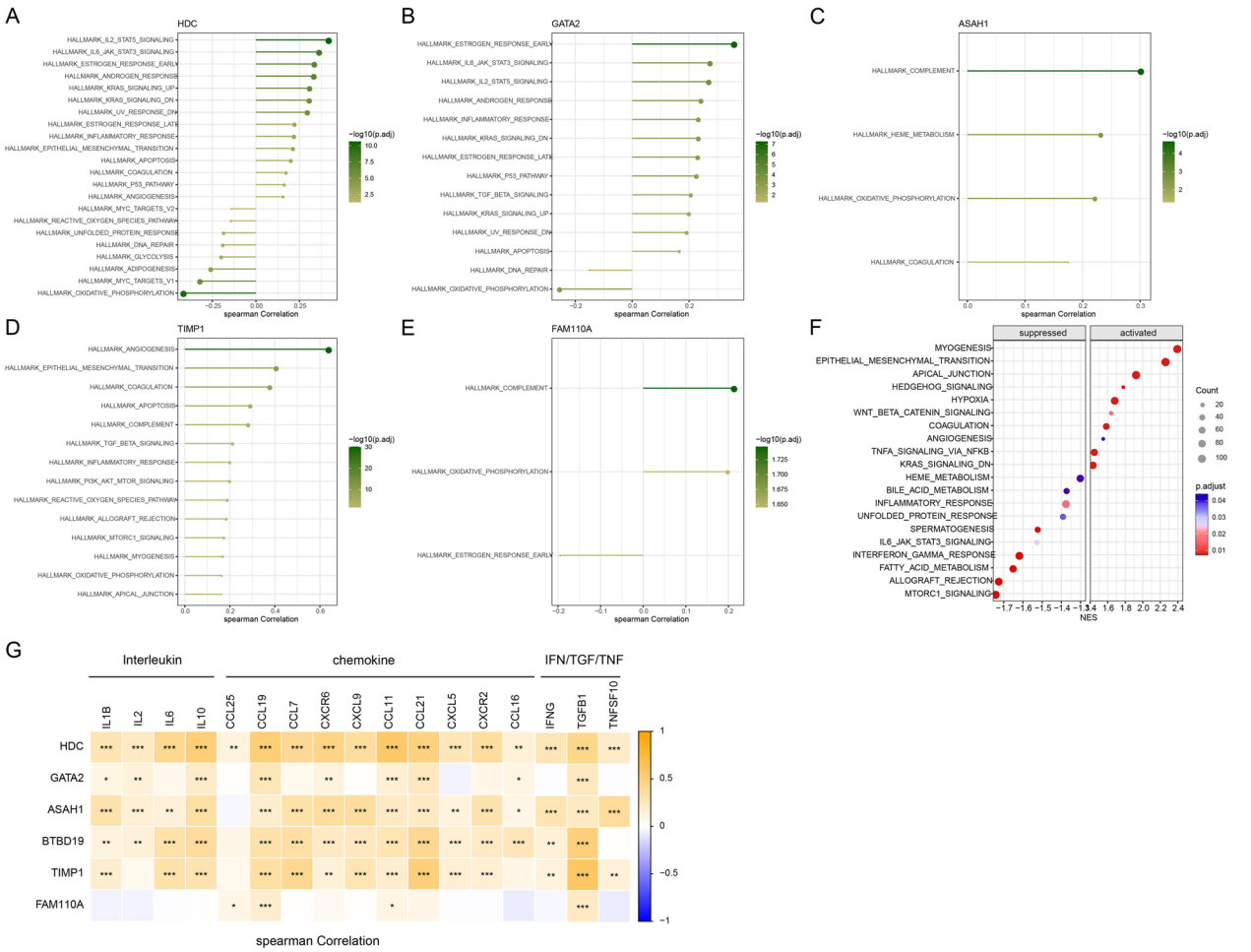

Fig. 6 Classification model and signal regulation of molecules in the model. **A**-**E** Association of HDC, GATA2, ASH1, TIMP, and FAM110A with the hallmark pathway. **F** GSEA analysis of the classification model. The "activated" pathways with scores greater than 0 are those that are activated in the high-risk group but suppressed in the low-risk group, while the "suppressed" pathways with scores less than 0 represent those that are more active in the low-risk group and suppressed in the high-risk group. **G** Association of HDC, GATA2, ASH1, TIMP, and FAM110A with the inflammatory factors. And "count" represents the number of enriched genes

(Fig. 7G-H). Those experiments indicated that mast cell influenced phenotype of CRC.

Discussion

Tumor heterogeneity is common in CRC patients. Various immune cells and inflammatory chemokines in the TME interact and influence each other to regulate tumor progression, thereby affecting tumor recurrence and treatment response, and having a great impact on the prognosis of patients [27]. Here, we dissected 9 cell types in CRC, including B cells, CD8 + T cells, naive T cell, Tprolif, Tregs, epithelial cells, CAFs, macrophage, and mast cells. In the tumor tissue, Tprolif, mast cells and CAFs, Tregs, macrophages, mast cells and epithelial cells were overdistributed, while B cells and CD8 + T cells were underrepresented. Among them, CAFs [28], Tregs, macrophages [29] and epithelial cells that have undergone EMT [30] are all immunosuppressive cells, and therefore, the CRCs we analyzed are likely to exhibit immunosuppressive microenvironment.

The number of mast cells, the so-called mast cell density (MCD), has been increased in CRC [31]. We also found excessive accumulation of mast cells in CRC. Within the tumor, mast cells interact with infiltrating immune cells, tumor cells and extracellular matrix, which is achieved through direct cell–cell interaction or the release of multiple mediators capable of remodeling the TME, and ultimately affect TME remodeling and the fate of tumor cells [32, 33]. In this study, we distinguished malignant cells in CRC and found that mediators mediating the interaction between mast cells and them included ligand-receptor pairs in WNT signal transduction and ligand-receptor pairs related to immune/inflammatory regulation. The effect of MCS

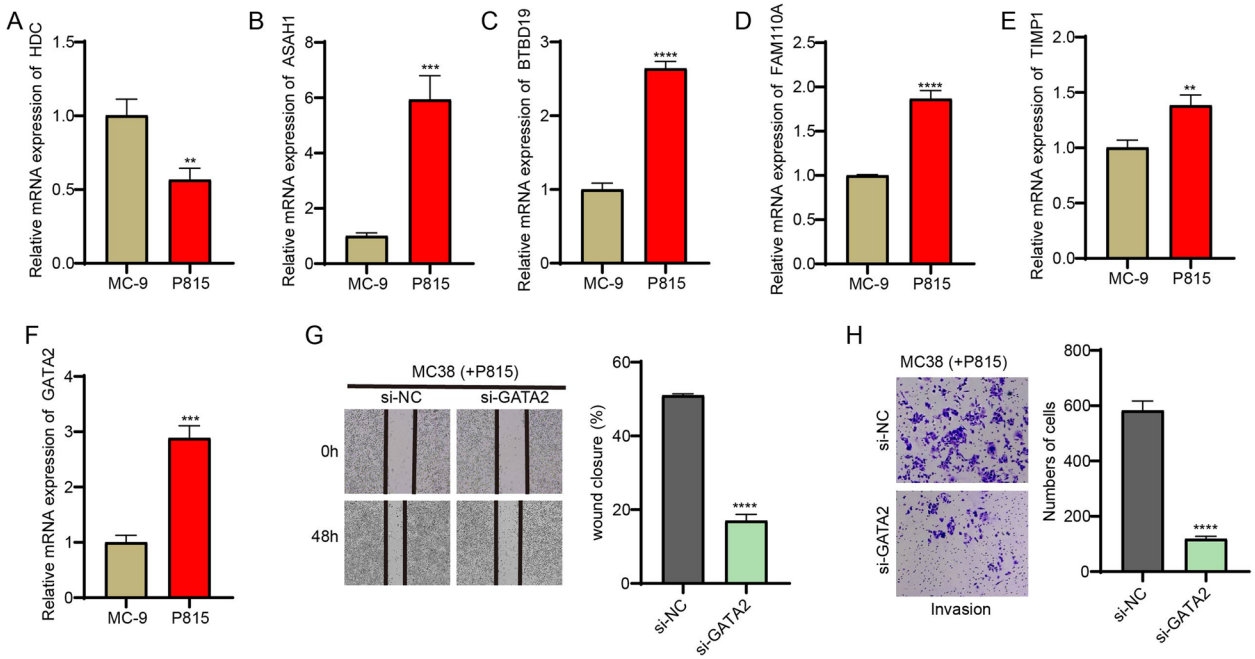

Fig. 7 Expression and pathological function of the molecules in the classification model in CRC cells**. A**-**F** The mRNA levels of molecules in the classification model detected by qRT-PCR between MC-9 cells and P815 cells, including HDC, ASAH1, BTBD19, FAM110A, TIMP1 and GATA2. **G** Wound area images and would closure rate for si-GATA2 and control groups in co-culture system of MC38 and P815 cells. **H** Cell invasion of si-GATA2 and control in co-culture system of MC38 and P815 cells

is caused by their localization, density, activation and degranulation state, secretion of cytokines and/or proteases, and proximity to other immune cells and cancer cells, as stated in previous studies [15]. We identified the cytokine TNFSF10 with its receptor TNFRSF10B [34], cathepsin G and its receptors PARD3, F2RL1, F2RL3 in the ligand-receptor pairs that mediate the interaction between mast cells and malignant cells. A study as early as 10 years ago found that LGALS9 increased the stability and function of induced Treg cells by directly binding to its receptor CD44 [35]. This also partially supports the immunomodulatory role of mast cells in CRC.

Mast cells have been recognized as independent prognostic markers of cancer for many years [36]. However, no drug targeting mast cells alone and selectively has been reported [37]. Recent studies have suggested the inclusion of molecular markers of immune cells involved in CRC progression to aid cancer prognosis and patient follow-up during treatment [38]. For this reason, we extracted mast cell markers and used bulk RNA-seq analysis to identify the biomarkers that play a role in CRC prognosis. Although this approach does not fully probe the complexity of mast cell mechanisms in CRC, it is a first step towards targeting mast cells for therapeutic CRC analysis. 6 of the mast cell markers were selected to construct a prognostic classification model, including HDC, GATA2, ASAH1, BTBD19, TIMP1, and FAM110A. In addition to identifying the prognosis of CRC patients, the classification model also identified the immunosuppressive microenvironment in the high-risk group. On a single molecule basis of the classification model, mast cells are the major source of HDC in humans, which are underexpressed and exert anticancer efficacy in experimental tumor models [39, 40]. GATA2 is highly expressed in CRC cells, and knockdown of GATA2 inhibits the proliferation, invasion, EMT and cancer stemness of CRC cells [41]. ASAH1 is overexpressed in human CRC cases, and silencing its expression leads to immune cell death and induction of mitochondrial stress [42]. Hermann's study found that systemic upregulation of TIMP1 in male pancreatic cancer patients was associated with shortened survival time and increased liver metastases, and noted that this phenomenon of high TIMP1 levels was also present in male CRC patients [43]. In this study, we observed that HDC were very low expressed in mouse tumor mast cells, while BTBD19, FAM110A, GATA2, ASAH1 and TIMP1 showed excessive expression in mouse tumor mast cells. We also knocked down GATA2 in the co-culture system formed

by CRC cells and tumor mast cells, which resulted in the suppression of cell migration and invasion.

Limitations of this study were also listed. Firstly, spatial information cannot be probed from single-cell data, which leads to a one-sided understanding of cell functions and the tumor microenvironment. A combination of sequencing methods including batch sequencing, spatial transcriptomics, and single-cell sequencing technologies may be a solution and a future direction [27, 44]. Secondly, the loss of some cell populations may have biased the results. Therefore, the observed effects might be influenced by unconsidered factors, such as the presence or absence of other immune cell types and the heterogeneity within the mast cell population. In the future, it is necessary to validate the existing findings in larger and independent cohorts, incorporate more diverse samples to eliminate the risks of sample size and selection bias, and strengthen the robustness of the conclusions. Finally, this study only verified the function of one molecule in the classification model and lacked more experimental verification, which led to the limited practical application value of the research conclusions. In the future, it's essential to apply mouse models s in vivo, observe their impact on tumor growth, metastasis and host immune responses, and obtain data via histological analysis and biomarker detection. Also, we'll closely collaborate with clinical studies. We'll collect extensive samples from diverse CRC patients, including different subtypes, stages, and genetic backgrounds, to expand marker gene validation.

In summary, this study identified 9 specific cell types within CRC, elaborated on the cellular functions and interactions with malignant cells of mast cells in CRC development, and highlighted the molecular components and biological features that constitute a promising prognostic classification model.

Abbreviations

CRC	Colorectal cancer
LASSO	Least absolute shrinkage and selection operator
TME	Tumor microenvironment
TCGA	The Cancer Genome Atlas
GEO	Gene Expression Omnibus
ROC	Receiver Operating Characteristic
AUC	Area under the curve
MSigDB	Molecular Signatures Database
GSEA	Gene Set Enrichment Analysis
NES	Normalized enrichment score
UMAP	Uniform Manifold Approximation and Projection
DCA	Decision curve analysis
CAF	Cancer-associated fibroblast
Treg	Regulatory T cell
MCD	Mast cell density
DEG	Differentially expressed gene

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12920-025-02117-4.

Supplementary Material 1

Acknowledgements
Not applicable.

Authors' contributions
XYC and ZJL designed the research. JNG and JC performed the analyses. YTW and SSS performed the experiments. XMB, HMF, KX and HYW collected and analyzed part of the data. JNG wrote the manuscript. MXH critically commented and edited the manuscript. All authors read and approved the final manuscript.

Funding
This work was supported by the National Natural Science Foundation of China 82072725 (to Xiaoyuan Chu), the Key Program of Jiangsu Provincial Health Commission ZD2021039 (to Xiaoyuan Chu), the China Postdoctoral Science Foundation 2020M670090ZX (to Zengjie Lei).

Data availability
The datasets generated and/or analyzed during the current study are available in the [GSE146771] repository, [https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc= GSE146771] and [GSE38832] repository, [https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc= GSE38832].

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Department of Medical Oncology, Jinling Hospital, Nanjing Medical University, Nanjing, Jiangsu Province, China. [2]Department of Medical Oncology, Affiliated Hospital of Medical School, Nanjing Jinling Hospital, Nanjing University, Nanjing, Jiangsu Province, China. [3]Department of Medical Oncology, Jinling Hospital, Nanjing University of Chinese Medicine, Nanjing, Jiangsu Province, China. [4]Department of Medical Oncology, the First School of Clinical Medicine, Jinling Hospital, Southern Medical University, Nanjing, Jiangsu Province, China. [5]305 Zhongshan East Road, Xuanwu District, Nanjing, Jiangsu Province 210000, China.

Received: 15 April 2024 Accepted: 27 February 2025
Published online: 08 April 2025

References

1. Siegel RL, Miller KD, Wagle NS, Jemal A. Cancer statistics, 2023. CA Cancer J Clin. 2023;73(1):17–48.
2. Klimeck L, Heisser T, Hoffmeister M, Brenner H. Colorectal cancer: a health and economic problem. Best Pract Res Clin Gastroenterol. 2023;66:101839.
3. Salleh EA, Lee YY, Zakaria AD, Che Jalil NA, Musa M. Cancer-associated fibroblasts of colorectal cancer: translational prospects in liquid biopsy and targeted therapy. Biocell. 2023;47(10):2233–44.

4. Ruan Y, Lu G, Yu Y, Luo Y, Wu H, Shen Y, et al. PF-04449913 Inhibits Proliferation and Metastasis of Colorectal Cancer Cells by Down-regulating MMP9 Expression through the ERK/p65 Pathway. Curr Mol Pharmacol. 2023. https://doi.org/10.2174/1874467217666230915125622.
5. Eng C, Jacome AA, Agarwal R, Hayat MH, Byndloss MX, Holowatyj AN, et al. A comprehensive framework for early-onset colorectal cancer research. Lancet Oncol. 2022;23(3):e116–28.
6. Lifei L, Zhang J, Li X, Zhu Y, Wang Y, Liu D. Sericic acid ameliorates dss-induced ulcerative colitis in mice by modulating the NF-κB and Nrf2 Pathways. Curr Mol Pharmacol. 2023;16(7):759–70.
7. Bou Malhab LJ, Abdel-Rahman WM. Obesity and inflammation: colorectal cancer engines. Curr Mol Pharmacol. 2022;15(4):620–46.
8. Andrei P, Battuello P, Grasso G, Rovera E, Tesio N, Bardelli A. Integrated approaches for precision oncology in colorectal cancer: The more you know, the better. Semin Cancer Biol. 2022;84:199–213.
9. Romero-Garmendia I, Garcia-Etxebarria K. From omic layers to personalized medicine in colorectal cancer: the road ahead. Genes (Basel). 2023;14(7):1430.
10. Aponte-López A, Fuentes-Pananá EM, Cortes-Muñoz D, Muñoz-Cruz S. Mast cell, the neglected member of the tumor microenvironment: role in breast cancer. J Immunol Res. 2018;2018:2584243.
11. Derakhshan T, Boyce JA, Dwyer DF. Defining mast cell differentiation and heterogeneity through single-cell transcriptomics analysis. J Allergy Clin Immunol. 2022;150(4):739–47.
12. Segura-Villalobos D, Ramirez-Moreno IG, Martinez-Aguilar M, Ibarra-Sanchez A, Munoz-Bello JO, Anaya-Rubio I, et al. Mast cell-tumor interactions: molecular mechanisms of recruitment, intratumoral communication and potential therapeutic targets for tumor growth. Cells. 2022;11(3):349.
13. Mao Y, Feng Q, Zheng P, Yang L, Zhu D, Chang W, et al. Low tumor infiltrating mast cell density confers prognostic benefit and reflects immunoactivation in colorectal cancer. International journal of cancer. 2018;143(9):2271–80.
14. Xie Z, Niu L, Zheng G, Du K, Dai S, Li R, et al. Single-cell analysis unveils activation of mast cells in colorectal cancer microenvironment. Cell Biosci. 2023;13(1):217.
15. Molfetta R, Paolini R. The controversial role of intestinal mast cells in colon cancer. Cells. 2023;12(3):459.
16. Ma F, Wang S, Xu L, Huang W, Shi G, Sun Z, et al. Single-cell profiling of the microenvironment in human bone metastatic renal cell carcinoma. Commun Biol. 2024;7(1):91.
17. Zulibiya A, Wen J, Yu H, Chen X, Xu L, Ma X, et al. Single-Cell RNA sequencing reveals potential for endothelial-to-mesenchymal transition in tetralogy of fallot. Congenit Heart Dis. 2023;18(6):611–25.
18. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat Biotechnol. 2018;36(5):411–20.
19. Engebretsen S, Bohlin J. Statistical predictions with glmnet. Clinical epigenetics. 2019;11(1):123.
20. Liang K, Guo Z, Zhang S, Chen D, Zou R, Weng Y, Peng C, Xu Z, Zhang J, Liu X, Pang X, Ji Y, Liao D, Lai M, Peng H, Ke Y, Wang Z, Wang Y. GPR37 expression as a prognostic marker in gliomas: a bioinformatics-based analysis. Aging. 2023;15(19):10146–67.
21. Eng KH, Schiller E, Morrell K. On representing the prognostic value of continuous gene expression biomarkers with the restricted mean survival curve. Oncotarget. 2015;6(34):36308–18.
22. Van de Sande B, Flerin C, Davie K, De Waegeneer M, Hulselmans G, Aibar S, et al. A scalable SCENIC workflow for single-cell gene regulatory network analysis. Nat Protoc. 2020;15(7):2247–76.
23. Sulsenti R, Scialpi GB, Frossi B, Botti L, Ferri R, Tripodi I, Piva A, Sangaletti S, Pernici D, Cancila V, Romeo F, Chiodoni C, Lecis D, Bianchi F, Fischetti I, Enriquez C, Crivelli F, Bregni M, Renne G, Pece S, Tripodo C, Pucillo CE, Colombo MP, Jachetti E. Intracellular Osteopontin Promotes the Release of TNFα by Mast Cells to Restrain Neuroendocrine Prostate Cancer. Cancer Immunol Res. 2024;12(9):1147-1169. https://doi.org/10.1158/2326-6066.CIR-23-0792.
24. Justus CR, Marie MA, Sanderlin EJ, Yang LV. Transwell in vitro cell migration and invasion assays. Methods Mol Biol. 2023;2644:349–59.
25. George D, Mallery P. IBM SPSS statistics 26 step by step: A simple guide and reference. Routledge; 2019.
26. Rigoni A, Bongiovanni L, Burocchi A, Sangaletti S, Danelli L, Guarnotta C, et al. Mast cells infiltrating inflamed or transformed gut alternatively sustain mucosal healing or tumor growth. Can Res. 2015;75(18):3760–70.
27. Wen R, Zhou L, Peng Z, Fan H, Zhang T, Jia H, et al. Single-cell sequencing technology in colorectal cancer: a new technology to disclose the tumor heterogeneity and target precise treatment. Front Immunol. 2023;14:1175343.
28. Chen Y, McAndrews KM, Kalluri R. Clinical and therapeutic relevance of cancer-associated fibroblasts. Nat Rev Clin Oncol. 2021;18(12):792–804.
29. Tie Y, Tang F, Wei YQ, Wei XW. Immunosuppressive cells in cancer: mechanisms and potential therapeutic targets. J Hematol Oncol. 2022;15(1):61.
30. Taki M, Abiko K, Ukita M, Murakami R, Yamanoi K, Yamaguchi K, et al. Tumor Immune Microenvironment during Epithelial-Mesenchymal Transition. Clin Cancer Res. 2021;27(17):4669–79.
31. Marech I, Ammendola M, Gadaleta C, Zizzo N, Oakley C, Gadaleta CD, et al. Possible biological and translational significance of mast cells density in colorectal cancer. World J Gastroenterol. 2014;20(27):8910–20.
32. Komi DEA, Redegeld FA. Role of Mast Cells in Shaping the Tumor Microenvironment. Clin Rev Allergy Immunol. 2020;58(3):313–25.
33. Shi S, Ye L, Yu X, Jin K, Wu W. Focus on mast cells in the tumor microenvironment: Current knowledge and future directions. Biochim Biophys Acta Rev Cancer. 2023;1878(1):188845.
34. Burgaletto C, Platania CBM, Di Benedetto G, Munafo A, Giurdanella G, Federico C, et al. Targeting the miRNA-155/TNFSF10 network restrains inflammatory response in the retina in a mouse model of Alzheimer's disease. Cell Death Dis. 2021;12(10):905.
35. Wu C, Thalhamer T, Franca RF, Xiao S, Wang C, Hotta C, et al. Galectin-9-CD44 interaction enhances stability and function of adaptive regulatory T cells. Immunity. 2014;41(2):270–82.
36. Johansson A, Rudolfsson S, Hammarsten P, Halin S, Pietras K, Jones J, et al. Mast cells are novel independent prognostic markers in prostate cancer and represent a target for therapy. Am J Pathol. 2010;177(2):1031–41.
37. Tsukamura M, Ichiyama S. Mycobacterium avium complex bacteria probably live longer than other slowly growing mycobacteria. Kekkaku. 1988;63(9):597–600.
38. Roliano GG, Azambuja JH, Brunetto VT, Butterfield HE, Kalil AN, Braganhol E. Colorectal Cancer and Purinergic Signalling: An Overview. Cancers (Basel). 2022;14(19). https://doi.org/10.3390/cancers14194887.
39. Francis T, Graf A, Hodges K, Kennedy L, Hargrove L, Price M, et al. Histamine regulation of pancreatitis and pancreatic cancer: a review of recent findings. Hepatobiliary Surg Nutr. 2013;2(4):216–26.
40. Grauers Wiktorin H, Nilsson MS, Kiffin R, Sander FE, Lenox B, Rydstrom A, et al. Histamine targets myeloid-derived suppressor cells and improves the anti-tumor efficacy of PD-1/PD-L1 checkpoint blockade. Cancer Immunol Immunother. 2019;68(2):163–74.
41. Pan Y, Zhu Y, Zhang J, Jin L, Cao P. A feedback loop between GATA2-AS1 and GATA2 promotes colorectal cancer cell proliferation, invasion, epithelial-mesenchymal transition and stemness via recruiting DDX3X. J Transl Med. 2022;20(1):287.
42. Vijayan Y, James S, Viswanathan A, Aparna JS, Bindu A, Namitha NN, et al. Targeting acid ceramidase enhances antitumor immune response in colorectal cancer. J Adv Res. 2023. https://doi.org/10.1016/j.jare.2023.12.013.
43. Hermann CD, Schoeps B, Eckfeld C, Munkhbaatar E, Kniep L, Prokopchuk O, et al. TIMP1 expression underlies sex disparity in liver metastasis and survival in pancreatic cancer. J Exp Med. 2021;218(11):e20210911.
44. Xie F, Xi N, Han Z, Luo W, Shen J, Luo J, et al. Progress in research on tumor microenvironment-based spatial omics technologies. Oncol Res. 2023;31(6):877–85.

Publisher's Note