BMC Gastroenterology

RESEARCH

Open Access

Identification of the "Collagen-Macrophage" sub-category of patients with colorectal cancer as an extension of the CMS4 subtype with THBS2 as a therapeutic target

Shuwen Chen[1,2†], Zhaoyan Jiang[1†], Wanxuan Song[2†], Chuqiao Lu[2], Yanbing Lin[3], Shiyao Xu[2], Kunxin Xie[5], Li Wan[4*] and Xiaoqin Yuan[1*]

Abstract

We identified a subset of patients with colorectal cancer (CRC) enriched with "collagen-TAMs," designated the CM class, using large integrated colon cancer transcriptome and single-cell transcriptome datasets. This classification system could be used as an extension of the traditional CMS classification system for CRC to guide more accurate classification and treatment.

We also screened CAF-derived *THBS2* as a potential biomarker for CM and found that it plays an important role in CRC disease models in vitro and in vivo, promoting tumor development and metastasis as well as TAM recruitment. Targeting *THBS2* combined with PD-1 therapy effectively improved the therapeutic effect of immunotherapy in vivo. The CM classification provides a new perspective for CRC treatment, and *THBS2*, which is highly expressed in CM cases, can be used as a new potential combined target for immunotherapy.

Keywords Colorectal cancer, Immunotherapy, Classification system, THBS2

Introduction

Colorectal cancer (CRC) is the third most common malignant cancer globally, with consistently high annual morbidity and mortality [1]. Its malignant progression (including the epithelial-mesenchymal transition of tumor cells [2, 3] and stemness enhancement [4]) and the transition of the tumor microenvironment (an "inflammatory tumor microenvironment [5]" or "immunosuppressive tumor microenvironment [6]") can affect tumor fate and patient prognosis. Heterogeneity is highly present in patients with CRC [7], leading to high variation in the effectiveness of therapies that target programmed T cell death (the PD-1/PD-L1 neutralization strategy) [8–10]. Identification of the reason behind this failure of some patients to respond to treatment is urgently needed.

†Shuwen Chen, Zhaoyan Jiang and Wanxuan Song contributed equally to this work.

*Correspondence:
Li Wan
doctorwl@njmu.edu.cn
Xiaoqin Yuan
yuanxq@njmu.edu.cn
[1] Department of Anatomy, Histology and Embryology, Nanjing Medical University, Nanjing 211166, China
[2] Department of Clinical Medicine, First Clinical Medicine College, Nanjing Medical University, Nanjing 211166, China
[3] Research Center for Environment and Female Reproductive Health, the Eighth Affiliated Hospital, Sun Yat-Sen University, Shenzhen 518033, China
[4] Department of Oncology, The Affiliated Huaian No.1 People's Hospital of Nanjing Medical University, Huai'an 223302, China
[5] Department of Biochemistry and Molecular Biology, Key Laboratory of Human Functional Genomics of Jiangsu Province, Nanjing Medical University, Nanjing 211166, China

The heterogeneous disease landscape of CRC leads to different patient responses to drugs. To address inconsistencies between gene expression-based CRC classifications and facilitate clinical conversion, previous investigators have proposed a consensus molecular subtype [11]. Using this system, CMS4 CRC with mesenchymatization was identified as having significant fibroblast proliferation, accompanied by TGFβ signal activation, angiogenesis, and matrix invasion [12, 13]; these patients are prone to liver metastasis, resulting in poor prognosis, have poor immune cell infiltration, and do not respond well to immunotherapy [13, 14].

In CRC, the tumor boundaries are mostly surrounded by the interstitial region [15], where recruitment of immunosuppressive cells and activation of cancer-associated fibroblasts occur [16–18]. Because of dimensionality reduction, traditional high-throughput sequencing (including bulk RNA and single-cell RNA sequencing) eliminates the spatiotemporal information inherent to the tumor itself, making it difficult to determine the tumor boundary [19, 20]. With the rise of spatial transcriptome sequencing technologies, in situ sequencing (ISS)-based methods have been able to identify biological information from the tumor boundary region [21], including Liu's identification of the tumor immune barrier (TIB) in the HCC microenvironment [22] and Ye's research on the interaction of FAP(+) fibroblasts and SPP1(+) macrophages in CRC [23]. However, how this immunosuppressive tumor boundary forms and the intrinsic molecular mechanisms and dynamic migration processes behind cell states remain unknown.

In our study, we identified the existence of a "collagen-macrophage" subclass of patients with CRC, mainly composed of traditional CMS4 patients but also including some patients from other CMS categories. These patients showed recruitment of CAFs and SPP1 + macrophages, as well as a decrease in toxic immune cells, possibly contributing to their poor response to immunotherapy. Multi-cohort single-cell sequencing data showed that the TSP2 protein secreted by CAFs in CM CRC was significantly higher than that in non-CM CRC. In vivo, TSP2 promoted the recruitment of SPP1 + macrophages,and blocking both PD1 and TSP2 or TSP2 alone showed strong anti-tumor effects while enhancing the efficacy of conventional immunotherapy. Removal of the CAF-TAM system at the tumor edge of CM CRC effectively improved the effect of PD1 drug use. Our study makes up for the deficiencies of CMS classification in clinical practice and proposes effective molecular targets, providing a new perspective for the combined application of colon cancer immunotherapy.

Results

Spatially enriched SPP1$^+$ macrophages and FAP$^+$ fibroblasts at the edge of the tumor mass contribute to poor prognosis in patients with CRC

Many previous single-cell sequencing studies have focused on the comparison between tumors and paratumors; none have paid attention to the cell composition and signal transduction in the tumor margin region, or its impact on patient prognosis and clinical treatment. We integrated two cohorts with clear sampling information (EMTA-B8107 [24] and GSE144735 [25] cohorts) (Table 1) and obtained a total of 13 paratumor, tumor margin, and non-margin tumor samples. After quality control and dimensionality reduction, 61,698 valid cells were obtained, and cell types were annotated according to previous studies [26] (Fig. 1A). Following the results obtained by a previous study, we screened the genes in the TCGA-CRC cohort that contribute to poor prognosis; 2,529 genes were identified and used for subsequent analysis. After we used MAGIC to interpolate the expression values of these genes and calculate the sum value of their expression in a single cell [27], we constructed a risk landscape at the single-cell level to more easily observe which types of tumor or microenvironmental cells contribute to poor prognosis in patients with CRC (Fig. 1B).

According to the classification of epithelium, stroma, and immunity, myeloid cells were the most important immune cells causing poor prognosis in patients, which was consistent with many previous reports (Fig. 1C), while fibroblasts were identified as the most important stromal cells responsible for poor patient outcomes (Fig. 1D). This co-localization relationship has been reported in several previous studies [22, 23]. Counting the proportion of cells in the samples from different tissue sites, FAP$^+$ fibroblasts and SPP1$^+$ macrophages were significantly increased at the edge of the tumor (Fig. 1E-F); their intercellular communication was also significantly increased (Figure S1A-D). FAP and SPP1 also had expression specificity in subsets of fibroblasts and macrophages (Fig. 1G, I).

We validated the association of the two types of cells with the highest risk scores (SPP1$^+$ macrophages and FAP$^+$ cancer-associated fibroblasts) in the bulk cohort, and found a high correlation between the two, indicating spatial co-localization ($R = 0.78$, $p < 0.001$) (Fig. 1H). We verified the localization of the two types of cells by deconvolution in Ye's and Gao's cohorts [23, 28], and found two types of patients with CRC: those with CAFs-TAMs with high co-localization, and those with CAFs-TAMs with low co-localization (Fig. 1J).

Table 1 Public datasets and sample information used in this study

dataset Bulk RNA-seq	tumor	normal	
TCGA-COAD	481	40	
GSE37892	130	0	
GSE14333	290	0	
GSE17536	177	0	
GSE39582	566	19	
GSE9348	70	12	
GSE8671	32	32	
GSE62932	64	4	
GSE198697	72	36	
GSE23878	35	24	
GSE72718	9(LIVER-colorectal cancer liver metastasis)	9(COLON-colorectal cancer liver metastasis)	10(primary colorectal cancer without liver metastasis)
GSE89393	5(colon tumor)	6(normal colon)	6(liver metastasis)
GSE144259	3(primary colon tumor)	3(normal colon)	3(liver metastasis)
GSE100243	8(lymph node metastasis)	7(normal colon)	8(primary colon tumor)
GSE179979	4(primary tumor tissue)	3(liver metastatic tissue)	
GSE44076	98(Tumor sample from patient)	98(Normal paired sample from patient)	50(Mucosa sample from healthy donnor)
dataset Single-cell			
EMTA-B8107	7	7	7(border)
Syn26844071	378	189	189(LymphNode)
GSE144735	6	6	6(border)
GSE166555	13	12	
GSE178318	6(primary CRC)	6(liver metastasis)	3(PBMC)
GSE200997	16	7	
OEP001756			
GSE158636	1	1	
dataset Spatial			
Ye's Cohort	Obtained from the original author		
Gao's Cohort			

Through cell communication analysis, we found that the communication between $SPP1^+$ macrophages and FAP^+ fibroblasts was enhanced at the tumor margin; we defined this type of CRC as CM.

Defining the "Collagen-Macrophage" patients and confirm their unique characteristics

We used WGCNA to identify the two modules most correlated with CAFs and $SPP1^+$ macrophages, performing functional enrichment analysis (Fig. 2A). Module 1 was enriched in the "Extracellular matrix organization," "Cell-substrate adhesion," and "Collagen metabolism" processes, while Module 2 was enriched in the "Macrophage activation," "Glycolysis and Gluconeogenesis," and "Myeloid leukocyte activation" processes, which correlate with the functions of macrophages (Fig. 2B).

We integrated the two modules and conducted consensus molecular clustering on a total of 1,337 patients from TCGA, GSE39582, and GSE14333, finding that these genes could distinguish between CM and non-CM CRC in the cohort (Fig. 2C). Three-dimensional PCA analysis was used to verify the differentiation between CM and non-CM CRC (Fig. 2D). To verify the consistency of our clustering in different cohorts, we took the intersection of differential genes ($logFC \geq 1$, $p < 0.05$) between CM and non-CM CRC in the three cohorts; the up-regulated genes in CM CRC in the three cohorts were consistent (inconsistency rate in the three cohorts: 13.3%, 27.3%, 0.045%). A total of 254 genes (Table 2) (Fig. 2E) co-upregulated in all three cohorts were enriched in the extracellular matrix-, TGF-β-, and immune cell activation-related pathways (Fig. 2F). The prognosis of patients with CM CRC

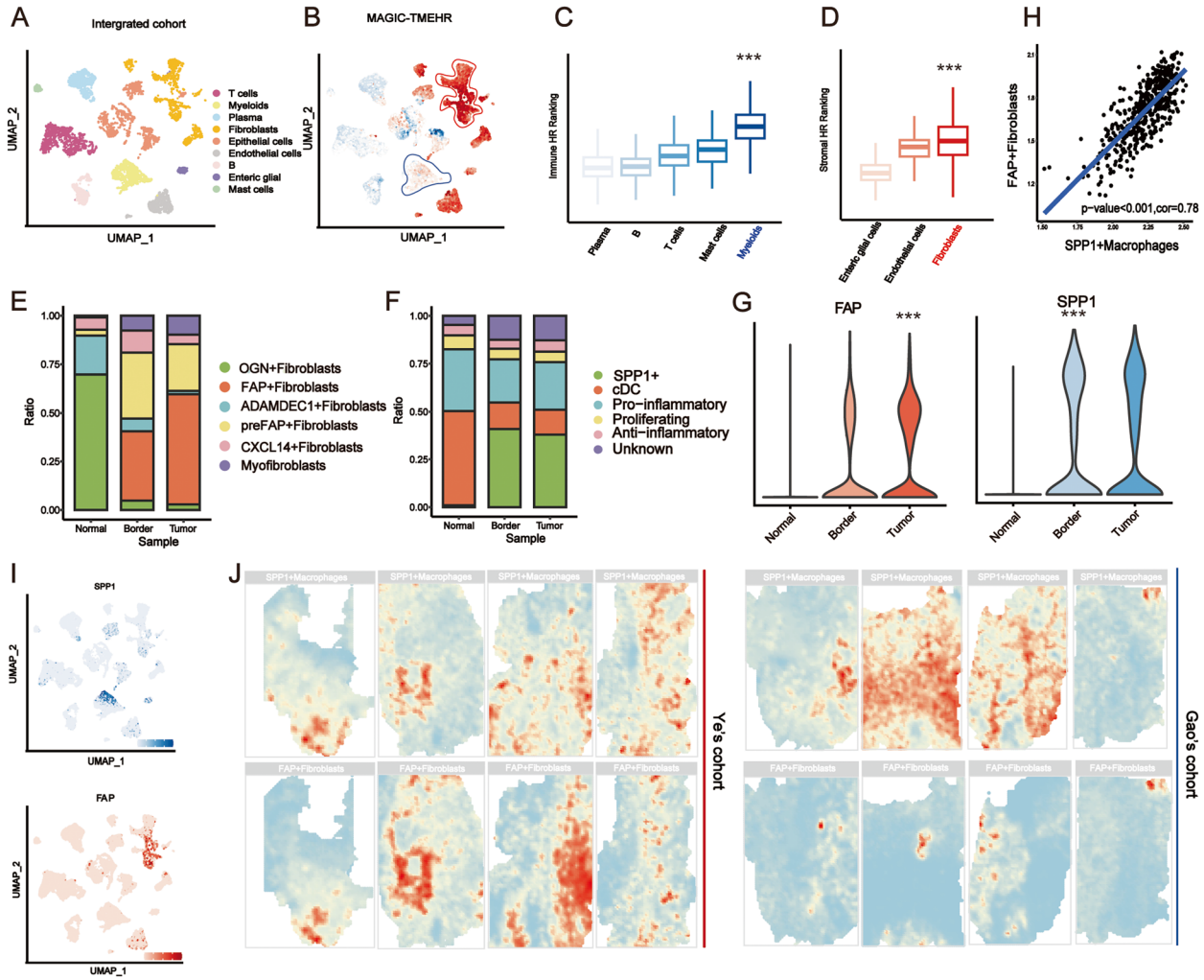

Fig. 1 Spatially enriched SPP1 + macrophages and FAP + fibroblasts at the edge of the tumor mass contribute to poor prognosis in patients with CRC

was significantly worse than those with non-CM CRC ($p = 0.022$) (Fig. 2G). We calculated the differences of all hallmarks between patients with CM and non-CM CRC, and found that epithelial-mesenchymal transformation, angiogenesis, the IL6-STAT3 pathway, the TGF-β pathway, and complement signaling were significantly up-regulated in those with CM CRC; oxidative phosphorylation, fatty acid metabolism, DNA repair, and MYC signaling were significantly up-regulated in non-CM CRC (Fig. 2H).

In the previously proposed CMS clustering method, CMS4 patients showed activation of TGF-β signaling, abnormal matrix proliferation, and increased metastasis of tumor cells. To explore the similarities and differences between CM and CMS classification, we observed the composition of the CM patient subclass. Patients with CM CRC were mainly composed of CMS4 cases, but also included some CMS1, CMS2, and CMS4 cases (Fig. 2I). After the use of UMAP to reduce the dimensionality, we found that CM classification could be used as an extension for CMS4 cases in CMS classification (Fig. 2J). Patients who are excluded from the CMS4 classification but who still belong to the CM class cannot be effectively treated.

We confirmed the molecular phenotypes of CM CRC from the perspective of immune infiltration and pathway activation. In the microenvironment of CM CRC, M0 macrophages, M2 macrophages, and tumor-associated fibroblasts were enriched, while $CD8^+$ T cells, plasma cells, and $CD4^+$ memory T cells representing anti-tumor immunity were significantly reduced (Fig. 2K). When we used all biologically relevant pathways in the Reactome database [29] to conduct ssGSEA score analysis and compare CM with

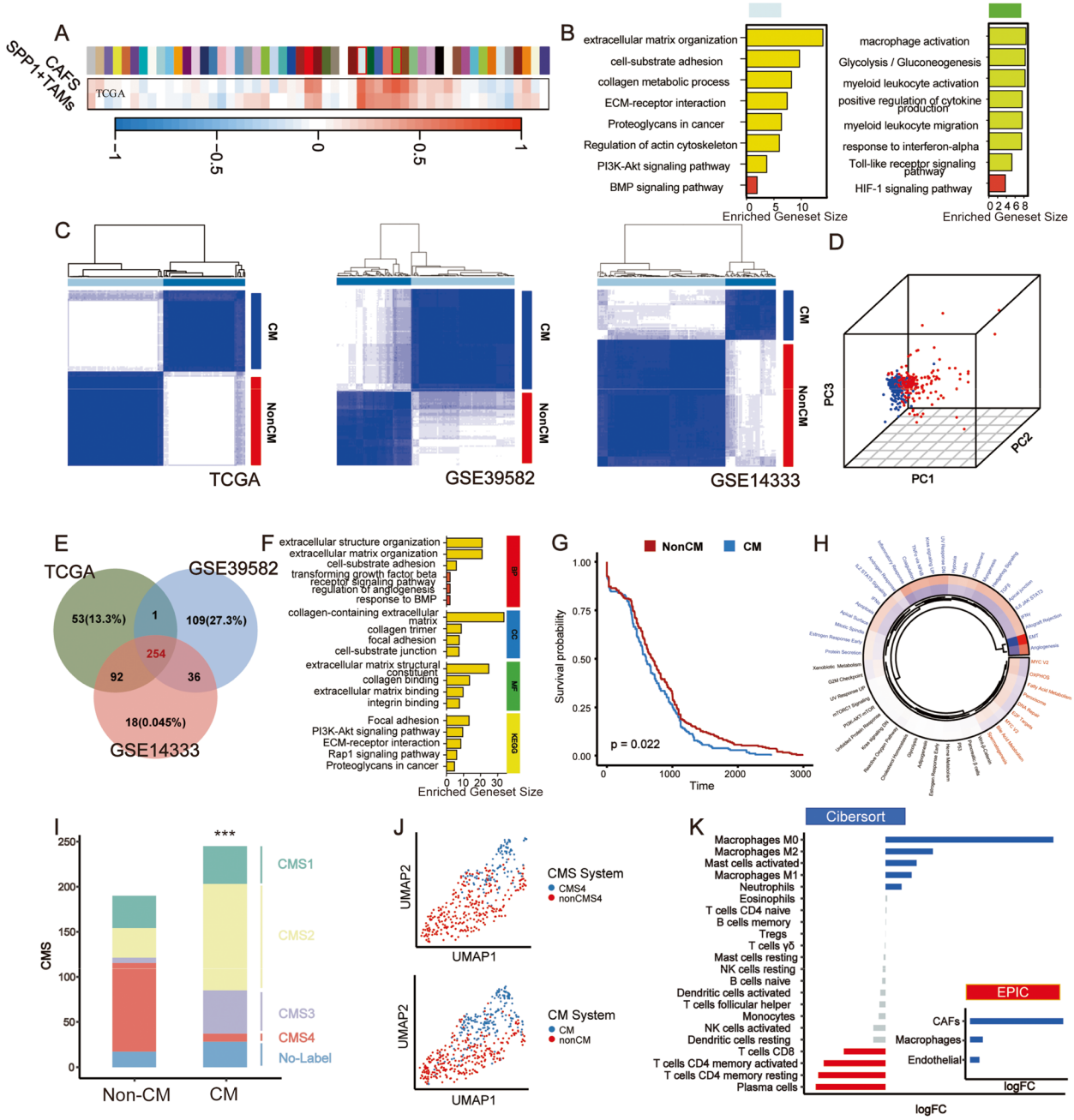

Fig. 2 Defining the "Collagen-Macrophage" patients and confirm their unique characteristics

non-CM CRC, the highest-ranked pathways in CM patients were extracellular matrix remodeling, collagen synthesis, and glycosaminoglycan synthesis, which was also seen in the GSEA results. We selected pathways related to collagen synthesis in the Biological Process category in the GO database [30], and found that collagen synthesis, collagen chain assembly, enzyme modification in collagen synthesis, and collagen crosslinking were all up-regulated in CM CRC. Together, these results indicate a distinct landscape of collagen dysplasia and immunosuppression in CM CRC (Figure S2A-C).

Table 2 List of genes used for consensus clustering to distinguish patients with CM

Genes							
SFRP2	GPC6	SGCE	PLEKHO1	COL5A1	NCF2	GFPT2	HCLS1
COL10A1	EMP3	COL15A1	EGR2	COL1A2	PTRF	ARMCX1	HLA-DRA
GREM1	CDH11	TIMP2	C3orf80	C1S	IGFBP6	PRKCDBP	IL6
FNDC1	CPA3	GUCY1A3	CD14	LOC101928916	SHISA2	POSTN	ZNF532
THBS2	INHBA	C1R	OLFML1	MFAP5	DPYSL3	GUCY1B3	GLIPR2
SPOCK1	FCER1G	COL3A1	C1QB	RP3-428L16.2	LAPTM5	SYNC	HLA-DQA1
CTHRC1	FN1	MEIS1	FCGR2B	SERPING1	NUAK1	RGS2	VGLL3
ASPN	GEM	TMEM47	FCGR1B	BGN	C1orf162	NDN	ENTPD1
SFRP4	EFEMP1	EPYC	CXCR4	FERMT2	PXDN	MEIS2	TMEM119
SULF1	UBE2E2	FCGR3A	ADAM12	NEXN	CRISPLD1	THY1	CXCL8
LINC01279	DPT	CLMP	CD163	MAB21L2	COL12A1	CCL21	TIMP1
FAP	MAFB	TIMP3	WISP1	GAS1	DACT1	TNFSF4	CCL8
PLN	BEX4	MRC1	NR3C1	OLFML2B	ALOX5AP	ITGA5	MITF
AEBP1	APOE	FLNA	RAI14	SERPINF1	MMP9	ANGPTL2	ELTD1
VCAN	COL1A1	GPX3	CSF2RB	GLT8D2	TCEAL7	TMEM45A	HCK
COMP	PDGFRB	EFEMP2	CXCL13	PCOLCE	FBLN2	PLXDC2	EVI2A
ISLR	LOX	DDR2	HSD17B6	SRPX	SLC2A14	MNDA	CCL3
TNFAIP6	BCAT1	CFH	HS3ST3A1	MRGPRF	VSIG4	PKD2	ROBO1
PPAPDC1A	SPP1	F13A1	TREM1	MXRA5	C5AR1	GPX8	MS4A7
RAB31	TNC	CYR61	PDGFRL	COLEC12	ITGAM	LY96	C1QC
FBN1	C11orf96	MGP	PLAU	CTSK	SLC2A3	MXRA8	PBX3
LOXL1	HMCN1	PDGFC	ADAMTS2	COL6A3	CYP1B1	LGALS1	C3AR1
COL11A1	FIBIN	TGFB1I1	S100A4	SPARC	COL8A1	CTGF	PTGS2
FRMD6	DFNA5	GPNMB	KIAA1462	SPARCL1	NAP1L3	SDC2	HSD11B1
LOC102725271	GLIS2	SLIT2	VIM	CCDC80	GJA1	CXCL12	EVI2B
RSPO3	VCAM1	TWIST1	HCAR3	EMILIN1	MSRB3	FAM129A	HLA-DPA1
NXN	RBMS1	MIR100HG	TNFSF13B	TUBB6	TUBA1A	ZNF521	GLIPR1
C3	RARRES2	COL6A2	PPP1R18	MFAP2	NID2	AKAP12	CSF1R
CCL18	TYROBP	ZFPM2	CFL2	LUM	FCGR3B	TSPYL5	SLFN11
COL5A2	KCTD12	PRRX1	MS4A4A	CCL2	ARL4C	COX7A1	CECR1
TAGLN	BCL2A1	SNAI2	C1QA	NARR	IQCJ-SCHIP1	CYBRD1	APOC1
HTRA1	DCN	OGN	SPHK1	CTSL	PNMA1		

Constructing CM-clustering differential expression gene scores (CSDGS) to distinguish CRC patients, its relationship to immunotherapy response prediction, and identification of thrombospondin-2 as the CM signature protein

The traditional theory of M1-M2 macrophage polarization of macrophages has been validated in many solid tumors, but it cannot fully elucidate the changes in macrophage status in the tumor microenvironment [31, 32]. The phenotypic landscape of macrophages based on CM typing needs to be explained. To do this, we used Cibercortx [33] and the integrated cohort we constructed to analyze differences in the macrophage landscape between CM and non-CM CRC. $SPP1^{+}$ macrophages were the only macrophage subpopulation that was enriched in CM patients (Figure S3A-C).

To better quantify CM patients, we reduced the dimensions of 254 CSDGs using the previously described PCoA dimensionality reduction, and found that PCoA1 could distinguish CM from non-CM CRC; we defined PCoA1 as a CM-clustering differential expression gene score system (CSDGS). The CSDGS score of patients with CM CRC was significantly lower than that of patients with non-CM CRC; those with lower CSDGS scores had more obvious CM characteristics (Fig. 3A). The TIDE score is a recognized immunotherapy-related score that has been effectively validated in many solid tumors and is considered an indicator of tumors with a "cold" immune environment. Analyzing the correlation between the CSDGS we obtained and the TIDE score, there was a significant negative correlation between the two ($R = -0.64, p < 0.001$,)

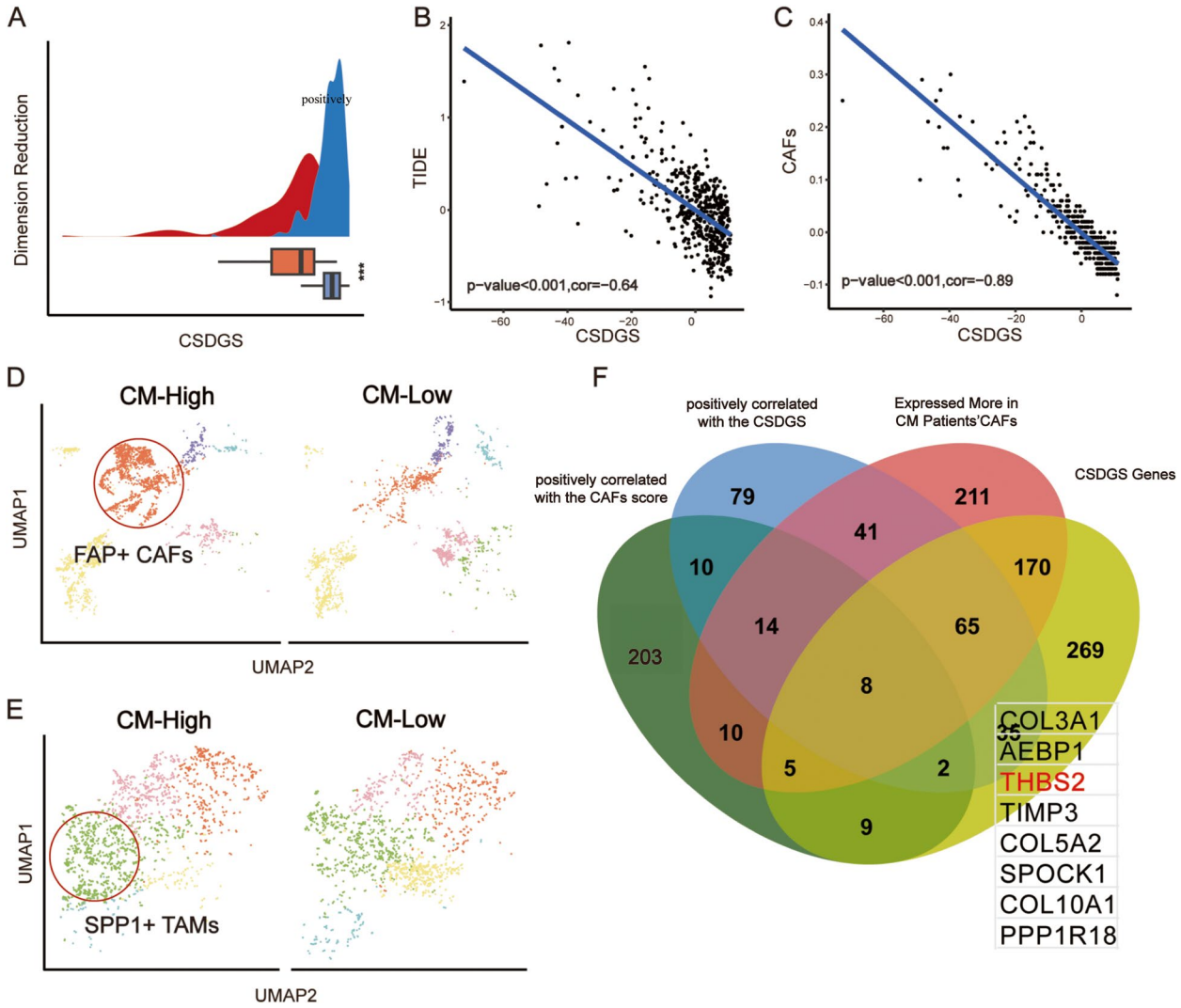

Fig. 3 Constructing CM-clustering differential expression gene scores (CSDGS) to distinguish CRC patients, its relationship to immunotherapy response prediction, and identification of thrombospondin-2 as the CM signature protein

(Fig. 3B), verifying the association between CM typing and immune-cold environment. CSDGS was also significantly correlated with the CAFs score obtained in the TIDE score (Fig. 3C).

After mapping CSDGS to our integrated cohort, we found that patients with CM-High had significantly more FAP^+ CAFs and $SPP1^+$ TAMs than patients with CM-Low (Fig. 3D-E). To find effective biomarkers to distinguish CM-High from CM-Low, we screened genes positively correlated with the CAFs score, those significantly positively correlated with the reciprocal of CSDGS, genes expressed more in FAP^+ fibroblasts in patients with CM-High than those with CM-Low in the single-cell integration cohort, and the 254 genes used to build the CSDGS (Fig. 3F). We screened out 8 biomarkers that could effectively distinguish CM from non-CM CRC; of these, *THBS2*, encoding thrombospondin-2(TSP2) protein, was considered a biomarker for CM patients.Among the eight candidate genes, excluding THBS2, most encode extracellular matrix components or lack direct involvement in intercellular communication within the tumor microenvironment, while THBS2-encoded thrombospondin-2 (TSP2) serves as a canonical secretory protein and signaling molecule that critically mediates intercellular crosstalk across distinct cellular populations, thus we hypothesize that THBS2 may function as a pivotal biomarker for CM patients.

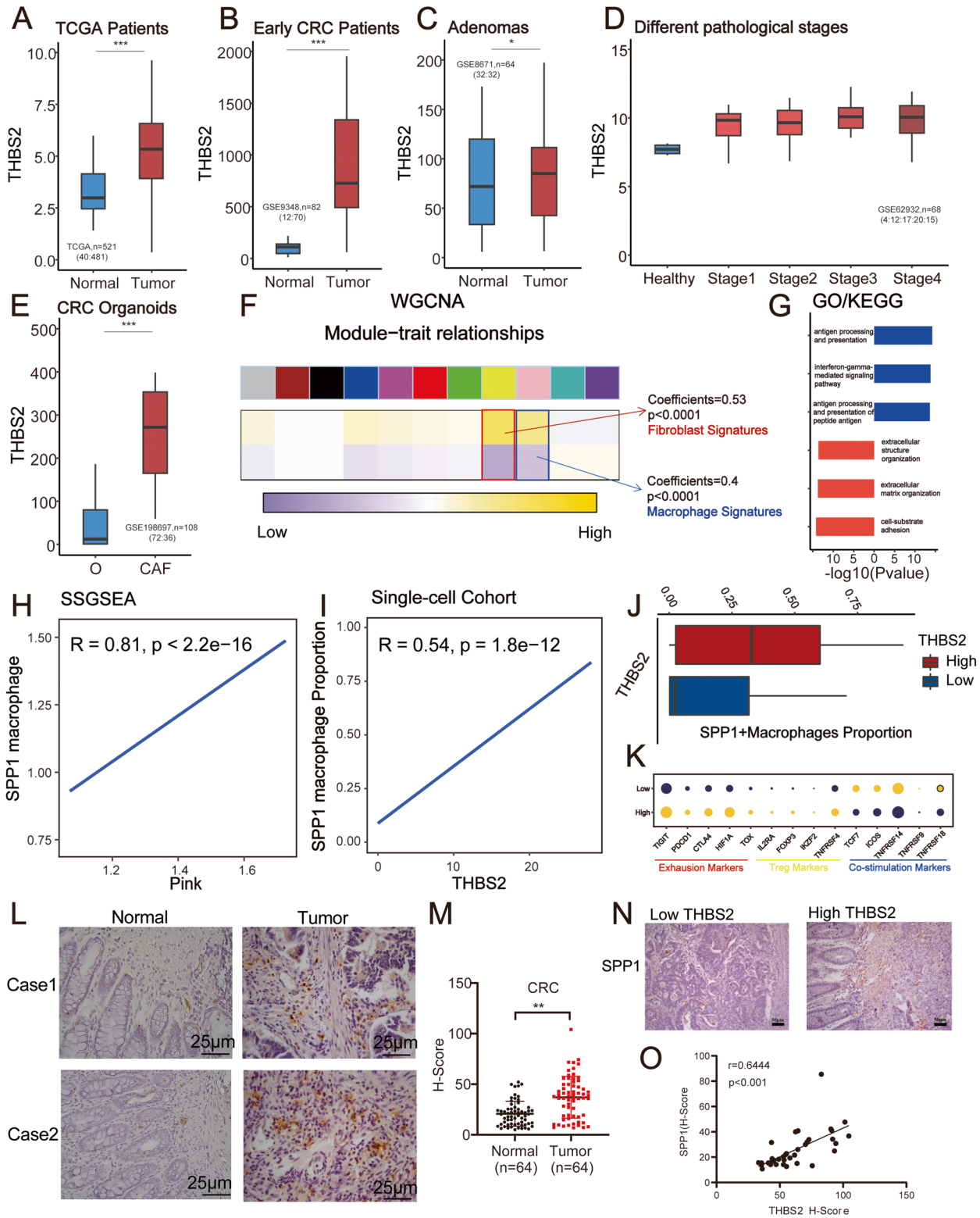

Fig. 4 Multiple cohort study analysis confirmed the role of THBS2 in CRC progression

Multiple cohort study analysis confirmed the role of *THBS2* in CRC progression

To further confirm the role of *THBS2* in CRC progression, we investigated the transcriptome sequencing cohort from current CRC-related studies. We confirmed the elevation of *THBS2* expression in tumor tissue (Fig. 4A), as well as in early tumors (Fig. 4B). We then investigated adenomas (a benign precursor form of colon cancer) and found that *THBS2* was also upregulated in early adenoma (Fig. 4C). The differences in expression were mainly reflected between healthy tissues and tumor tissues, while differences in different pathological stages were not obvious (Fig. 4D), suggesting that upregulation of *THBS2* is particularly important during oncogenesis. We also investigated data from a cohort of organoid co-cultures and found that the expression level of *THBS2* was significantly increased within the system co-cultured with CAFs (Fig. 4E), validating our previous screening and confirming that *THBS2* is a functional protein specifically secreted by CAFs.

To explore the role of *THBS2*, we constructed a gene co-expression network in TCGA-COAD with genes encoding the top 5000 proteins; a total of 11 co-expression modules were constructed, in which the yellow (coefficients = 0.53, $p < 0.0001$) and pink modules (coefficients = 0.4, $p < 0.0001$) were significantly correlated with *THBS2* expression (Fig. 4F). We performed functional enrichment for the two modules, and found the yellow module was mainly enriched in pathways such as extracellular matrix remodeling and cell–matrix adhesion, as well as most of the genes specifically expressed by fibroblasts; the pink module was mainly enriched in antigen processing and presentation and IFNγ-related signaling pathways, with most of the genes being those highly expressed in macrophages (Fig. 4G). These results further suggested that *THBS2* may act as a bridging molecule between CAFs and macrophages. We scored the genes in the pink module by ssGSEA, and the score was highly associated with SPP1 + macrophage expression ($R = 0.81$, $p < 0.00001$) (Fig. 4H). We subsequently verified that the level of CAFs expressing *THBS2* in patients in the single-cell cohort was highly correlated with the proportion of $SPP1^+$ macrophages (Fig. 4I-J).

In previous studies, $SPP1^+$ macrophages mediated T cell exhaustion so that T cell cytotoxicity to tumor cells was suppressed. We verified this in a single-cell cohort, finding that patients with high *THBS2* expression had high expression of exhausted T cell markers such as TIGIT, PDCD1, CTLA4, HIF1A, and TOX [34, 35]. Similarly, markers from T_{reg} cells that mediate immunosuppression and immune evasion were also highly expressed in patients with high *THBS2* expression; however, molecules that mediate co-stimulatory activated T cell activity, such as ICOS, TNFRSF14, TNFRSF9, had lower expression in patients with high *THBS2* expression (Fig. 4K). We subsequently validated THBS2 expression in our own CRC tissue samples and found that it was significantly up-regulated in tumor tissues (Fig. 4L-M). The expression of SPP1 in tissues of patients with high THBS2 expression was significantly higher than in those with low THBS2 expression, and the expression of THBS2 and SPP1 were highly correlated in human samples (Fig. 4N-O). These results all validated that THBS2 may serve as a functional protein unique to CM CRC.

The role of THBS2 in CRC disease progression has not been fully described. To comprehensively investigate the effects of THBS2 on the tumor and tumor microenvironment, we performed in vivo and in vitro experiments. CRC cell lines HCT116 and RKO cells were stimulated with THBS2 recombinant protein, and CCK8 and transwell experiments were performed successively; We found that THBS2 could effectively promote the proliferation, migration and invasion of CRC cell lines (Figure S4A-C). We then constructed a mouse subcutaneous tumor model by injecting 5×10^5 MC38 cells into the back of each mouse. When the tumor volume of MC38 reached ~ 100 mm^3, the tumor was injected with adeno-associated virus (AAV) with specific knockdown *THBS2* (Adv-*THBS2*) or control AAV (Adv-Ctrl) at 0, 3, and 6 days at a viral dose of 3×10^8 PFU/injection. As expected, the tumor growth of the mice injected with Adv-*THBS2* was significantly slower than that of the Adv-Ctrl group (Figure S4D), and the tumor weight was smaller (Figure S4E). An increase in the expression of CD8, a tumor infiltration marker, was observed in mice in the Adv-*THBS2* group (Figure S4F-G). These results indicate that *THBS2* can promote tumor development in vivo and in vitro.

FAP^+CAF-secreted *THBS2* promotes the recruitment of $SPP1^+$TAMs, predicts the efficacy of immunotherapy in CRC patients, and is a target for combination immunotherapy

We extracted the fibroblasts from the integrated data and divided them into seven groups based on cell markers. $ADAMDEC1^+$ fibroblasts were previously found to significantly promote inflammation [36], while $CXCL14^+$ fibroblasts were also reported as a cluster characterized by high expression of BMP family proteins [37]; in our data, this cluster also promoted the development of inflammation. OGN-labeled fibroblasts are mainly expressed in normal tissues and were considered to be normal [37]. FAP-labeled fibroblasts were considered cancer-associated [38], while the cluster with slightly lower levels of FAP expression was considered precursors for FAP-labeled fibroblasts ($preFAP^+$ fibroblasts).

THBS2 was specifically expressed in FAP$^+$ fibroblasts and its expression was enhanced with the activation of CAFs (Figure S5A). A pseudotime analysis of the fibroblasts found that the pro-inflammatory fibroblasts labeled with CXCL14 and ADAMDEC1 reached different developmental ends from the immunosuppressive fibroblasts labeled with FAP, similar to the M1-M2 macrophage subtyping system. Based on these results, antagonism of two types of fibroblasts appears to exist in CRC patients: those expressing ADAMDEC1 and CXCL14 to promote the development of inflammation, and immunosuppressive fibroblasts characterized by FAP expression (Figure S5B). FAP$^+$ fibroblasts had a higher correlation with SPP1$^+$ TAMs relative to preFAP$^+$ fibroblasts in both the transcriptome and the single-cell transcriptome cohorts (Figure S5C-D). In a large colon cancer single-cell cohort provided by Dan et al., pro-inflammatory and immunosuppressive fibroblasts were negatively correlated in patient tissues (R = -0.703, $p < 0.001$)(Figure S5E) [39]. High expression of FAP$^+$ fibroblasts was previously recognized as a marker associated with immunosuppression, affecting prognosis and causing poor survival outcomes; however, inflammatory fibroblasts highly expressing inflammatory molecules significantly improved patient outcomes. The imbalance of F1-F2 (anti-inflammatory and pro-inflammatory) fibroblast differentiation in CM patients promotes the formation of an immunosuppressive microenvironment, leading to poor prognosis (Figure S5E-F). We also used mIHC to confirm that *THBS2* was almost exclusively derived from FAP$^+$ fibroblasts and encapsulated in the marginal part of the tumor (Figure S5H).

Traditional enrichment analysis is based on the differentially expressed genes. Some molecules specifically expressed in some cells may indicate gene biomarkers inherent to the cluster itself (Fig. 5A). Analyzing differentially expressed genes between the *THBS2* high/low group in the TCGA-COAD cohort, genes such as COL1A1 and COL3A1 had high logFC values, though this did not mean that they had biological functions associated with *THBS2*. We decreased the selected interval, choosing genes with a logFC between 1 and 1.5 that were significantly enriched in pathways related to intercellular communication, including regulation of macrophage activation, endothelial cell proliferation, and T cell activation. The results demonstrated that TSP2 had a regulatory effect on the immune microenvironment (Fig. 5B-C). We then compared the cellular composition of patients with high and low *THBS2* expression in a single-cell cohort; the infiltration proportion of SPP1$^+$ cells was not significantly different between groups, but the expression of SPP1 was significantly increased in those with high *THBS2* expression. Additionally, both preFAP$^+$ and FAP$^+$ fibroblasts were significantly increased in those with high *THBS2* expression. In cell communication analysis, the communication between SPP1$^+$ macrophages and FAP$^+$ fibroblasts was significantly enhanced in the high *THBS2* expression group, including MIF and SPP1-mediated ligand-receptor signaling, further suggesting the potential role of *THBS2* in the formation of the TIB (Fig. 5D-E).

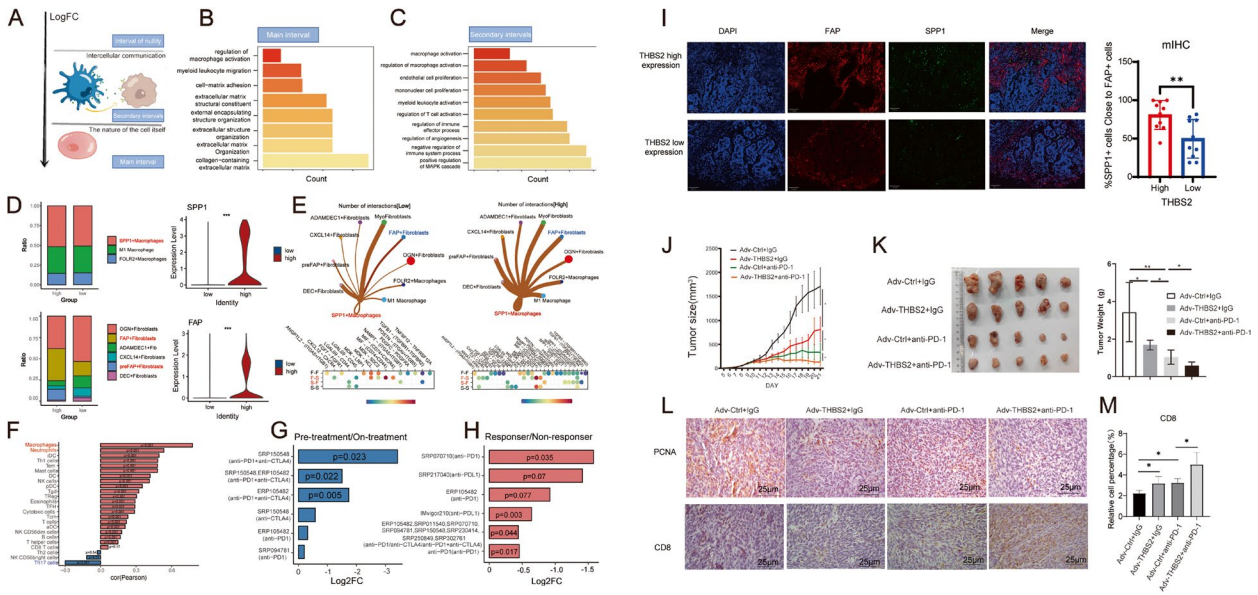

Fig. 5 FAP + CAF-secreted THBS2 promotes the recruitment of SPP1 + TAMs, predicts the efficacy of immunotherapy in CRC patients, and is a target for combination immunotherapy

We calculated the correlation between *THBS2* and different components in the cell fraction of the immune microenvironment and found that it was highly correlated with the infiltration level of macrophages (cor > 0.7, $p < 0.0001$) (Fig. 5F). To verify the effect of *THBS2*on patient response to immunotherapy, we retrospectively selected patient data containing responders and non-responders pre- and post-treatment [40]. *THBS2* expression was significantly higher in the non-responders than in responders and tended to decrease after the treatment (Fig. 5G-H). The differences in the expression of SPP1 and FAP in patients with high and low TSP2 expression were verified by mIHC, and SPP1 and FAP were enriched at the edge of the tumor mass in tissues with high TSP2 expression (Fig. 5I). Combined treatment with *THBS2* adenovirus and PD-1 monoclonal antibody in mice showed significant tumor control (Fig. 5J-K), demonstrating the potential of TSP2 as a combined target for PD-1 treatment; infiltration of CD8T cells in the tumor tissue of the combined treatment group was significantly increased. These results demonstrated that TSP2 blockade could break the CAFs-TAMs system at the tumor edge, promoting infiltration of CD8T cells and enhancing the immune response (Fig. 5L-M).

THBS2 promotes tumor metastasis by activating MMP expression and is an effective biomarker for metastasis

To elucidate the role of *THBS2* in tumor progression, we scored all of the hallmark events in tumorigenesis, finding that *THBS2* was significantly correlated with EMT, angiogenesis, and hypoxia in tumors (Fig. 6A), three biological processes involved in metastasis. To verify whether *THBS2* promoted the metastasis of tumor cells, we retrieved data from TCGA and GSE72718. The expression of *THBS2* in patients with lymphatic invasion was significantly higher than that in patients without, and the expression of *THBS2* in colon primary tumors of patients with liver metastasis was also significantly higher than in those without liver metastasis (Fig. 6B). This demonstrated a significant association between high *THBS2* expression and metastatic events in colon tumors. To explore the reason behind this, we investigated the classical mechanisms of tumor metastasis from reviews [41]. Secretion of MMPs [42] was closely linked to the expression of *THBS2* (Fig. 6C-D). MMP family members, such as MMP7, MMP11, MMP2, MMP1, and MMP3, had significantly elevated expression in patients with high *THBS2* expression; this was also true in a single-cell cohort. MMP1 and MMP7, specifically secreted by tumor cells, were highly correlated with *THBS2* expressed in CAFs (Fig. 6E-H). In vitro cell experiments verified that the expression of MMP1 and MMP7 in tumor cells (HCT116 and RKO) stimulated with TSP2 recombinant protein was significantly higher than that in the control group (PBS) at the mRNA and protein levels (Fig. 6I-J). These results demonstrated the role of *THBS2* as a key dual switch molecule in the development of CRC. TSP2 is a biomarker of CM; it is secreted by CAFs and plays an important role in promoting the immunosuppressive barrier constructed by TAMs as well as in tumor metastasis. TSP2 can be used as a molecular target for PD-1 immunotherapy (Fig. 6K).

Discussion

Molecular classification of tumors has become the current mainstream process [43, 44], prolonging patient prognosis through clinical classification under the guidance of precision medicine. CRC is one of the most common cancers; it has a poor prognosis and great intra-patient heterogeneity. A molecular classification system for CRC has been proposed for a long time; however, although molecular clustering based on the whole transcriptome can effectively distinguish the differences in biological pathways between different patients, it cannot be effectively translated into clinical practice [45, 46]. Because of this, it is urgent to identify more biologically significant molecular typing methods, find corresponding sensitive biomarkers, and identify molecular targets.

PD-1 therapy has started a new era in the treatment of tumors, but it has encountered many difficulties in treating solid tumors. In CRC, previous studies have shown that T cell infiltration is minimal in the vast majority of patients, especially those with MSS-CRC [47]. A large number of immunosuppressive cells infiltrate these tumors, directly leading to the exhaustion of T cells and the failure of PD-1 monoclonal antibody treatment. In our study, we found that patients with this "cold microenvironment," who we classified as CM, had significant stromal hyperplasia (CAF enrichment). Additionally, we demonstrated that FAP^+ CAFs and $SPP1^+$ TAMs formed an immune barrier around the tumor in such patients through multi-cohort and multi-omics studies. In patients with a high degree of CAF-TAM co-localization, such coupling bands are thought to cluster around the malignant tumor, possibly contributing to the formation of an immunosuppressive environment and poor response to immunotherapy. The pathway landscape of patients with CM CRC was different from that of those with non-CM CRC, which may be the reason behind collagen formation and TAM recruitment. We developed a score (CSDGS) that effectively differentiated patients with CM CRC. As a new type of colon cancer classification, the CM system breaks through the limitations in the traditional CMS classification for patients with CMS4 and reveals a group of patients with silent response to

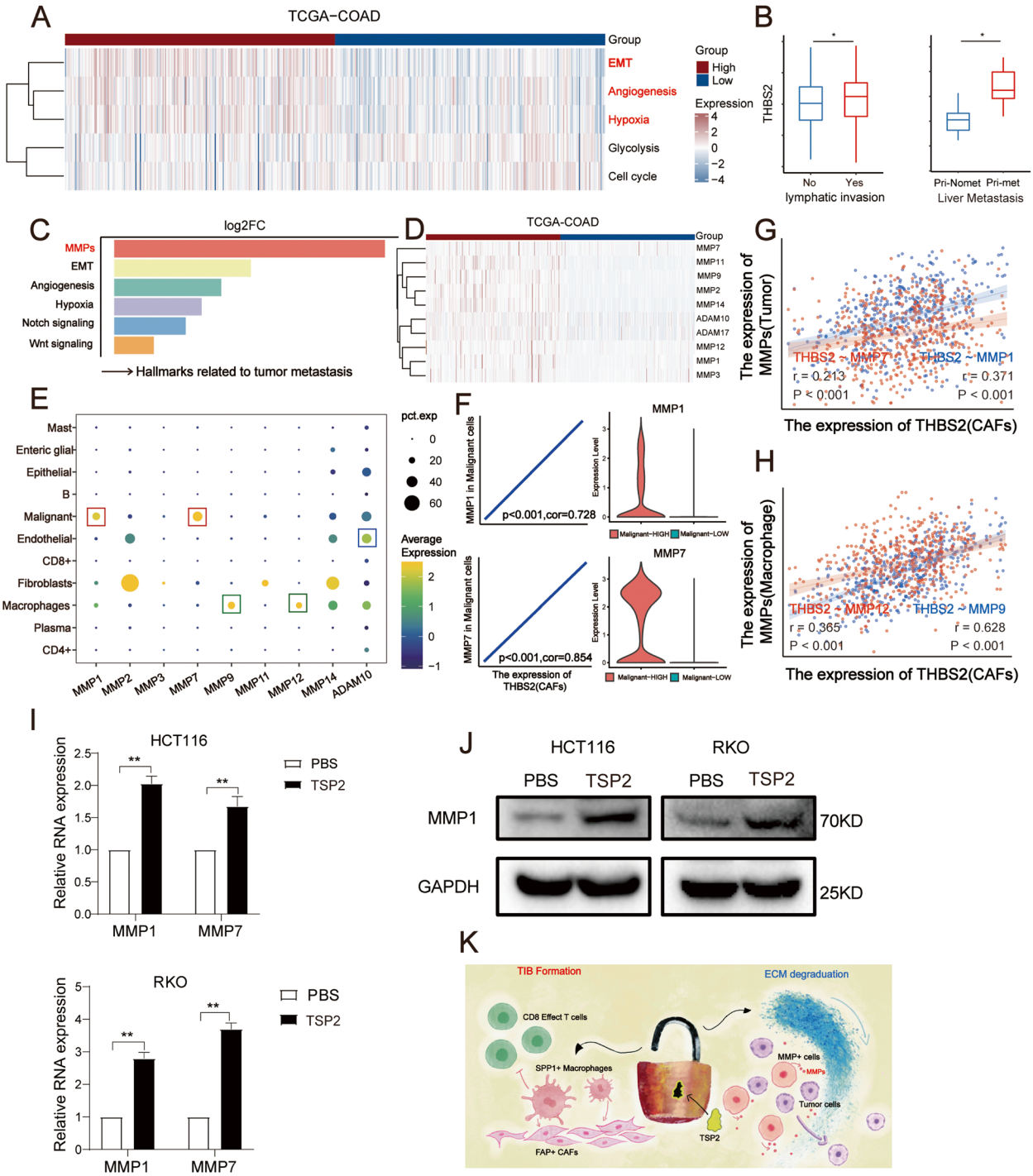

Fig. 6 THBS2 promotes tumor metastasis by activating MMP expression and is an effective biomarker for metastasis

immunotherapy because of abnormal TGFβ signal transduction and immunosuppressive barrier formation.

We also confirmed that TSP2, a CAF-derived secreted protein, could be a useful biomarker for CM patients. In previous studies, *THBS2* was reported to be a biomarker and combined therapeutic target for lung adenocarcinoma immunotherapy [48]. In pancreatic cancer, it has been reported to play an important role in the tumor-matrix loop to activate downstream MAPK signaling [49]. However, its role in CRC, especially in the formation

of the tumor immune microenvironment, has not been fully elucidated. We confirmed the biological function of TSP2 through in vivo and in vitro experiments, as well as clinical sample verification. The results suggested that *THBS2* may form an immune barrier by collecting $SPP1^+$ macrophages, promoting the exhaustion of toxic T cells. TSP2 can promote the recruitment of $SPP1^+$ TAMs to block immunotherapy efficacy, and also promote the invasion and metastasis of tumor cells by promoting the secretion of MMPs. These results suggest that TSP2 plays a pro-tumor role not only in the growth of tumor cells but also in promoting the formation of immunosuppressive barriers to create an immune microenvironment that promotes tumor growth. Its bidirectional effect suggests its potential to be translated into a combination therapeutic target for PD-1 mAb therapy.

Our study provides a new way of analysis where molecular typing of patients is no longer macroscopically based on genome-wide characteristics but is extended to more accurate molecular typing based on specific biological backgrounds and pathways. Starting from the macroscopically microenvironment-expressed genes related to poor prognosis in patients with CRC, we identified fibroblasts and myeloid cells as the respective types of stromal and immune cells that lead to poor prognosis. We then precisely identified FAP^+ CAFs and $SPP1^+$ TAMs in these two types of cells, finding that they coupled at the tumor edge in selected patients. This classification method is innovative in that it does not use the traditional posterior model of "classification first, interpretation later"; it makes a division based on known important biological phenomena, and then simultaneously verifies the correctness of the division and discovers new biological targets and markers. Such a typing method can aid in finding biologically valuable conclusions using patient genetic information, providing a new perspective for clinical translation.

As a group of newly identified patients through retrospective research, patients with CM CRC need more study using proteomics, metabolomics, and other omics techniques. As TSP2 is secreted, whether its plasma level can be correlated with the relevant clinical features and molecular phenotypes of CM cases needs to be examined; however, the identification of CM CRC, the identification of TSP2 molecular targets, and the attempt of combined TSP2 PD-1 therapy in our study will provide a new perspective for more comprehensive and precise treatment of CRC.

Our study is constrained by several limitations and shortcomings, notably the lack of in-depth investigation into the molecular mechanisms underlying THBS2-induced upregulation of MMPs expression and enhanced invasive capacity in colorectal cancer cells, as well as the unexplored clinical applicability of the CM classification system, where the potential of CM subtyping to guide clinical therapeutic decision-making in colorectal cancer necessitates further empirical validation. To address these gaps, we plan to conduct future studies focusing on THBS2 as both a biomarker and a targetable molecule for CM-subtyped patients, aiming to establish its pathophysiological significance.

Methods

Clinical samples

A total of 100 pairs of paraffin-embedded human CRC tissue samples and corresponding adjacent normal mucosal samples were collected at the First Affiliated Hospital of Nanjing Medical University and the Nanjing Drum Tower Hospital. Before surgery, no patients received chemotherapy or radiotherapy.A total of 52 pairs of CRC samples from the First Affiliated Hospital of Nanjing Medical University were quickly frozen in liquid nitrogen immediately after collection and stored at -80 °C until RNA extraction. This project was approved by the research ethics committees of Nanjing Medical University (approval ID: (2016)640) and Nanjing Drum Tower Hospital (approval ID: 2020–378-01). All patients signed the written informed consent form before participating in this study.

Cell culture

Human CRC cell lines HCT116 and RKO were purchased from the Shanghai Institute of Cell Biology (Chinese Academy of Sciences, China). The cell lines were characterized by STR profiling and confirmed to be mycoplasma-free. Cells were cultured in RPMI 1640 medium (HyClone, South Logan, UT, USA) supplemented with 10% FBS (Gibco, Grand Island, NY, USA), 100 U/mL penicillin, and 0.1 mg/mL streptomycin (Gibco). All cell lines were cultured in a humidified incubator at 37 °C with 5% CO_2. After treatment of cells with Verteporfin (MedchemExpress, Monmouth Junction, NJ, USA) or XMU-MP-1 (MedchemExpress), cells were collected for further study. RKO and HCT116 cells were stimulated with 100 ng/mL TSP2 recombinant protein for 48 h to collect protein and RNA.

CCK8 assay

Cell proliferation was measured using the Cell Counting Kit 8 (Medchem Express). Briefly, CRC cells transfected with siRNA or plasmids were seeded into 96-well plates at a density of 1×10^3 cells per well. After cell adhesion, medium containing TSP2 recombinant protein at 100 ng/mL was added to the corresponding cells. Cell viability was measured by CCK-8 (Dojindo) after 0, 24, 48, 72, and 96 h according to the method provided by

the manufacturer. Absorbance at 450 nm was measured using a microplate reader (Bio-Tek). Each experiment was performed in triplicate and repeated at least three times.

Transwell

Cell migration and invasion assays were performed using 24-well Transwell chambers (polycarbonate membrane with 8 μm pore size, Corning, New York, USA). Cells (6×10^4 HCT116 or 6×10^4 RKO) in 200 μL serum-free medium were inoculated into the upper well chambers uncoated or coated with Matrigel for migration or invasion assays, respectively. The lower chamber was filled with 600 μL of medium containing 10% FBS. TSP2 recombinant protein at a concentration of 100 ng/mL was added and incubated for 24–36 h. Cells on the upper surface of the membrane that had not migrated or infiltrated were wiped away. Cells on the submembrane surface were fixed with 95% ethanol and stained with 0.1% methanol/PBS crystalline violet, and then observed with a Zeiss microscope (Melville, New York, USA). Five stained areas were randomly selected in each group and stained cells were counted to assess cell migration and invasion. Each experiment was repeated at least three times.

Immunohistochemistry

Primary antibodies used for immunohistochemical staining were anti-TSP2 (Biodragon), anti-PCNA (Abcam), anti-FAP (Abcam), anti-CD8a (proteintech), and anti-SPP1 (abcam). The secondary antibody was horseradish peroxidase (HRP)-labeled anti-rabbit IgG (Dako Cytomation). Paraffin-embedded human and xenograft mouse CRC tissues were cut into 5 μm thick sections. All sections were sequentially deparaffinized, rehydrated, antigen recovered, and then incubated with primary and corresponding secondary antibodies. Sections were color-developed with 3,3'-diaminobenzidine (DAB) and counterstained with hematoxylin. The intensity of staining was categorized into four levels: 0 (negative), 1 (weakly positive), 2 (moderately positive), and 3 (strongly positive). The percentage of cells with different immunostaining results was assessed. The H-score was determined according to the following formula: [(percentage of weak staining)×1]+[(percentage of moderate staining)×2]+[(percentage of strong staining)×3], with a range of results from 0 to 300. The expression level of each protein was categorized as low or high according to the median H-score. Protein expression levels in CRC tissues were categorized as low or high based on the median H score.

Quantitative analysis of multiplex immunohistochemistry

Whole tumor tissue sections were processed using multiplex Immunohistochemistry (mIHC). Whole-slide imaging was performed on a Vectra Polaris system (PerkinElmer) at 20×resolution. The entire tumor area was then analyzed using QuPath v0.4.3. Nuclei were segmented with a pre-trained U-Net model ($\sigma=1.5$ μm, background radius = 8 px), and cell phenotypes were defined based on marker expression and morphometric criteria. Spatial co-localization was quantified as the percentage of $SPP1^+$ macrophages within 20 μm edge-to-edge distance of FAP^+CAFs.

Tumor formation in C57BL/6 mouse models

Male C57BL/6 mice (4–5 weeks old) were purchased from the Model Animal Research Center of Nanjing University (Nanjing, China) and randomly divided into two or four groups (5 mice per group). Mice were kept under pathogen-free conditions in a 12-h light/12-h dark cycle. Mice were injected subcutaneously with mouse colorectal cancer MC38 cells (5×10^5) in the hind abdomen (5 per group). Tumor growth was checked every other day and tumor volume was assessed according to the following formula: volume = length × width^2 × 0.5. When a tumor volume of approximately 100 mm^3 was reached, simple intra-tumoral injections were performed at three injection time points on days 0, 3, and 6 to knock out *THBS2* using adenovirus; when the tumor volume reached approximately 1,000 mm^3, the mice were necropsied, and the xenografts were dissected and weighed for further studies. For further anti-PD-1 treatment by injection of knockout *THBS2* virus, the time points of virus injection were 0, 3, and 6 days, the time point of intraperitoneal injection of anti-PD-1 injection was 1 day, the mice were necropsied 6 days after the injection of anti-PD-1, and the xenografts were dissected and reweighed for further study. Mice were sacrificed by spinal dislocation after anesthesia with sodium pentobarbital at an injection dose of 50mg/kg by intraperitoneal injection.Animal welfare and experimental procedures were approved by the Animal Care and Use Committee of Nanjing Medical University (approval number: 1601080). All experiments were performed in accordance with the Guide for the Care and Use of Laboratory Animals. No blinding was done. The sequence used for adv-*THBS2* was: GGACAA UUGCCAGCUUCUCUUTT.

Multiplex immunohistochemistry and image acquisition

Histologic analysis was performed on formalin-fixed paraffin-embedded (FFPE) colon cancer tissues. FFPE tissues were sectioned to 5 μm thick using a manual deparaffinizing slicer (LAIKA) and subjected to multiple stains. After initial deparaffinization, heat-induced epitope retrieval (HIER) was performed on all sections in ethylenediaminetetraacetic acid (EDTA) buffer (pH 8) at 57 °C for 120 min; this was followed by multiple

rounds of staining including antigen retrieval, blocking, primary antibody, HRP-coupled secondary antibody, and tyramine signal amplification (TSA) steps. Each round of staining was followed by an antibody stripping step (pH 8, 57 °C). To visualize the immunofluorescence signal, the AFIHC TSA stains TYR-520, 570, and 690 were used. Spectral DAPI (AFIHC) was used for nuclear counterstaining. Slice imaging was performed on a Vectra 3.0 automated imaging system. The entire tissue was scanned (10X magnification) before areas within the tumor and peri-tumor tissue (tumor margins) were selected based on pan-melanoma staining specific for *THBS2*, SPP1, and FAP and imaged at high resolution using Phenochart at 20X magnification.

Western blot analysis

Protein lysates from cells or tumor tissues were separated by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE), transferred to polyvinylidene difluoride (PVDF) membranes, and blotted using the following antibodies: MMP1 (proteintech), MMP7 (proteintech), *THBS2* (biodragon), FAP (Abcam), and SPP1 (Abcam).

RNA extraction and reverse-transcription quantitative PCR

Total RNA (1 μg) was reverse transcribed using the PrimeScript™ 1st Strand cDNA Synthesis Kit (Takara Bio, Shiga, Japan) according to the user's manual and the expression level of each RNA was detected with an ABI 7300 Sequence Detector (Applied Biosystems, Foster City, CA, USA) using a SYBR Green Master Mix (Takara Bio). The relative fold change in RNA expression was then calculated using actin normalization by the ΔΔCt method. Each experiment was performed in triplicate and repeated at least three times. The PCR primers used in this study were human-actin-F (TCATGAAGTGTG ACGTGGACAT), human-actin-R (CTCAGGAGGAGC AATGATCTTG), human-MMP1-F (AAAATTACACGC CAGATTTGCC), human-MMP1-R (GGTGTGACATTA CTCCAGAGTTG), human-MMP7-F (GAGTGAGCT ACAGTGGGAACA), and human-MMP7-R (CTATGA CGCGGGAGTTTAACAT).

Data source and processing

This study includes data from 17 cohorts (Table 1). Data from GEO were downloaded from (https://www.ncbi.nlm.nih.gov/geo/), and data from the TCGA database were downloaded from (https://portal.gdc.cancer.gov/). Data from Synapse were obtained from the Synapse website as agreed to by the research group owning the source data (https://www.synapse.org/#!Synapse:syn26720761/ and https://www.synapse.org/#!Synapse:syn26844071/).

For single-cell data derived from tumor and normal tissue within the dataset, quality control was implemented using the Seurat package, applying filtration criteria including sequencing depth, total gene count, and mitochondrial gene percentage, with thresholds set as 500–2,000 genes per cell and ≤ 10% mitochondrial gene content. Following quality control, clustering analysis was performed on the processed data, after which the top 2000 genes demonstrating maximum expression variability across all samples were selected for normalization. Principal component analysis (PCA) utilizing the top 20 principal components achieved spatial dimensionality reduction in principal component clustering with a resolution parameter set to 0.2. Automated cell annotation was conducted via SingleR, complemented by manual annotation to categorize cells into six major groups: B cells, T cells, myeloid cells, endothelial cells, epithelial cells, and other subtypes.

As previously described, the FPKM values of TCGA-CRC (COAD) were converted to transcripts per million (TPM), and the same approach was used for the transcriptome data from GEO. Clinical information of the patients in the cohorts, including survival time, survival status, tumor pathology stage, sex, lymphatic invasion, and metastasis, are available from the datasets or can be acquired from the corresponding author.

Gene set acquisition

The ligand genes of the $SPP1^+$ macrophage subtype involved in the formation of the tumor immune barrier (TIB) were obtained from previous studies. Gene sets from the $SPP1^+$ macrophages and CAFs in CRC were all analyzed and acquired from the EMTA-B8107 data. Relevant genes regarding tumor progression (such as EMT, hypoxia, and angiogenesis) were acquired from the MsigDB database (https://www.gsea-msigdb.org/gsea/msigdb), while genes related to tumor metastasis were collected from pertinent literature. Gene lists have been available in the supplementary material of the article.

Dimensional reduction and clustering of the single-cell RNA sequencing data

Seurat v4 [50, 51] in R 4.4.0 was used for dimensionality reduction, clustering, and visualization of the scRNA-Seq data. Principal component analysis (PCA) was used to reduce the dimensionality of the high-dimension data. Based on the top 30 principal components (PCs), the integrated dataset was further reduced to 2D space and visualized by Uniform Manifold Approximation and Projection (UMAP). FindNeighbors(), FindClusters(), and RunUMAP() were used to perform these analyses at 0.1–0.5 resolution. Annotations for cells derived from

previous studies are all presented in the supplementary material and relevant references have been provided.

Pseudotime analysis

Monocle2 was used to analyze the specified pseudotime of the cell along the trajectory in scRNA-seq data [52]. The input for Monocle2 was the TPM matrix, and the cells were annotated with the cluster numbers identified by Seurat. Next, we estimated the size and dispersion of the data. Genes required for the pseudotime analysis were obtained using the differentialGeneTest function, all with $p < 0.05$. Following this, the data were reduced to 2D by the DDRTree algorithm and cells were sorted along the trajectory of the specified starting point.

Cellchat analysis

Cellchat [53] was used to calculate and analyze intercellular communications among macrophage clusters. Cellchat results were used to reveal afferent communication patterns of target cells and efferent communication patterns of secretory cells. We selected the communication patterns of SPP1 + macrophages for visualization to demonstrate the changes in communication between SPP1 + macrophages and TBS2 + fibroblasts.

WGCNA analysis

The "WGCNA" package [54] was used to screen hub genes that were significantly associated with *THBS2* expression level. The expression profiles of the top 30% of the variance in the TCGA-COAD cohort were the input. According to our previous study, a soft threshold was then determined, an adjacency matrix was clustered, and a hub module was determined. The strongest positive correlations were selected for further analysis by calculating Pearson's correlation coefficient between the modules and *THBS2* expression level. We then measured gene significance (GS) for each gene's traits and module membership (MM) in the hub module. Genes in the module were screened for relation to *THBS2* using MM > 0.6 and GS > 0.6 as thresholds.

GSEA and GSVA analysis

The h.all.v7.4.symbols gene set was downloaded from the MSigDB database for enrichment analysis using the "GSVA" package [55, 56]; an adj.*P* value < 0.05 was considered statistically significant. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses were conducted using the "clusterProfiler" package; an adj.*P*-value < 0.05 and adj.q-value < 0.05 were considered statistically significant.

ICB response prediction

The immunotherapy prediction results presented in the study, including comparisons between pre-treatment and post-treatment patients and between responsive and non-responsive patients, were derived from the ICBatlas database; $P < 0.05$ was considered significant. The results were presented using the R package ggplot2. The TIDE algorithm [57] was used to model distinct tumor immune evasion mechanisms, with higher TIDE scores predicting immune evasion and poor immunotherapy response; we used the TIDE algorithm to measure the effectiveness of the CSGDS generated in this study.

Statistical methods and presentation

All statistical analyses used R software (v.4.0. 1); $p < 0.05$ was considered statistically significant.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12876-025-03918-8.

Supplementary Material 1.
Supplementary Material 2.
Supplementary Material 3.
Supplementary Material 4.
Supplementary Material 5.
Supplementary Material 6.

Acknowledgements

We thank Professor Ye Youqiong from Shanghai Jiao Tong University for providing us with spatial transcriptome sequencing resources.We thank Lisa Oberding, MSc, from Liwen Bianji (Edanz) (www.liwenbianji.cn) for editing the English text of a draft of this manuscript.

Authors' contributions

XY,KX and LW designed the study and wrote the manuscript/revised manuscript. SC and ZJ developed the methodology and performed the analyses. YL and LW collected and analyzed the clinical data. LW,CL and WS performed the statistical analysis. SC,SX and WS performed the bioinformatics analysis. The authors read and approved the final manuscript.

Funding

This work was supported by the National Natural Science Foundation of China (no. 81972288).

Data availability

All data generated or analyzed during this study have been included in this manuscript.

Declarations

Ethics approval and consent to participate

This project was conducted in accordance with the Declaration of Helsinki and was approved by the research ethics committees of Nanjing Drum Tower Hospital (approval ID: 2020–378-01) and Nanjing Medical University (approval ID: (2016)640). All of the patients signed the written informed consent form before enrolled in this study.

Consent for publication

Not applicable.

Competing interests
The authors declare no competing interests.

Received: 4 July 2024 Accepted: 21 April 2025
Published online: 08 May 2025

References

1. Sung H, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 2021;71:209–49. https://doi.org/10.3322/caac.21660.
2. Shin AE, Giancotti FG, Rustgi AK. Metastatic colorectal cancer: mechanisms and emerging therapeutics. Trends Pharmacol Sci. 2023;44:222–36. https://doi.org/10.1016/j.tips.2023.01.003.
3. Zhao H, et al. Wnt signaling in colorectal cancer: pathogenic role and therapeutic target. Mol Cancer. 2022;21:144. https://doi.org/10.1186/s12943-022-01616-7.
4. Du L, et al. Targeting stemness of cancer stem cells to fight colorectal cancers. Semin Cancer Biol. 2022;82:150–61. https://doi.org/10.1016/j.semcancer.2021.02.012.
5. Schmitt M, Greten FR. The inflammatory pathogenesis of colorectal cancer. Nat Rev Immunol. 2021;21:653–67. https://doi.org/10.1038/s41577-021-00534-x.
6. Ganesh K. Optimizing immunotherapy for colorectal cancer. Nat Rev Gastroenterol Hepatol. 2022;19:93–4. https://doi.org/10.1038/s41575-021-00569-4.
7. Keum N, Giovannucci E. Global burden of colorectal cancer: emerging trends, risk factors and prevention strategies. Nat Rev Gastroenterol Hepatol. 2019;16:713–32. https://doi.org/10.1038/s41575-019-0189-8.
8. Zhao W, et al. Colorectal cancer immunotherapy-Recent progress and future directions. Cancer Lett. 2022;545:215816. https://doi.org/10.1016/j.canlet.2022.215816.
9. Ganesh K, et al. Immunotherapy in colorectal cancer: rationale, challenges and potential. Nat Rev Gastroenterol Hepatol. 2019;16:361–75. https://doi.org/10.1038/s41575-019-0126-x.
10. Lizardo DY, et al. Immunotherapy efficacy on mismatch repair-deficient colorectal cancer: From bench to bedside. Biochim et Biophys Acta Rev Cancer. 2020;1874:188447.
11. Guinney J, et al. The consensus molecular subtypes of colorectal cancer. Nat Med. 2015;21:1350–6. https://doi.org/10.1038/nm.3967.
12. Küçükköse E, et al. KIT promotes tumor stroma formation and counteracts tumor-suppressive TGFβ signaling in colorectal cancer. Cell Death Dis. 2022;13:617. https://doi.org/10.1038/s41419-022-05078-z.
13. Ros XR, Vermeulen L. Turning Cold Tumors Hot by Blocking TGF-β. Trends in cancer. 2018;4:335–7. https://doi.org/10.1016/j.trecan.2018.03.005.
14. van den Bulk J, et al. Neoantigen-specific immunity in low mutation burden colorectal cancers of the consensus molecular subtype 4. Genome medicine. 2019;11:87. https://doi.org/10.1186/s13073-019-0697-8.
15. Xu M, Zhang T, Xia R, Wei Y, Wei X. Targeting the tumor stroma for cancer therapy. Mol Cancer. 2022;21:208. https://doi.org/10.1186/s12943-022-01670-1.
16. Hinshaw DC, Shevde LA. The Tumor Microenvironment Innately Modulates Cancer Progression. Can Res. 2019;79:4557–66. https://doi.org/10.1158/0008-5472.Can-18-3962.
17. Tsoumakidou M. The advent of immune stimulating CAFs in cancer. Nat Rev Cancer. 2023;23:258–69. https://doi.org/10.1038/s41568-023-00549-7.
18. Pei L, et al. Roles of cancer-associated fibroblasts (CAFs) in anti- PD-1/PD-L1 immunotherapy for solid cancers. Mol Cancer. 2023;22:29. https://doi.org/10.1186/s12943-023-01731-z.
19. Lei Y, et al. Applications of single-cell sequencing in cancer research: progress and perspectives. J Hematol Oncol. 2021;14:91. https://doi.org/10.1186/s13045-021-01105-2.
20. Zhang Y, et al. Single-cell RNA sequencing in cancer research. J Exp Clin Cancer Res. 2021;40:81. https://doi.org/10.1186/s13046-021-01874-1.
21. Rao A, Barkley D, França GS, Yanai I. Exploring tissue architecture using spatial transcriptomics. Nature. 2021;596:211–20. https://doi.org/10.1038/s41586-021-03634-9.
22. Liu Y, et al. Identification of a tumour immune barrier in the HCC microenvironment that determines the efficacy of immunotherapy. J Hepatol. 2023;78:770–82. https://doi.org/10.1016/j.jhep.2023.01.011.
23. Qi J, et al. Single-cell and spatial analysis reveal interaction of FAP(+) fibroblasts and SPP1(+) macrophages in colorectal cancer. Nat Commun. 2022;13:1742. https://doi.org/10.1038/s41467-022-29366-6.
24. Qian J, et al. A pan-cancer blueprint of the heterogeneous tumor microenvironment revealed by single-cell profiling. Cell Res. 2020;30:745–62. https://doi.org/10.1038/s41422-020-0355-0.
25. Lee HO, et al. Lineage-dependent gene expression programs influence the immune landscape of colorectal cancer. Nat Genet. 2020;52:594–603. https://doi.org/10.1038/s41588-020-0636-z.
26. Cañellas-Socias A, et al. Metastatic recurrence in colorectal cancer arises from residual EMP1(+) cells. Nature. 2022;611:603–13. https://doi.org/10.1038/s41586-022-05402-9.
27. van Dijk D, et al. Recovering Gene Interactions from Single-Cell Data Using Data Diffusion. Cell. 2018;174:716-729.e727. https://doi.org/10.1016/j.cell.2018.05.061.
28. Wu Y, et al. Spatiotemporal Immune Landscape of Colorectal Cancer Liver Metastasis at Single-Cell Level. Cancer Discov. 2022;12:134–53. https://doi.org/10.1158/2159-8290.Cd-21-0316.
29. Gillespie M, et al. The reactome pathway knowledgebase 2022. Nucleic Acids Res. 2022;50:D687-d692. https://doi.org/10.1093/nar/gkab1028.
30. Liberzon A, et al. The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst. 2015;1:417–25. https://doi.org/10.1016/j.cels.2015.12.004.
31. Zhu X, et al. Tumor-associated macrophage-specific CD155 contributes to M2-phenotype transition, immunosuppression, and tumor progression in colorectal cancer. J Immunother Cancer. 2022;10:e004219. https://doi.org/10.1136/jitc-2021-004219.
32. Bao X, et al. An immunometabolism subtyping system identifies S100A9(+) macrophage as an immune therapeutic target in colorectal cancer based on multiomics analysis. Cell reports Medicine. 2023;4:100987. https://doi.org/10.1016/j.xcrm.2023.100987.
33. Rusk N. Expanded CIBERSORTx. Nat Methods. 2019;16:577. https://doi.org/10.1038/s41592-019-0486-8.
34. Chow A, Perica K, Klebanoff CA, Wolchok JD. Clinical implications of T cell exhaustion for cancer immunotherapy. Nat Rev Clin Oncol. 2022;19:775–90. https://doi.org/10.1038/s41571-022-00689-z.
35. Bell HN, et al. Microenvironmental ammonia enhances T cell exhaustion in colorectal cancer. Cell Metab. 2023;35:134-149.e136. https://doi.org/10.1016/j.cmet.2022.11.013.
36. Jasso GJ, et al. Colon stroma mediates an inflammation-driven fibroblastic response controlling matrix remodeling and healing. PLoS Biol. 2022;20:e3001532. https://doi.org/10.1371/journal.pbio.3001532.
37. Pelka K, et al. Spatially organized multicellular immune hubs in human colorectal cancer. Cell. 2021;184:4734-4752.e4720. https://doi.org/10.1016/j.cell.2021.08.003.
38. Feig C, et al. Targeting CXCL12 from FAP-expressing carcinoma-associated fibroblasts synergizes with anti-PD-L1 immunotherapy in pancreatic cancer. Proc Natl Acad Sci USA. 2013;110:20212–7. https://doi.org/10.1073/pnas.1320318110.
39. Joanito I, et al. Single-cell and bulk transcriptome sequencing identifies two epithelial tumor cell states and refines the consensus molecular classification of colorectal cancer. Nat Genet. 2022;54:963–75. https://doi.org/10.1038/s41588-022-01100-4.
40. Yang M, et al. ICBatlas: A Comprehensive Resource for Depicting Immune Checkpoint Blockade Therapy Characteristics from Transcriptome Profiles. Cancer Immunol Res. 2022;10:1398–406. https://doi.org/10.1158/2326-6066.Cir-22-0249.
41. Xie YH, Chen YX, Fang JY. Comprehensive review of targeted therapy for colorectal cancer. Signal Transduct Target Ther. 2020;5:22. https://doi.org/10.1038/s41392-020-0116-z.
42. Pezeshkian Z, et al. Insights into the Role of Matrix Metalloproteinases in Precancerous Conditions and in Colorectal Cancer. Cancers. 2021;13:6226. https://doi.org/10.3390/cancers13246226.
43. Collisson EA, Bailey P, Chang DK, Biankin AV. Molecular subtypes of pancreatic cancer. Nat Rev Gastroenterol Hepatol. 2019;16:207–20. https://doi.org/10.1038/s41575-019-0109-y.

44. Lehmann BD, et al. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. J Clin Investig. 2011;121:2750–67. https://doi.org/10.1172/jci45014.
45. Stintzing S, et al. Consensus molecular subgroups (CMS) of colorectal cancer (CRC) and first-line efficacy of FOLFIRI plus cetuximab or bevacizumab in the FIRE3 (AIO KRK-0306) trial. Ann Oncol. 2019;30:1796–803. https://doi.org/10.1093/annonc/mdz387.
46. Lenz HJ, et al. Impact of Consensus Molecular Subtype on Survival in Patients With Metastatic Colorectal Cancer: Results From CALGB/SWOG 80405 (Alliance). J Clin Oncol. 2019;37:1876–85. https://doi.org/10.1200/jco.18.02258.
47. Parikh AR, et al. Radiation therapy enhances immunotherapy response in microsatellite stable colorectal and pancreatic adenocarcinoma in a phase II trial. Nat Cancer. 2021;2:1124–35. https://doi.org/10.1038/s43018-021-00269-7.
48. Yang H, et al. Multi-scale integrative analyses identify THBS2(+) cancer-associated fibroblasts as a key orchestrator promoting aggressiveness in early-stage lung adenocarcinoma. Theranostics. 2022;12:3104–30. https://doi.org/10.7150/thno.69590.
49. Nan P, et al. Tumor-stroma TGF-β1-THBS2 feedback circuit drives pancreatic ductal adenocarcinoma progression via integrin α(v)β(3)/CD36-mediated activation of the MAPK pathway. Cancer Lett. 2022;528:59–75. https://doi.org/10.1016/j.canlet.2021.12.025.
50. Hao Y, et al. Integrated analysis of multimodal single-cell data. Cell. 2021;184:3573-3587.e3529. https://doi.org/10.1016/j.cell.2021.04.048.
51. Korsunsky I, et al. Fast, sensitive and accurate integration of single-cell data with Harmony. Nat Methods. 2019;16:1289–96. https://doi.org/10.1038/s41592-019-0619-0.
52. Trapnell C, et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat Biotechnol. 2014;32:381–6. https://doi.org/10.1038/nbt.2859.
53. Jin S, et al. Inference and analysis of cell-cell communication using Cell Chat. Nat Commun. 2021;12:1088. https://doi.org/10.1038/s41467-021-21246-9.
54. Kakati T, Bhattacharyya DK, Barah P, Kalita JK. Comparison of Methods for Differential Co-expression Analysis for Disease Biomarker Prediction. Comput Biol Med. 2019;113:103380. https://doi.org/10.1016/j.compbiomed.2019.103380.
55. Foroutan M, et al. Single sample scoring of molecular phenotypes. BMC Bioinformatics. 2018;19:404. https://doi.org/10.1186/s12859-018-2435-4.
56. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinformatics. 2013;14:7. https://doi.org/10.1186/1471-2105-14-7.
57. Jiang P, et al. Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. Nat Med. 2018;24:1550–8. https://doi.org/10.1038/s41591-018-0136-1.

Publisher's Note