Cell & Bioscience

RESEARCH

Open Access

Comprehensive analysis of the critical role of the epithelial mesenchymal transition subtype - TAGLN-positive fibroblasts in colorectal cancer progression and immunosuppression

Junli Zhang[1,2,3], Xinxin Jin[3], Yachao Hou[3], Biao Gu[1], Hongwei Li[1], Li Yi[1], Wenjuan Wu[3,5*] and Shangshang Hu[1,4,5*]

Abstract

Epithelial-mesenchymal transition (EMT) plays a pivotal role in tumor metastasis and immune suppression in colorectal cancer (CRC). However, the specific mechanisms of EMT and its relationship with the clinical prognosis and immunotherapy response in CRC patients remain unclear. In this study, we identified TAGLN-positive fibroblasts (TAGLN$^+$Fib) as a cancer-associated fibroblast (CAF) subtype within the tumor microenvironment (TME) that promotes tumor metastasis and immune evasion. High EMT scores, strongly associated with TAGLN expression, were correlated with advanced tumor stages, poor prognosis, and resistance to immunotherapy. Functional experiments demonstrated that TAGLN knockdown significantly reduced CRC cell proliferation, migration, and EMT phenotypes in vitro and suppressed tumor growth in vivo. Furthermore, TAGLN$^+$Fib closely interacted with MMP7-positive tumor epithelial cells and SPP1-positive macrophages, forming a pro-metastatic and immunosuppressive network. An EMT-TME risk model constructed using TAGLN$^+$Fib exhibited robust predictive power for CRC prognosis and immunotherapy response. This study reveals the association of EMT scores with CRC prognosis and immunotherapy response, highlights TAGLN$^+$Fib's critical role in tumor progression, and develops an EMT-TME risk model, offering insights for personalized CRC treatment and precision medicine.

Keywords Epithelial-mesenchymal transition, TAGLN, Cancer-associated fibroblasts, Immunosuppression

*Correspondence:
Wenjuan Wu
wuwj_2012@126.com
Shangshang Hu
230229079@seu.edu.cn
[1]Department of Blood transfusion, The Third People's Hospital of Bengbu Affiliated to Bengbu Medical University, No. 38 Shengli Road, Bengshan District, Bengbu City, Anhui Province, China
[2]Anhui Provincial Key Laboratory of Tumor Evolution and Intelligent Diagnosis and Treatment, Anhui, China
[3]Bengbu Medical University Key Laboratory of Cancer Research and Clinical Laboratory Diagnosis, Bengbu Medical University, Bengbu, Anhui 233030, China
[4]School of Medicine, Southeast University, Nanjing, Jiangsu 210009, China
[5]Department of Biochemistry and Molecular Biology, School of Laboratory Medicine, Bengbu Medical University, Bengbu, Anhui 233030, China

Introduction

Colorectal cancer (CRC) is among the most common malignant tumors of the digestive system worldwide, characterized by persistently high incidence and mortality rates [1]. Despite advancements in surgery, chemotherapy, targeted therapy, and immunotherapy in recent years, survival outcomes for late-stage CRC patients remain poor, with significant heterogeneity in responses to immunotherapy [2]. This variability is thought to stem from the complex tumor microenvironment (TME) of CRC, within which epithelial-mesenchymal transition (EMT) plays a pivotal role in tumor progression and immune evasion [3].

EMT represents a dynamic process where tumor cells transition from an epithelial to a mesenchymal phenotype, conferring enhanced invasiveness and migratory capabilities. As a critical driver of metastasis [4], EMT also remodels the TME, disrupting interactions between tumor cells and the immune system. This reprogramming fosters an immunosuppressive state that impairs the efficacy of immunotherapies, such as anti-PD-1/PD-L1 treatments [5–7]. While the central role of EMT is widely recognized, its precise mechanisms underlying CRC prognosis, immune evasion, and response to immunotherapy remain to be elucidated systematically.

Cancer-associated fibroblasts (CAFs) are key components of the TME, exerting multifaceted influences on tumor progression [8]. CAFs promote tumor cell proliferation, invasion, and metastasis through the secretion of cytokines and extracellular matrix (ECM) components. Additionally, they modulate the immune microenvironment to establish an immunosuppressive milieu [9]. Recent studies have revealed the heterogeneity of CAFs, with distinct subtypes playing diverse roles in tumor progression. Certain CAF subtypes are hypothesized to be closely linked with the EMT process [10]. However, the specific functions and molecular mechanisms of different CAF subtypes in CRC remain incompletely understood.

This study conducted a multidimensional analysis of multiple bulk and single-cell transcriptomic datasets, combined with in vitro and in vivo functional experiments, to systematically elucidate the relationship between EMT scores and CRC prognosis and immunotherapy response. It further explored the critical role of the EMT subtype TAGLN^{+}Fib in tumor metastasis and immune suppression. Building on these findings, machine learning algorithms were employed to develop a risk prognostic model based on EMT TME scores, enabling stratification of CRC patients by prognostic risk and immunotherapy sensitivity. These results not only offer novel insights into EMT-related basic research but also identify potential biomarkers and therapeutic targets for personalized treatment and precision medicine in CRC.

Materials and methods

Cell culture

The CRC cell line SW620 (purchased from the Shanghai Cell Bank) was cultured in RPMI-1640 medium (Gibco) supplemented with 10% fetal bovine serum (FBS, Gibco) and 1% penicillin-streptomycin at 37 °C in a humidified incubator with 5% CO_2. CAFs and normal fibroblasts (NFs) were isolated and cultured following the protocols described in our previous study [11].

CRISPR-Cas9-mediated TAGLN knockdown

The TAGLN gene in CAFs was knocked out using the CRISPR-Cas9 system. Single-guide RNA (sgRNA) sequences were designed using the CRISPOR online tool to target the exonic region of the TAGLN gene. The sgRNA sequences were cloned into the px458 vector, which contains the Cas9 and EGFP genes (Addgene, plasmid #48138). Transfection was performed using Lipofectamine 3000 (Thermo Fisher). After 48 h, EGFP-positive cells were sorted by flow cytometry, and TAGLN-knockdown clones were selected. Western blot analysis was used to verify knockdown efficiency. The sequences were as follows:

sgRNA_1-F: CACCGGTACCCTGATGGCTCCAAGC,
sgRNA_1-R: AAACGCTTGGAGCCATCAGGGTACC;
sgRNA_2-F: CACCGATGTGGGCCGCCCAGACCGT,
sgRNA_2-R: AAACACGGTCTGGGCGGCCCACATC;
sgRNA_3-F: CACCGAATCGAGAAGAAGTATGACG,
sgRNA_3-R: AAACCGTCATACTTCTTCTCGATTC.

In vitro assays

The protocols for total protein extraction, Western blotting, immunohistochemistry (IHC), cell proliferation assay (CCK-8), colony formation assay, and Transwell migration and invasion assays were performed as described in a previous study [12]. The primary antibodies used in this study included TAGLN (Abclonal, cat. A25499PM), E-Cadherin (Abclonal, cat. A20798), N-Cadherin (Abclonal, cat. A3045), Vimentin (Abclonal, cat. A19607), and GAPDH (Proteintech, cat. #60004-1-IG). Secondary antibodies were goat anti-mouse IgG and goat anti-rabbit IgG (Biosharp).

Xenograft tumor model

Six-week-old male BALB/c nude mice were randomly divided into three groups ($n = 4$ per group). CAFs were mixed with SW620 cells at a 1:4 ratio, and 200 μL of the cell suspension (containing 1×10^7 cells) was subcutaneously injected into each mouse. Tumor volume was measured every three days. After five weeks, the mice were

euthanized, and tumors were excised for weight measurement and histological analysis. All animal experiments were approved by the Institutional Animal Ethics Committee and performed in compliance with animal welfare guidelines.

RT-qPCR analysis of M2 macrophage marker expression in mouse tumor tissues

In this study, RT-qPCR was employed to assess the mRNA expression levels of M2 macrophage-related markers (Arg1, CD206, Ym1, and Fizz1) in mouse tumor tissues. Specifically, total RNA was extracted from tumor tissues using the TRIzol method, followed by cDNA synthesis via reverse transcription. Real-time quantitative PCR was then performed using the SYBR Green fluorescence dye method. Primer sequences for each gene are provided in Supplementary Table 1. Gapdh was used as the internal control, and the relative expression levels of the target genes were calculated using the 2^-ΔΔCt method.

Tissue sample collection

Seven pairs of CRC tumor tissues and adjacent non-cancerous tissues were collected as formalin-fixed paraffin-embedded (FFPE) samples to assess TAGLN protein expression. Ethical approval for the use of human specimens was obtained from the Ethics Committee of Nanjing First Hospital, Nanjing Medical University, and written informed consent was obtained from all patients.

Transcriptomic data collection and analysis for CRC

Six bulk transcriptome datasets with survival information were collected from the TCGA and GEO databases, including TCGA-CRC (TCGA-COAD + TCGA-READ, referred to as bData_1), GSE39582 [13] (bData_2), and the integrated datasets of GSE72970, GSE14333 [14], GSE17538 [15], and GSE29621 [16], collectively referred to as bData_3. Batch effect correction and integration were performed using the "sva" R package. Four CRC single-cell transcriptome datasets were integrated using the "harmony" R package [17], with data quality control performed using the "Seurat" R package [18]. Additionally, four CRC spatial transcriptomics (ST) datasets from Wu et al. were utilized [19]. Cell-cell communication analysis was conducted using the "CellChat" R package [20]. High-dimensional weighted gene co-expression network analysis (hdWGCNA) was performed with the "hdWGCNA" R package [21], while conventional WGCNA analysis employed the "WGCNA" R package [22]. Survival analysis, including overall survival (OS) and relapse-free survival (RFS), was performed using the "survival" and "survminer" R packages. Gene Set Variation Analysis (GSVA) was conducted using the "GSVA" R package, employing the ssGSEA algorithm for gene set scoring. Gene Set Enrichment Analysis (GSEA), Gene Ontology (GO), and KEGG pathway analysis were executed using the "clusterProfiler" R package. The "KEGG" and "HALLMARK" gene sets were retrieved from the Molecular Signatures Database [23] (MSD). The gene set used to calculate the EMT score in this study was derived from the "HALLMARK" gene set. Immunohistochemical analysis of TAGLN expression in CRC tissues was conducted using The Human Protein Atlas [24]. The detailed methodologies for single-cell and spatial transcriptomic data processing, cell annotation, cell-cell communication analysis, hdWGCNA, and WGCNA are described in a previous study [25]. Based on the CIBERSORT algorithm, the infiltration levels of 22 immune cell types in bulk sequencing data were deconvoluted [26]. The CPTAC (Clinical Proteomic Tumor Analysis Consortium) database (https://www.pdc.cancer.gov/pdc/browse/filters/disease_type:Colon-20Adenocarcinoma) was used to analyze the differences in TAGLN protein expression between normal colon tissues and CRC tissues.

Immunotherapy analysis

In this study, the Tumor Immune Dysfunction and Exclusion (TIDE) database was utilized to evaluate immune dysfunction and exclusion scores in CRC samples. This analysis aimed to investigate the mechanisms of immune evasion in CRC and their correlation with EMT scores. The TIDE algorithm calculates two critical components: T-cell dysfunction in tumors with high infiltration of cytotoxic T lymphocytes (CTLs) and T-cell exclusion in tumors with low CTL infiltration. Additionally, TIDE predicts the response of CRC patients to immune checkpoint blockade therapy, specifically anti-PD-L1 treatment [27]. Furthermore, the study analyzed the Imvigor210 cohort, a publicly available dataset comprising urothelial carcinoma patients treated with anti-PD-L1 therapy (atezolizumab) [28].

Immune cell infiltration analysis

This study employed three immune infiltration assessment algorithms—Xcell, EPIC, and MCPcounter—to quantitatively analyze the composition of immune and stromal cells in the TME of CRC based on bulk transcriptomic data. The corresponding R packages, "xCell," "EPIC," and "MCPcounter," were used for these analyses.

Consensus molecular subtypes (CMSs)

The CMS classification proposed by Justin Guinney et al. in 2015, categorizing CRC samples into four subtypes—CMS1 (immune subtype), CMS2 (classical subtype), CMS3 (metabolic subtype), and CMS4 (mesenchymal subtype)—was applied in this study [29]. Subtype assignment for the CRC cohorts was performed using the

"Lothelab/CMScaller" R package, which classifies patient samples based on TCGA CRC data.

Construction and comparison of prognostic risk models

The prognostic risk model in this study was developed using the leave-one-out cross-validation (LOOCV) framework established by Zaoqu Liu et al. [30]. We systematically searched the PubMed database using the keywords "Colorectal Neoplasms"[Mesh] OR "colorectal cancer" OR "colon cancer" OR "rectal cancer" OR "CRC" AND risk signature, and collected published CRC prognostic gene signatures (inclusion criteria: prognostic models comprised solely of protein-coding genes). Subsequently, the feature genes from these models were matched against the expression profiles from three independent CRC cohorts (bData_1, bData_2, and bData_3) in our study. Finally, using the C-index as an evaluation metric, we performed a comprehensive performance comparison among the published prognostic models, our EMT-TME score, and the EMT-TME risk model.

Statistical analysis

All statistical analyses were conducted using R version 4.3.0. Differential gene expression analysis was performed with the "limma" R package. Correlations were assessed using Spearman's rank correlation test, and group comparisons were conducted using the Wilcoxon rank-sum test. All in vitro experiments were performed in triplicate, with results expressed as mean ± standard deviation. Statistical significance was set at $P < 0.05$ ($^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$).

Results

Association of EMT scores with clinical progression and immune suppression

To investigate the clinical relevance and biological characteristics of EMT, we analyzed three bulk transcriptome datasets of CRC: the TCGA dataset (bData_1), GSE39582 dataset (bData_2), and the integrated dataset comprising GSE72970, GSE14333, GSE17538, and GSE29621 (bData_3) (Fig. 1A). Our findings demonstrated that CRC patients with higher EMT scores exhibited significantly reduced OS and RFS (Fig. 1B). EMT scores were notably lower in patients over 65 years of age and increased progressively with tumor pathological characteristics, including T, N, M stages, and clinical staging (Fig. 1C). HALLMARK gene set enrichment analysis revealed significant enrichment of pathways related to tumor metastasis and immune suppression in the high EMT score group, including IL6_JAK_STAT3_SIGNALING, IL2_STAT5_SIGNALING, HYPOXIA, TGF_BETA_SIGNALING, and KRAS_SIGNALING_UP (Fig. 1D). Additionally, EMT scores were positively correlated with tumor immune dysfunction and exclusion (TIDE) scores (Fig. 1E). Notably, EMT scores were significantly higher in patients resistant to immune therapy (anti-PD-1 treatment) compared to those who were sensitive, a finding validated in the Imvigor210 anti-PD-1 cohort (Fig. 1F).

CAFs are the cell population most closely associated with EMT

To investigate the cell populations closely linked to EMT, this study integrated four independent single-cell transcriptomic datasets, encompassing 47 samples. After quality control, a total of 120,582 cells were retained for analysis (Fig. 2A). Using classical cell-type markers, eight major cell populations were identified: T/NK cells (CD3D), epithelial cells (KRT19), plasma cells (MZB1), myeloid cells (LYZ), B cells (MS4A1), fibroblasts (COL1A1), endothelial cells (PLVAP), and mast cells (CPA3) (Fig. 2B–C; Supplementary Fig. 1). Further analysis revealed a significant increase in EMT scores in tumor samples, with the highest scores observed in late-stage (stage IV) patients (Fig. 2D). Among the identified cell types, fibroblasts exhibited the highest EMT scores, markedly surpassing those of other cell populations (Fig. 2E). By combining these findings with bulk sequencing data and applying three immune infiltration algorithms (Xcell, EPIC, and MCPCount), we found that CAFs showed the strongest correlation with EMT scores (Fig. 2F).

Identification of the EMT subtype TAGLN^{+}Fib

In this study, fibroblast populations were reclustered and stratified into Low-EMT and High-EMT groups based on the median EMT score (Fig. 3A–B). Differential gene analysis between these two groups identified distinct gene sets (Fig. 3C). To pinpoint core regulatory genes in the High-EMT group, high-dimensional weighted gene co-expression network analysis (hdWGCNA) was performed on the fibroblast population (Supplementary Fig. 2A). This analysis revealed 14 co-expression modules (Fig. 3D), among which modules 2, 4, 5, 6, 7, and 8 were significantly enriched in the High-EMT group (Supplementary Fig. 2B–C, Fig. 3E). Further analysis intersected the top 25 hub genes from these six modules with the genes upregulated in the High-EMT group ($Log2FC > 1$, $P < 0.05$), resulting in seven core genes: SPARC, TAGLN, FN1, MYL9, POSTN, SULF1, and ACTA2 (Fig. 3F). All seven genes originate from the M7 module, and based on the KME (Module Eigengene Connectivity) parameter, TAGLN is positioned at a higher core regulatory location (Supplementary Fig. 2D, Fig. 3G). Therefore, TAGLN was identified as the core regulatory gene in the High-EMT group. Expression analysis revealed that TAGLN was significantly upregulated in CAFs compared to normal fibroblasts (NFs) (Fig. 3H). Single-cell and bulk transcriptomic data further demonstrated that

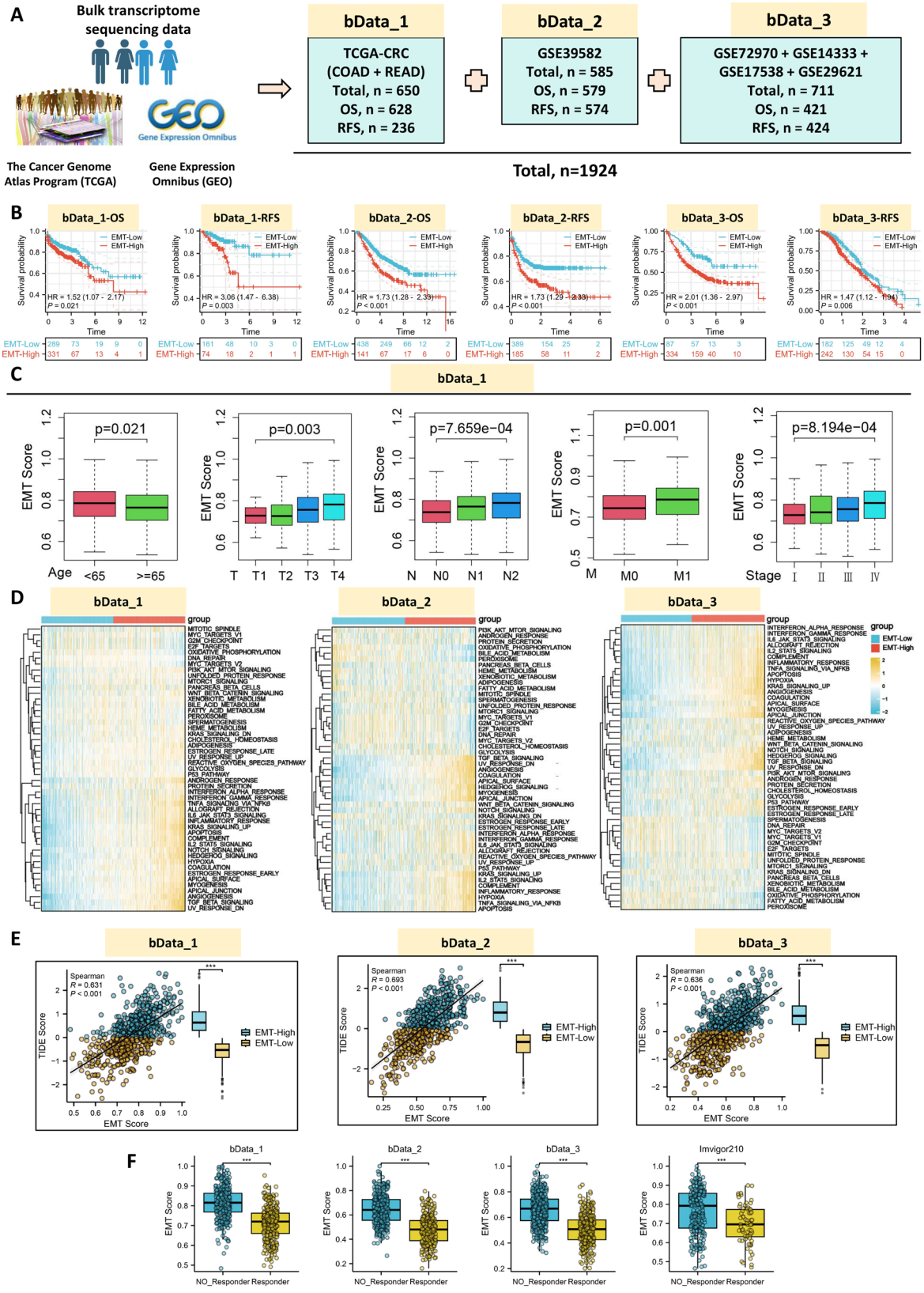

Fig. 1 Association of EMT Scores with Clinical Progression and Immune Suppression. **A.** Definition of three bulk transcriptome datasets (bData_1, bData_2, bData_3). **B**. OS and RFS survival curves stratified by EMT scores. **C**. EMT score differences across clinical parameters (age, T, N, M stages, and clinical staging). **D**. Heatmap of HALLMARK gene set enrichment in high and low EMT score groups. **E**. Correlation of EMT scores with TIDE scores. **F**. Differences in EMT scores between immune therapy-sensitive and -resistant patients

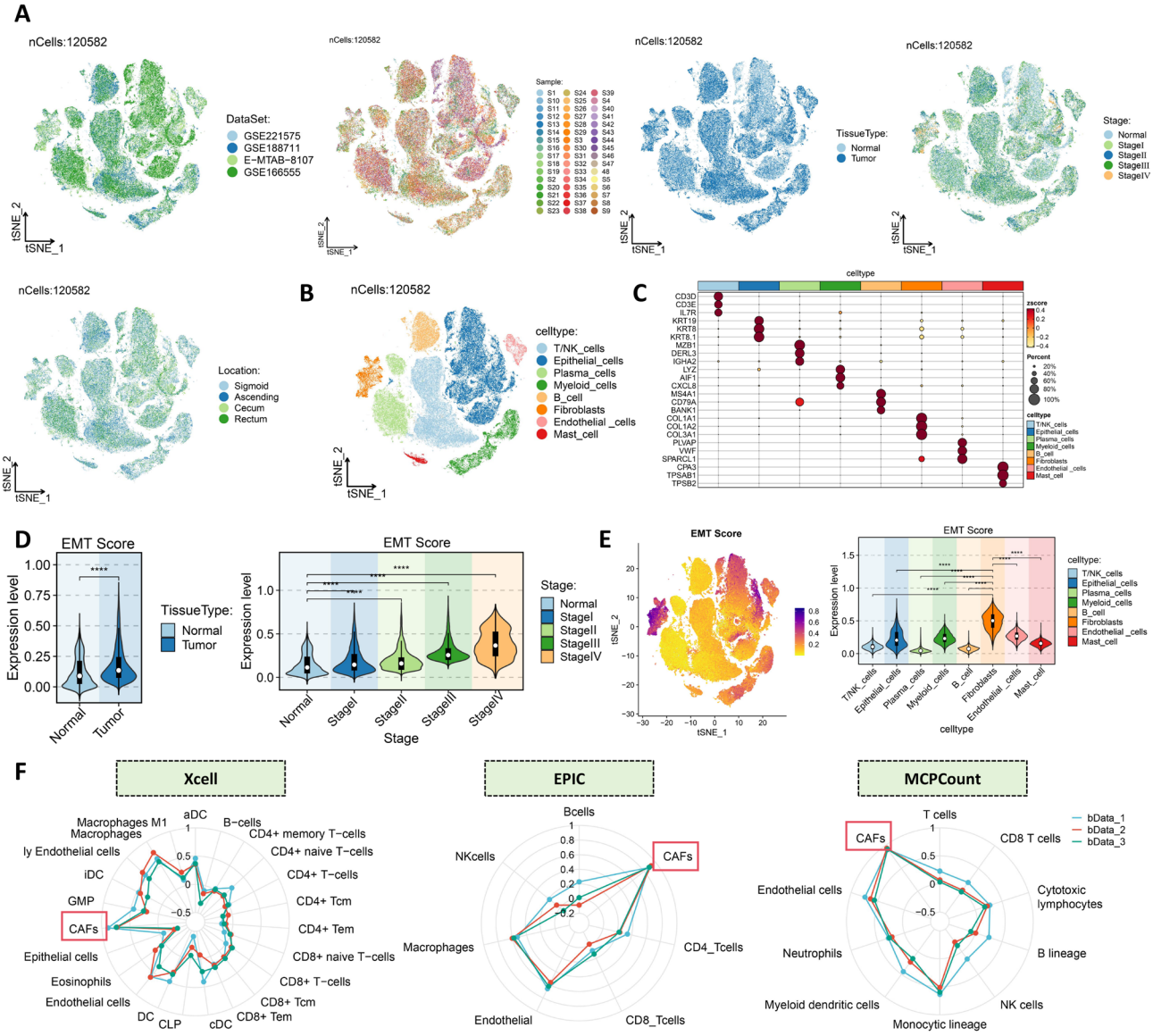

Fig. 2 CAFs Are the Cell Population Most Closely Associated with EMT. **A**. Single-cell transcriptomic landscape of four CRC cohorts, annotated by sample, tissue type, stage, and tumor location. **B**. Single-cell transcriptomic landscape showing eight distinct cell types. **C**. Top three classical markers for each cell type. **D**. EMT score differences across tissue types and disease stages. **E**. Distribution and comparison of EMT scores among cell types. F. Radar chart illustrating correlations between EMT scores and cell types as assessed by Xcell, EPIC, and MCPCount

TAGLN expression was highest in stage IV CRC patients (Fig. 3I, Supplementary Fig. 2E). TAGLN was predominantly expressed in fibroblasts (Fig. 3J), a finding corroborated by immunohistochemical analysis in CRC samples (Fig. 3K). Ruchi Shah et al. demonstrated that ACTA2 (α-SMA), FAP, MMP2, PDPN, and THY are important markers of CAFs [31]. We conducted a correlation analysis between TAGLN and these genes, and in three independent CRC cohorts, we found that TAGLN exhibited a significant positive correlation with all of them (Supplementary Fig. 3A). Furthermore, using spatial transcriptomics data, we observed that TAGLN and the markers ACTA2 (α-SMA), FAP, MMP2, PDPN, and THY1 display distinct spatial colocalization within CRC tissues, suggesting that these CAF markers may collaboratively participate in constructing the tumor microenvironment (Supplementary Fig. 3B). Based on these findings, fibroblasts were classified into TAGLN^{+}Fib and TAGLN^{-}Fib subpopulations (Fig. 3L). TAGLN expression was highly correlated with EMT scores in both subpopulations ($R > 0.8$, $P < 0.05$) (Fig. 3M–N). Spatial transcriptomic data revealed co-localization of TAGLN^{+}Fib and high EMT scores within the TME (Fig. 3O). ACTA2 (α-SMA) and FAP have been widely confirmed as classical markers of CAFs [32, 33]. Finally, multiplex immunofluorescence analysis confirmed the presence of TAGLN^{+}Fib in CRC

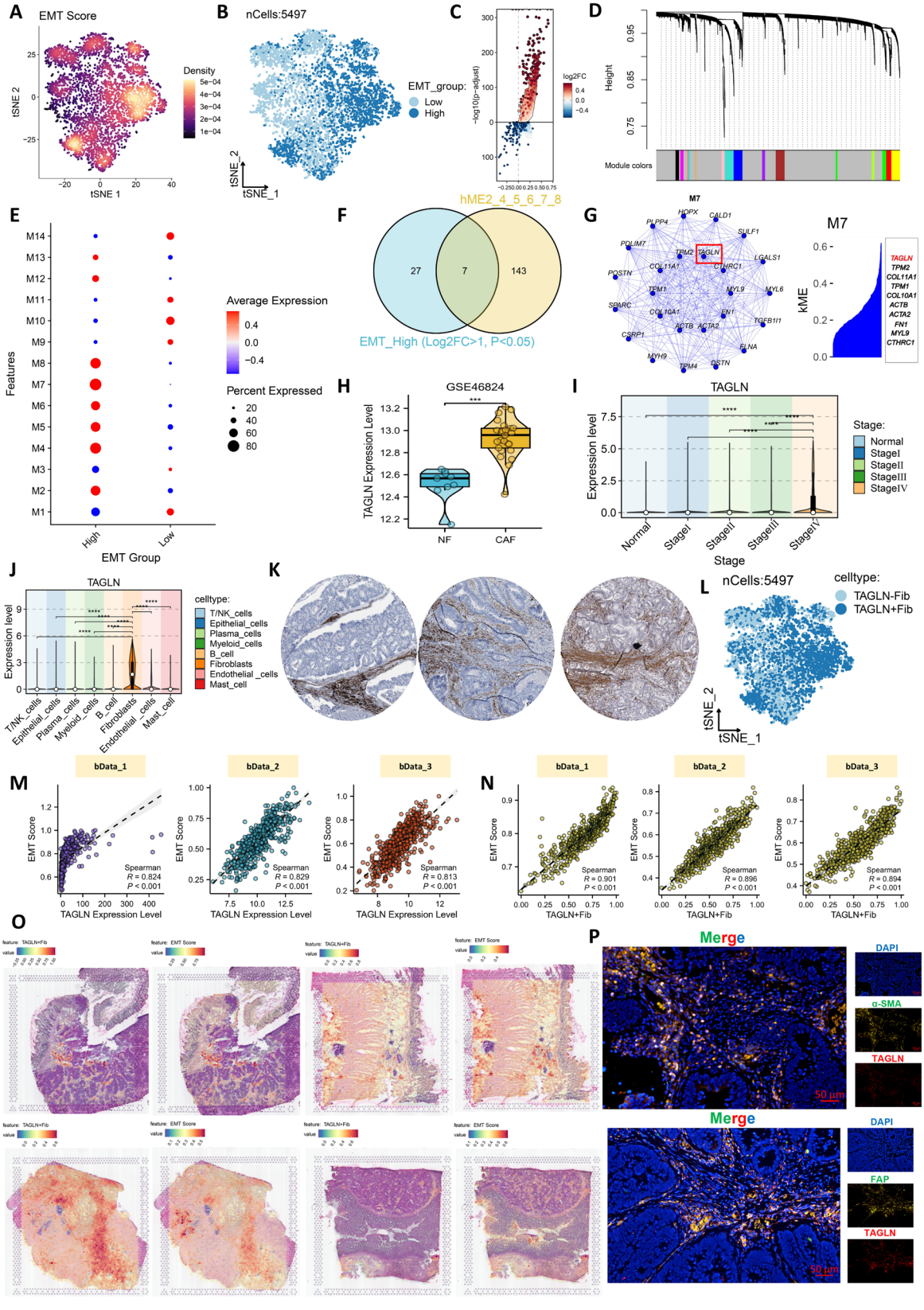

Fig. 3 (See legend on next page.)

(See figure on previous page.)

Fig. 3 Identification of the EMT Subtype TAGLN⁺Fib. **A**. Distribution of EMT scores in fibroblasts. **B**. Stratification of fibroblasts into Low-EMT and High-EMT groups. **C**. Volcano plot showing differentially expressed genes between the two groups. **D**. Hierarchical clustering of co-expression modules in fibroblasts. **E**. Enrichment of 14 co-expression modules in Low-EMT and High-EMT groups. **F**. Venn diagram of the top 25 hub genes from six modules and highly expressed genes in the High-EMT group. **G**. Co-expression network of the top 25 genes in the M7 module and analysis of KME (Module Eigengene Connectivity). **H**. Differential expression of TAGLN in NFs and CAFs. **I**. TAGLN expression across clinical stages in single-cell datasets. **J**. TAGLN expression among eight cell types. **K**. Immunohistochemical analysis of TAGLN in CRC tissues. **L**. Single-cell expression profiles of TAGLN⁺Fib and TAGLN⁻Fib. **M**. Correlation between TAGLN expression and EMT scores. **N**. Correlation between EMT scores and the TAGLN⁺Fib subpopulation. **O**. Spatial transcriptomic co-localization of TAGLN⁺Fib and EMT scores. P. Multiplex immunofluorescence analysis confirming TAGLN⁺Fib, with α-SMA/FAP as a CAF marker

tissues, with α-SMA serving as a CAF marker (Fig. 3P). These results collectively establish TAGLN⁺Fib as a EMT subtype and highlight its pivotal role in CRC progression.

TAGLN⁺Fib promotes EMT in CRC cells

Furthermore, in vitro experiments were conducted to validate the expression levels of TAGLN in adjacent non-tumor tissues and tumor tissues. The results revealed that TAGLN was significantly overexpressed in tumor tissues (Fig. 4A-B), which was further validated using the CPTAC database (Supplementary Fig. 4A). Additionally, TAGLN expression was markedly higher in CAFs than in normal fibroblasts (NFs), as determined using fibroblasts derived from four patients in previous studies (Fig. 4C–D). Subsequently, we utilized CRISPR/Cas9 technology to knock out TAGLN expression in CAFs. Among the tested single-guide RNAs (sgRNAs), sg_1 exhibited the highest knockdown efficiency and was selected for subsequent experiments (Fig. 4E–F). A co-culture system comprising CAFs and CRC cells was then established, dividing the samples into three groups: control (no modification), CAFs with TAGLN knockdown (CAF_Sg-TAGLN), and negative control CAFs (CAF_Sg-NC) (Fig. 4G). In vitro assays, including colony formation, CCK-8 proliferation, and Transwell migration assays, demonstrated that TAGLN knockdown significantly reduced CRC cell proliferation and migration when co-cultured with CAFs (Fig. 4H–J). Importantly, Western blot analysis revealed that TAGLN knockdown markedly suppressed the epithelial-to-mesenchymal transition (EMT) in CRC cells (Fig. 4K). To confirm these findings in vivo, we injected a mixture of CAFs and CRC cells (1:4 ratio) into mice, categorized into three corresponding groups. TAGLN knockdown significantly decreased tumor volume and weight in vivo (Fig. 4L–N). Ki67 staining further validated reduced tumor cell proliferation in the TAGLN knockdown group (Fig. 4O). Similarly, in vivo Western blot analysis confirmed that EMT markers were significantly downregulated in the TAGLN knockdown group (Fig. 4P). To further validate our in vivo and in vitro findings, we analyzed three independent CRC RNA-seq cohorts (bData_1, bData_2, and bData_3). Specifically, based on the median expression of TAGLN, CRC patient data were divided into a TAGLN-high expression group (TAGLN_High) and a TAGLN-low expression group (TAGLN_Low), and genes that were significantly upregulated in the TAGLN_High group were subsequently identified. GSEA results showed that these significantly upregulated genes were highly enriched in the "epithelial-mesenchymal transition" (EMT) pathway across all three cohorts (Supplementary Figs. 4B-D), supporting our hypothesis that TAGLN is involved in the EMT process. The list of differentially expressed genes between the TAGLN high and low expression groups in these cohorts is provided in Supplementary Tables 2, and the detailed GSEA results are shown in Supplementary Table 3. These results collectively indicate that TAGLN⁺Fib plays a pivotal role in promoting CRC cell EMT, proliferation, and metastasis, both in vitro and in vivo.

TAGLN + Fib as a poor prognostic factor in CRC and its association with immunosuppression

This study further explored the clinical relevance and biological characteristics of TAGLN + Fib in CRC. Analysis of bulk RNA-seq datasets with survival information revealed that high infiltration of TAGLN + Fib was significantly associated with reduced OS and RFS in CRC patients (Fig. 5A). Gene Ontology (GO) and KEGG pathway analyses indicated that TAGLN + Fib is primarily linked to extracellular matrix remodeling and immunosuppression (Fig. 5B). HALLMARK gene set analysis demonstrated significant positive correlations between TAGLN + Fib and multiple immunosuppressive pathways, including the TGF-β signaling pathway, TNF-α via NF-κB signaling, hypoxia response, IL6-JAK-STAT3 signaling, and IL2-STAT5 signaling (Fig. 5C, Supplementary Fig. 5). Moreover, TAGLN + Fib infiltration exhibited a strong positive correlation with TIDE scores, indicative of immunosuppressive activity (Fig. 5D). Consensus molecular subtypes (CMS) analysis showed that TAGLN + Fib levels were significantly elevated in the CMS4 subtype, known for its association with stromal activation and poor prognosis (Fig. 5E). This finding further supports the role of TAGLN + Fib in adverse CRC outcomes and immunosuppression. Regarding anti-PD-L1 immunotherapy, CRC patients were stratified into high and low TAGLN + Fib infiltration groups based on the median infiltration level. Results revealed that the high TAGLN + Fib group had a greater proportion of patients resistant to immunotherapy, a finding validated using the Imvigor210 cohort (Figs. 5F–G). We leveraged three independent CRC RNA-seq clinical cohorts

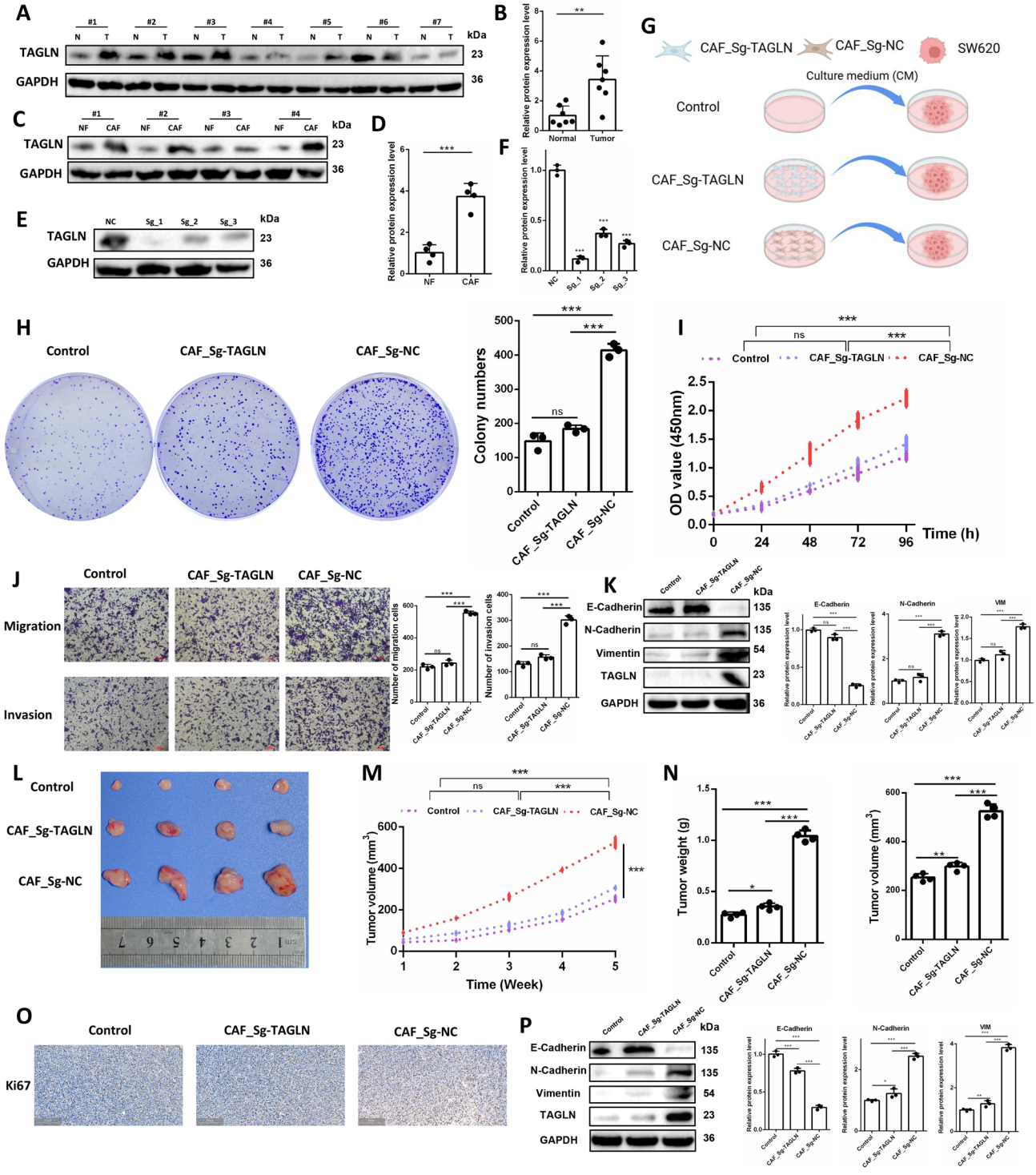

Fig. 4 TAGLN⁺Fib Promotes EMT in CRC Cells. **A**–**B**. TAGLN expression levels in adjacent non-tumor and tumor tissues. **C**–**D**. TAGLN expression in normal fibroblasts (NFs) and CAFs. **E**–**F**. Validation of TAGLN knockdown efficiency in CAFs using CRISPR/Cas9. **G**. Schematic representation of the CAF and CRC co-culture system. H. Results of the colony formation assay. **I**. Outcomes of the CCK-8 proliferation assay. **J**. Transwell migration assay results. **K**. Western blot analysis of EMT-related protein expression in different groups. **L**. In vivo tumorigenesis experiment using three mouse groups. **M**. Tumor volume changes over five weeks in each group. **N**. Differences in tumor volume and weight among the groups. **O**. Immunohistochemistry analysis of Ki67 expression across the groups. **P**. Western blot analysis of EMT-related proteins in tumor tissues from different groups

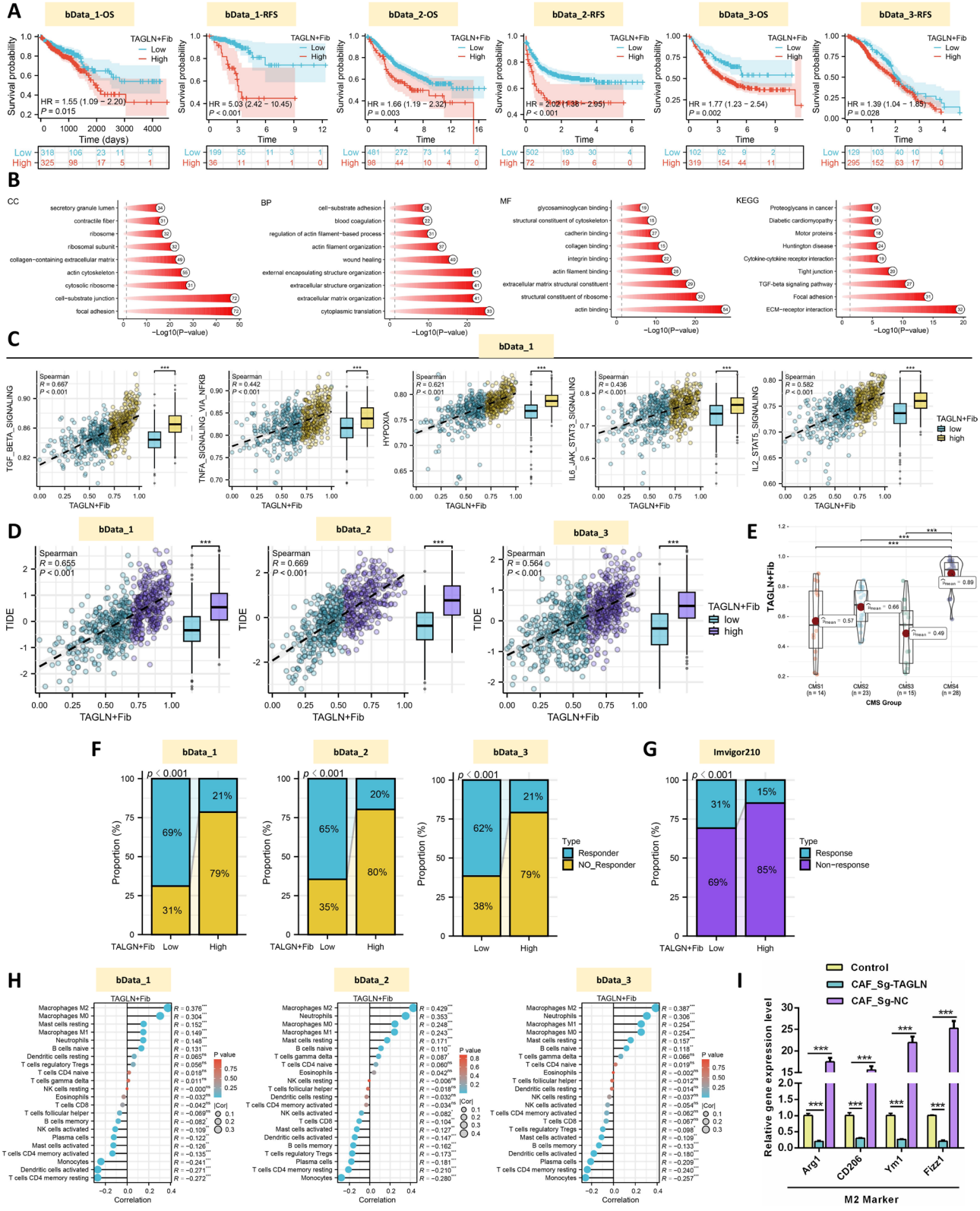

Fig. 5 (See legend on next page.)

(See figure on previous page.)

Fig. 5 TAGLN + Fib as a Poor Prognostic Factor in CRC and Its Association with Immunosuppression. **A**. Survival curves for OS and RFS based on TAGLN + Fib infiltration. **B**. GO and KEGG analyses of TAGLN + Fib-related pathways. **C**. Correlation between TAGLN + Fib and immunosuppressive pathways. **D**. Correlation of TAGLN + Fib with TIDE scores. **E**. TAGLN + Fib levels across CMS subtypes. **F**-**G**. Proportional distribution of immunotherapy response in high and low TAGLN + Fib groups. **H**. Correlation analysis between TAGLN⁺Fib and immune cells in three independent CRC cohorts based on the CIBERSORT algorithm. **I**. RT-qPCR analysis of the relative expression levels of M2-like macrophage markers (Arg1, CD206, Ym1, and Fizz1) in tumors from three mouse groups (Control, CAF_Sg-TAGLN, and CAF_Sg-NC)

(bData_1, bData_2, and bData_3) to further elucidate the correlation between TAGLN⁺Fib infiltration and immune cell infiltration levels in tumor tissues. We found that the degree of TAGLN⁺Fib infiltration was most strongly and significantly positively correlated with M2-like macrophages (Fig. 5H). Previous studies have demonstrated that M2 macrophages are closely associated with tumor immune suppression and resistance to immunotherapy [34]. Consequently, we performed RT-qPCR on tumor tissues from three groups in our mouse model (Control, CAF_Sg-TAGLN, and CAF_Sg-NC) to evaluate the expression changes of key M2 macrophage markers (Arg1, CD206, Ym1, and Fizz1). The results revealed that, compared with the Control group, the expression of M2 macrophage markers was significantly decreased in the CAF_Sg-TAGLN group and significantly increased in the CAF_Sg-NC group (Fig. 5I). These findings further support the immunosuppressive characteristics of TAGLN⁺Fib.

TAGLN + Fib is closely associated with Epi_3 (MMP7 + Epi) and Mye_5 (SPP1 + Macro)

To further explore the interaction between TAGLN + Fib and other cell populations, cellular communication analysis revealed that TAGLN + Fib exhibited stronger signaling output compared to TAGLN-Fib (Fig. 6A, Supplementary Fig. 6A), particularly toward tumor epithelial and myeloid cells (Fig. 6B-C, Supplementary Fig. 6B-E and 7 A-D). Subsequently, epithelial and myeloid cells were re-clustered into nine epithelial subclusters and eight myeloid subclusters, respectively (Fig. 6D). Notably, the Epi_3 (MMP7 + Epi) and Mye_5 (SPP1 + Macro) subclusters displayed significantly elevated EMT scores (Fig. 6E). Gene set variation analysis (GSVA) further indicated that the Epi_3 subcluster exhibited biological features associated with metastasis, whereas the Mye_5 subcluster showed characteristics related to immunosuppression (Figs. 6F-G). Correlation analyses demonstrated a strong positive relationship between both Epi_3 and Mye_5 and EMT scores as well as TAGLN + Fib (Fig. 6H, Supplementary Fig. 8A). Moreover, cellular communication analysis revealed that TAGLN + Fib had the strongest signaling interactions with Epi_3 and Mye_5 compared to other subclusters, exhibiting enhanced signal transmission capabilities (Fig. 6I, Supplementary Figs. 8B-C). To further investigate how TAGLN⁺Fib interacts with MMP7⁺Epi and SPP1⁺Macro, we performed an in-depth analysis using spatial transcriptomics data. The results clearly demonstrated that Epi_3 (MMP7⁺) cells exhibit obvious proximity or colocalization with TAGLN⁺Fib, and Mye_5 (SPP1⁺Macro) cells also show a colocalization pattern with TAGLN⁺Fib (Fig. 6J, Supplementary Figs. 8D and 3O). This close spatial relationship greatly enhances the potential for communication between TAGLN⁺Fib and both MMP7⁺Epi and SPP1⁺Macro cells. These findings establish a close association between TAGLN + Fib and Epi_3 (MMP7 + Epi) as well as Mye_5 (SPP1 + Macro), highlighting the importance of these interactions in the TME.

Development of a CRC prognostic risk model based on EMT TME scores

In previous analyses, TAGLN + Fib was identified as a poor prognostic factor for CRC and exhibited immunosuppressive characteristics. Furthermore, the communication between TAGLN + Fib and Epi_3 (MMP7 + Epi) and Mye_5 (SPP1 + Macro) was particularly significant, with Epi_3 demonstrating metastasis-related features and Mye_5 showing immunosuppressive traits, both displaying high EMT scores. To further elucidate these interactions, this study analyzed the ligands mediating the effects of TAGLN + Fib on Epi_3 and Mye_5. The analysis revealed that in Mye_5, the strongest ligand-receptor interactions involved COL1A2-CD44, COL1A1-CD44, and APP-CD74 (Fig. 7A). Similarly, in Epi_3, the most significant interactions included COL1A2/COL1A1-SCD4/SCD1/CD44/(ITGA3 + ITGB1) (Fig. 7A). Additionally, we found that both Epi_3 and Mye_5 are associated with poor prognosis in CRC (Supplementary Fig. 9A). Using these ligands and the characteristics of the three cell types (TAGLN + Fib, Epi_3, and Mye_5), the study constructed an EMT_TME Score based on the ssGSEA algorithm (Fig. 7B). The EMT_TME Score was found to be closely associated with CRC prognosis and identified as a marker of poor outcomes (Fig. 7C). Furthermore, the ligand signature showed the strongest correlation with the EMT_TME score ($R > 0.9$, $P < 0.05$), followed by TAGLN⁺Fib ($R > 0.8$, $P < 0.05$), Mye_5 ($R > 0.7$, $P < 0.05$), and Epi_3 ($R > 0.6$, $P < 0.05$) (Supplementary Fig. 9B). To further investigate, the EMT_TME Score was stratified into high and low groups based on the median in the bData_1 cohort, followed by WGCNA analysis (Fig. 7D). Among the results, the MEblue module was significantly enriched in the high EMT_TME

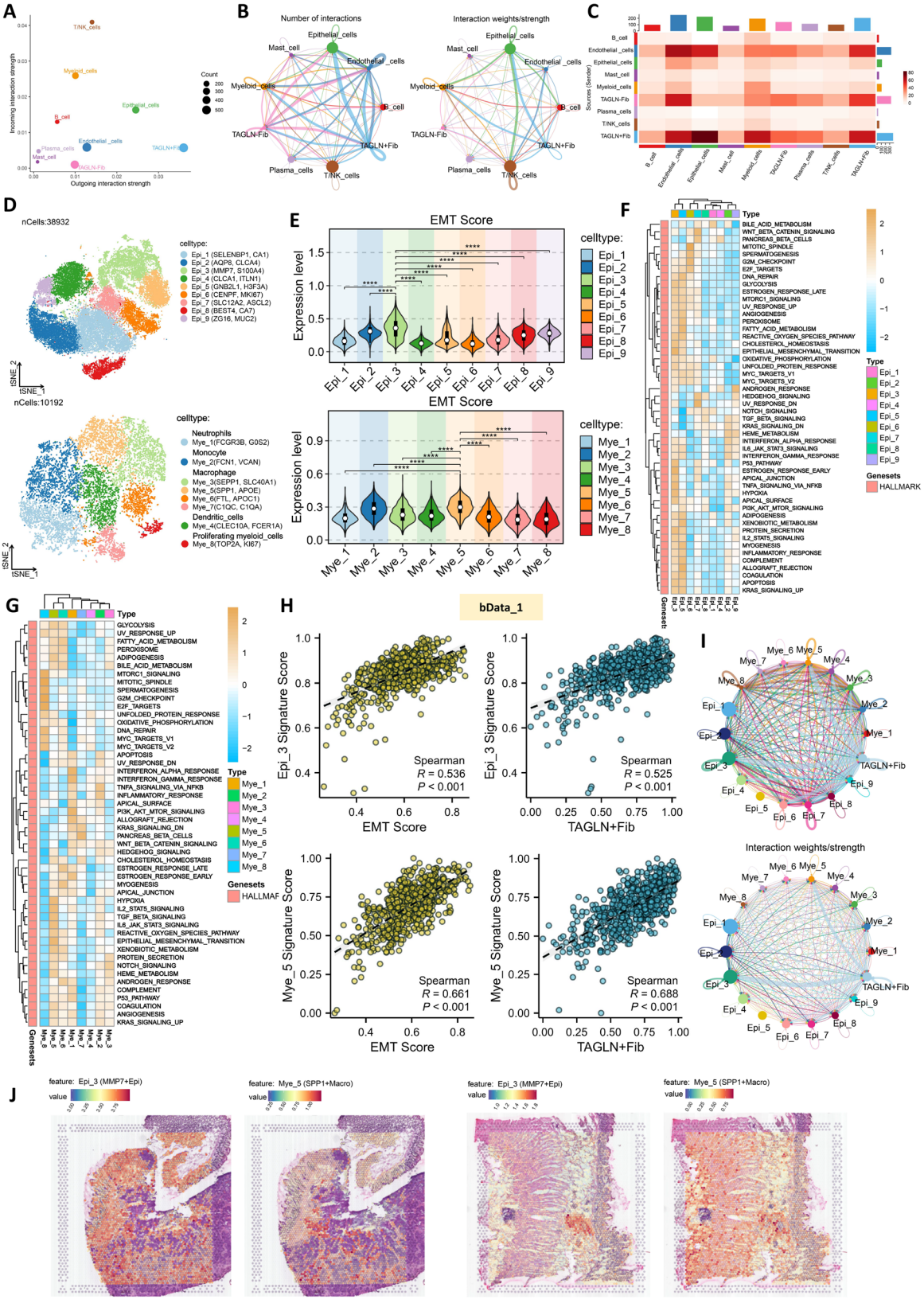

Fig. 6 (See legend on next page.)

(See figure on previous page.)

Fig. 6 TAGLN + Fib is Closely Associated with Epi_3 (MMP7 + Epi) and Mye_5 (SPP1 + Macro). **A**. 2D plot showing the signal reception and transmission strengths of each cell type. **B**. Number and weight of interactions among cell types. **C**. Heatmap illustrating communication strengths between cell types. **D**. Single-cell transcriptomic maps of nine epithelial and eight myeloid subclusters. **E**. Differences in EMT scores across epithelial and myeloid subclusters. **F**-**G**. Heatmaps of GSVA results for epithelial and myeloid subclusters. **H**. Correlation analysis between Epi_3, Mye_5, EMT scores, and TAGLN + Fib. **I**. Communication analysis of TAGLN + Fib with epithelial and myeloid subclusters. **J**. The spatial transcriptomics results demonstrate the spatial localization of Epi_3 (MMP7 + Epi) and Mye_5 (SPP1 + Macro)

Score group (Fig. 7E), and its scores were strongly positively correlated with the EMT_TME Score, with correlation coefficients exceeding 0.9 (Fig. 7F). GO and KEGG analyses of the MEblue module revealed biological characteristics related to extracellular matrix remodeling and immunosuppression (Fig. 7G), highlighting its potential for constructing a CRC prognostic risk model. The study initially used univariate Cox regression analysis to identify genes associated with CRC prognosis. Subsequently, 101 combinations of 10 machine learning algorithms were employed to develop a CRC prognostic risk model using the bData_1 cohort as the training set, with bData_2 and bData_3 serving as validation sets. The Lasso + plsRcox combination achieved the highest average C-index (Fig. 8A). The model included 10 risk genes, with detailed risk coefficients provided in Supplementary Table 4. Using the expression of these risk genes and their respective coefficients, risk scores were calculated for each patient in the training and validation sets. The results showed that higher risk scores were associated with increased mortality in CRC patients, a trend validated in the Meta cohort (integrating bData_1, bData_2, and bData_3) (Fig. 8B). Compared to the low-risk group, the high-risk group exhibited significantly reduced OS and RFS (Fig. 8C). Multivariate Cox regression analysis further confirmed that the risk score was an independent poor prognostic factor for OS and RFS in CRC patients, even when combined with clinical staging (Fig. 8D). In this study, we collected 195 published CRC prognostic gene signatures (Supplementary Table 5). Subsequently, we matched the expression profiles of the feature genes from these 195 models in our three independent CRC cohorts (bData_1, bData_2, and bData_3), among which 126 models had complete gene expression data. We then performed a comprehensive performance comparison using the C-index between these 126 published prognostic models and our EMT-TME score and EMT-TME risk model. The results demonstrated that the EMT-TME score exhibited robust prognostic performance across all three CRC cohorts as well as in the integrated Meta cohort, ranking 2nd in bData_1, 4th in bData_2, 7th in bData_3, and 5th in the Meta cohort (Supplementary Fig. 10). Moreover, our EMT-TME risk model showed outstanding prognostic performance, ranking 2nd in bData_2 and 1st in bData_1, bData_3, and the Meta cohort, indicating excellent prognostic predictive ability and significant clinical potential (Supplementary Fig. 10). Regarding immunotherapy, patients in the high-risk group demonstrated a higher proportion of treatment resistance (Fig. 8E). This finding was validated in the Imvigor cohort, which showed poorer prognosis and a higher proportion of immunotherapy-resistant patients in the high-risk group (Fig. 8F). In summary, this study developed a CRC prognostic risk model based on EMT TME scores. This model effectively stratifies CRC patients by prognostic risk and predicts their sensitivity to immunotherapy, providing a valuable tool for clinical decision-making and potentially broader applications in personalized CRC management.

Biological characteristics of the constructed risk model

The risk model developed in this study exhibits biological features associated with extracellular matrix (ECM) remodeling and immune suppression. To investigate these characteristics, we identified differentially expressed genes between the high- and low-risk groups in the training cohort (bData_1) and validation cohorts (bData_2 and bData_3), subsequently determining their shared genes (Figs. 9A–B). Gene set enrichment analysis (GSEA) based on the HALLMARK and KEGG datasets revealed that the risk model was primarily enriched in biological pathways such as EMT and the TGFβ signaling pathway (Fig. 9C). Furthermore, risk-associated genes, including MID2, KCNE4, SCG2, CILP, CRYAB, GJA1, and RGS16, as well as the risk scores, were predominantly enriched in fibroblast populations (Figs. 9D–E). Notably, compared to the low-risk group, the high-risk group demonstrated significantly enhanced cellular communication strength between fibroblasts, tumor epithelial cells, and myeloid cells (Fig. 9F). In conclusion, the risk model developed in this study not only highlights ECM remodeling and immune suppression as its primary biological characteristics but also underscores the heightened intercellular communication between fibroblasts, tumor epithelial cells, and myeloid cells, potentially driving CRC progression.

Discussion

This study systematically elucidates the relationship between EMT scores, CRC prognosis, and immunotherapy responsiveness by integrating six bulk transcriptomic datasets and four single-cell transcriptomic datasets. The core findings highlight the following: EMT scores are associated with poor prognosis and immune suppression

in CRC; TAGLN⁺Fib, a critical EMT subtype within CAFs, exhibit pro-metastatic and immunosuppressive functions; and TAGLN knockdown significantly inhibits EMT progression in CRC cells and reduces tumor burden. Furthermore, the risk model constructed based on EMT-TME scores derived from TAGLN⁺Fib demonstrated robust prognostic and immunotherapy response stratification capabilities across multi-center datasets. These findings provide valuable insights into CRC progression mechanisms and individualized therapeutic optimization.

An EMT score was developed based on an EMT gene set, revealing significant associations with OS, RFS, advanced tumor features, and immune suppression pathways in CRC patients. Notably, patients with high EMT scores exhibited greater resistance to immunotherapy, indicating that the EMT process plays a pivotal role in tumor progression and immune evasion, potentially by remodeling the TME [5, 7]. Tumor cells employ diverse immunosuppressive strategies to regulate immune responses and evade immune surveillance, such as upregulating immune checkpoint molecules, secreting immunosuppressive chemokines and cytokines, and recruiting immunosuppressive cells within the TME [35]. Research has shown that cancer cells leverage EMT as a survival strategy, enabling them to evade immune surveillance and resist therapy-induced cell death [36]. EMT enhances tumor cell migratory and invasive capacities, driving metastasis while reshaping the TME to induce an immunosuppressive state [37]. EMT has also been implicated in therapeutic resistance across various cancers, including resistance to anti-PD-L1 therapies (e.g., nivolumab and pembrolizumab) and anti-CTLA-4 therapies (e.g., ipilimumab) [38]. This resistance may be driven by mechanisms such as the activation of pro-survival signaling pathways, enhanced DNA repair mechanisms, and the acquisition of stem-like traits. For instance, upregulation of the β-catenin (CTNNB1) signaling pathway promotes immune evasion in hepatocellular carcinoma, contributing to resistance against anti-PD-L1 therapies like nivolumab [38]. The EMT process suppresses antitumor immune responses while enhancing immunosuppressive factors, such as TGF-β, IL-8, IL-10, and VEGF, which inhibit T-cell function and attract regulatory T cells (Tregs), myeloid-derived suppressor cells (MDSCs), and M2 macrophages. This creates an immunosuppressive microenvironment that shields cancer cells from recognition and attack by natural killer (NK) cells and cytotoxic CD8⁺ T cells. Furthermore, EMT facilitates the generation of chemoresistant cancer stem cells (CSCs) with immune evasion properties, further exacerbating therapeutic resistance [39].

Understanding the mechanisms of EMT-related immune evasion and therapeutic resistance is critical for developing effective strategies to address the challenges of cancer treatment. However, existing research has predominantly focused on tumor cells, with limited exploration of the roles of other cell types within the TME. In this study, we utilized single-cell and spatial transcriptomic data to comprehensively analyze EMT-associated cell subpopulations, identifying TAGLN⁺ fibroblasts (TAGLN⁺Fib) as a key CAF subtype involved in tumor progression. CAFs are not considered a homogenous population anymore, with newly research focusing on this. TAGLN (transgelin) is a protein closely associated with cytoskeletal remodeling [40]. While previous studies have linked TAGLN to CRC metastasis and poor prognosis [41, 42], its precise mechanism of action has remained unclear. However, TAGLN expression is not isolated; rather, it is closely associated with other widely recognized CAF markers and key drivers, such as ACTA2 (α-SMA), FAP, MMP2, PDPN, and THY1 [31–33]. Spatial transcriptomics data further clarified that these markers exhibit clear spatial colocalization within CRC tissues. ACTA2 (α-SMA) and FAP have been widely confirmed as classical markers of CAFs and are typically closely associated with high tumor invasiveness, immunosuppression, and poor prognosis [32, 33]. Multiplex immunofluorescence analysis confirmed the co-expression of TAGLN with ACTA2 (α-SMA) and FAP in CAFs. However, it is noteworthy that although TAGLN exhibits high network connectivity among multiple CAF-related genes, we do not contend that its function necessarily supersedes or replaces other classical CAF markers; rather, we emphasize its central regulatory role within a specific EMT-related CAF subpopulation. This understanding provides an important reference for subsequent targeted interventions aimed at modulating the function of specific CAF subpopulations. CRISPR-Cas9 knockdown experiments further revealed that TAGLN is a central regulator of EMT, significantly promoting tumor cell proliferation, migration, and EMT processes. In vivo experiments showed that TAGLN deletion markedly reduced tumor burden and suppressed EMT-associated phenotypes. Moreover, TAGLN⁺Fib was strongly associated with poor prognosis in CRC patients, underscoring the clinical significance of TAGLN as a potential therapeutic target. These findings have important implications for the clinical management of CRC. First, the functional role and potential mechanism of TAGLN⁺Fib in the EMT process of CRC were clarified. Elevated TAGLN expression not only serves as a predictor of poor clinical outcomes but also represents a potential biomarker for interventions targeting EMT-related pathological processes. In the future, targeted therapies against TAGLN⁺Fib, such as antibodies or small-molecule inhibitors, could provide innovative treatment options for patients with advanced CRC.

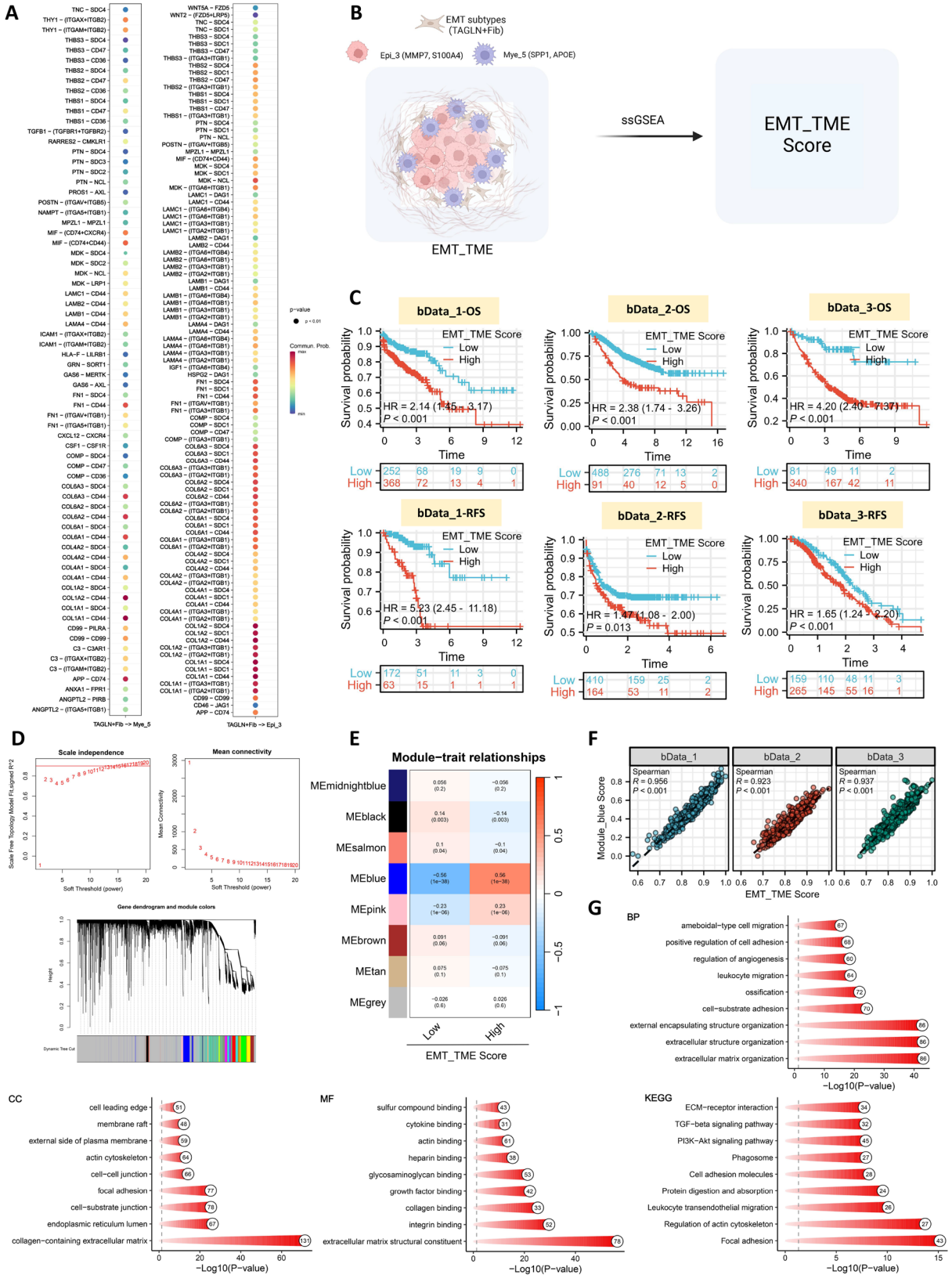

Fig. 7 Identification of Co-Expression Modules for EMT TME Scores. **A**. Heatmap of ligand-receptor interaction strength between TAGLN + Fib, Epi_3, and Mye_5. **B**. EMT_TME Score construction using the ssGSEA algorithm. **C**. Survival curve analysis of EMT_TME Score for OS and RFS. **D**. WGCNA analysis based on high and low EMT_TME Score groups. **E**. Heatmap of module correlations with EMT_TME Score groups. **F**. Correlation analysis of MEblue module scores with EMT_TME Score. **G**. GO and KEGG analysis of the MEblue module

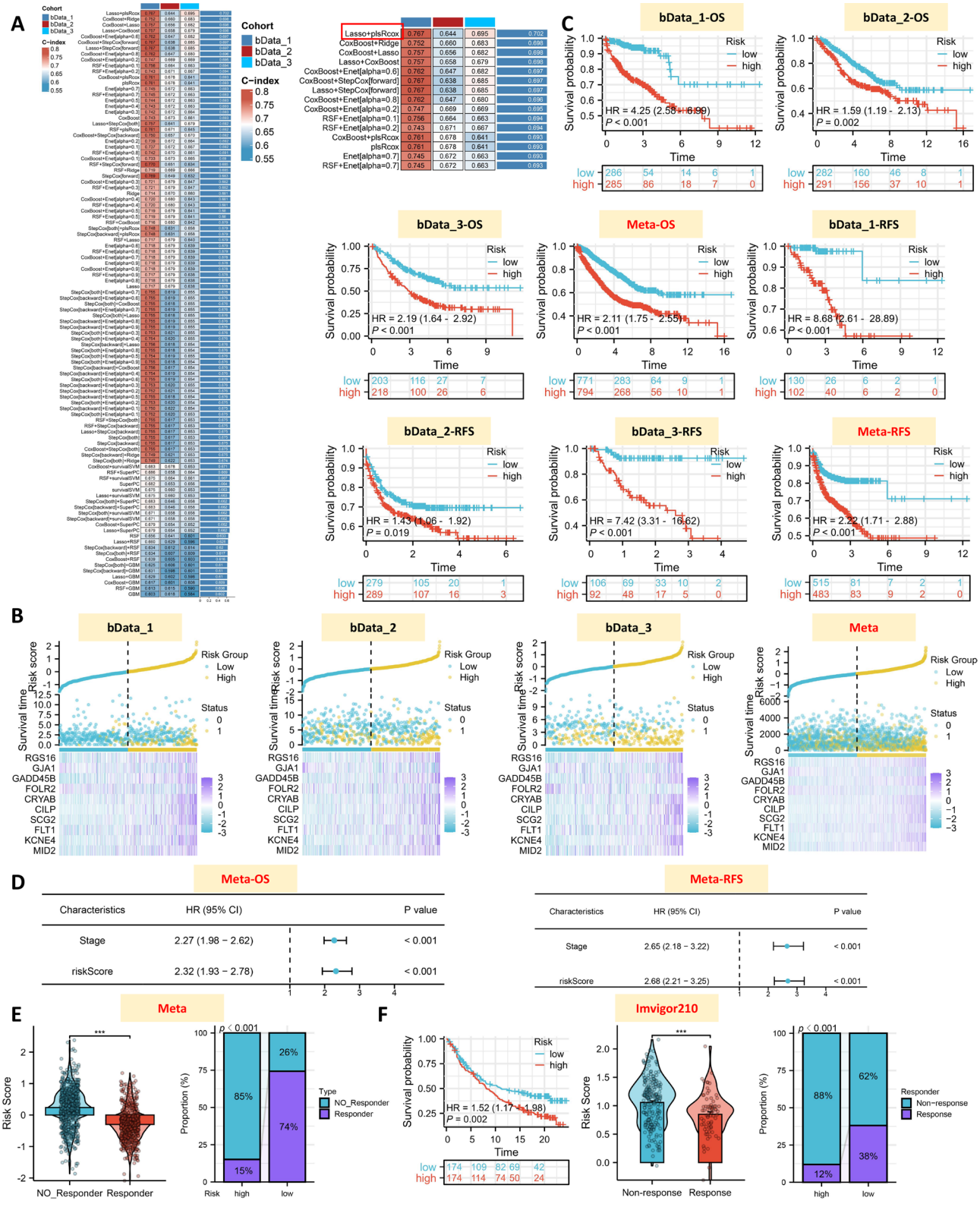

Fig. 8 Development of a CRC Prognostic Risk Model Based on the MEblue Module. **A** Performance of 101 machine learning algorithm combinations for CRC prognostic model construction. **B**. Correlation between CRC patient mortality, risk scores, and risk genes in the Meta cohort (bData_1, bData_2, bData_3). **C**. Survival curve analysis of risk scores for OS and RFS. **D**. Multivariate Cox regression analysis for OS and RFS. **E**-**F**. Differences in immunotherapy sensitivity and resistance between high-risk and low-risk groups

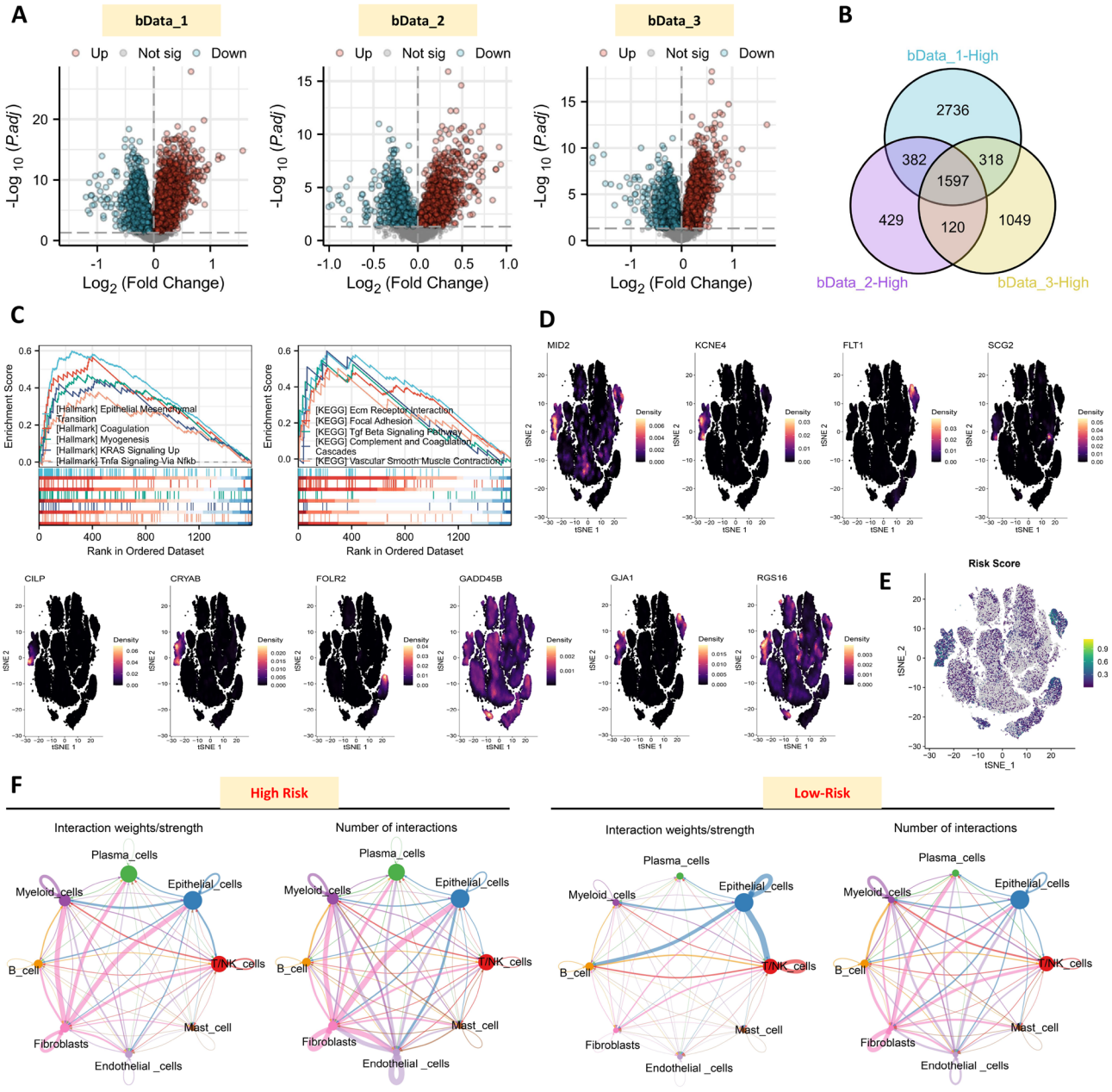

Fig. 9 Biological Characteristics of the Constructed Risk Model. (**A**) Volcano plot illustrating differentially expressed genes between high- and low-risk groups in training (bData_1) and validation (bData_2 and bData_3) cohorts. (**B**) Venn diagram showing shared differentially expressed genes across the three cohorts. (**C**) GSEA analysis of high- and low-risk groups based on HALLMARK and KEGG datasets. (**D**) Expression distribution of the 10 risk-associated genes. (**E**) Distribution of risk scores across cellular populations. (**F**) Cellular communication frequency and interaction weight among the eight major cell types in high- and low-risk groups

This study revealed that TAGLN⁺Fib exhibit significant immunosuppressive characteristics, with high infiltration of TAGLN⁺Fib correlating with greater resistance to immunotherapy. Notably, TAGLN⁺Fib was markedly enriched in the CMS4 subtype of CRC consensus molecular classification, which is associated with poor prognosis, increased metastasis, heightened immunosuppression, and suboptimal response to immunotherapy [29]. These findings further validate the immunosuppressive properties of TAGLN⁺Fib. How does TAGLN⁺Fib contribute to immunosuppression? Cellular communication analysis demonstrated significantly enhanced interactions between TAGLN⁺Fib and tumor epithelial cells as well as myeloid cells, particularly with Epi_3 (MMP7⁺Epi) and Mye_5 (SPP1⁺Macro). In this study, we found that TAGLN⁺Fib exhibited a close spatial colocalization with MMP7⁺ tumor epithelial cells (Epi_3) and SPP1⁺ macrophages (Mye_5), suggesting that

TAGLN⁺Fib may interact with the Epi_3 and Mye_5 subpopulations to promote the EMT process and remodeling of the tumor immune microenvironment in CRC. MMP7, a critical regulator of EMT and tumor metastasis, is highly expressed in poor-prognosis cases [43, 44]. SPP1⁺Macro promotes tumor progression through various mechanisms, including enhancing immunosuppression, driving tumor invasion and metastasis, and remodeling the TME [45]. Furthermore, SPP1⁺Macro has been shown to synergize with CAFs to establish immunosuppressive networks [46]. This study further identified that TAGLN⁺Fib exerts its effects through receptor-ligand axes, such as COL1A2/COL1A1-SCD4/SCD1/CD44/(ITGA3 + ITGB1) targeting MMP7⁺Epi, and COL1A2-CD44, COL1A1-CD44, and APP-CD74 targeting SPP1⁺Macro. These axes are well-documented drivers of tumor metastasis, immune evasion, and TME remodeling [47–50]. Therefore, we hypothesize that TAGLN⁺Fib may synergistically drive CRC progression and immune evasion through receptor–ligand interactions with MMP7⁺Epi and SPP1⁺Macro cells. Based on the involvement of TAGLN⁺Fib, MMP7⁺Epi, and SPP1⁺Macro, this study developed the EMT_TME Score, which emerged as a predictor of poor prognosis in CRC. The EMT_TME Score indicated that these three elements collectively drive CRC progression and holds potential as a tool for developing prognostic models. Using machine learning, we constructed a risk prognostic model based on the EMT_TME Score, which demonstrated robust predictive performance across multicenter validation cohorts. In recent years, numerous gene signature-based models have been proposed for prognostic prediction in CRC patients, yet systematic comparisons of their performance differences remain lacking [51, 52]. In this study, we rigorously compared the performance of our proposed EMT-TME score and its associated risk model with 126 published CRC prognostic gene models. The results demonstrated that the EMT-TME score performed well and ranked among the top models across three independent CRC cohorts as well as an integrated Meta cohort. Notably, the risk model based on the EMT-TME score consistently ranked first in multiple cohorts, exhibiting excellent predictive performance. These findings indicate that models built on EMT-related tumor microenvironment characteristics possess strong prognostic predictive capabilities. The model effectively stratified CRC patients by prognostic risk and predicted their sensitivity to immunotherapy, such as anti-PD-L1 therapy. This work provides an invaluable resource for clinicians in designing personalized treatment strategies. The EMT_TME Score has significant potential for broader clinical application, offering a practical tool to enhance CRC management and therapeutic decision-making in the future.

Although this study revealed the critical role of TAGLN⁺Fib in CRC progression and immunosuppression and developed an EMT-TME scoring risk model with clinical relevance, several limitations remain. First, although multiple CRC bulk transcriptomic and single-cell datasets were integrated to ensure the broad applicability of the results, the data were primarily derived from public databases and may exhibit sample heterogeneity and batch effects. Despite employing methods to remove batch effects, these factors might still influence the results. Second, while functional experiments both in vitro and in vivo validated the biological function of TAGLN, these experiments focused solely on a single aspect of TAGLN's function. Further systematic studies are needed to investigate the dynamic changes of TAGLN⁺Fib under different tumor environments or treatment conditions, as well as its interactions with other cells in the tumor microenvironment. In particular, regarding the regulatory effects of TAGLN⁺Fib on the functions of MMP7⁺Epi or SPP1⁺Macro cells, this study only proposed these hypotheses based on spatial transcriptomics and single-cell communication analyses, lacking definitive histological evidence and functional experimental validation. This represents an important limitation of the study; future research should include in vitro co-culture experiments (e.g., using blocking agents or specific gene knockout techniques) and more refined histological analyses to further elucidate the functional communication mechanisms between TAGLN⁺Fib and MMP7⁺Epi/SPP1⁺Macro cells, thereby providing a more solid theoretical basis for clinical interventions targeting CAF-mediated CRC progression. Finally, although the EMT-TME scoring risk model constructed in this study demonstrated good prognostic performance in multiple multicenter cohorts, its stability and generalizability still need to be validated in larger-scale prospective clinical studies.

Conclusion

In summary, this study establishes a strong association between EMT scores and CRC prognosis and immunotherapy response, identifies the crucial role of the TAGLN⁺Fib subtype in tumor progression and immunosuppression, and develops a EMT-TME Score-based risk model. These findings provide significant insights and a valuable reference for personalized treatment and precision medicine in CRC management.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13578-025-01405-x.

Supplementary Material 1: Supplementary Table 1. Primer sequences. Supplementary Table 2. Differential genes between TAGLN high expression group (TAGLN_High) and TAGLN low expression group (TAGLN_Low).

Supplementary Table 3. GSEA of upregulated genes in the TAGLN high expression group (TAGLN_High). Supplementary Table 4. Risk genes and corresponding risk coefficients. Supplementary Table 5. 195 published risk models.

Supplementary Material 2

Supplementary Material 3: Supplementary Fig. 1. Proportional distribution of the eight major cell populations. Supplementary Fig. 2. hdWGCNA and TAGLN Analysis. Supplementary Fig. 3. Correlation Analysis and Spatial Localization of CAF Marker Genes (ACTA2 (α-SMA), FAP, MMP2, PDPN, THY). Supplementary Fig. 4. Validation of TAGLN in CPTAC and GSEA. Supplementary Fig. 5. Correlation between TAGLN + Fib and immunosuppressive pathways. Supplementary Fig. 6. Cellular Communication and Epithelial Subcluster Analysis. Supplementary Fig. 7. Myeloid Subcluster Analysis. Supplementary Fig. 8. TAGLN + Fib Correlation Analysis. Supplementary Fig. 9. Survival and Correlation Analyses. Supplementary Fig. 10. Comparison of the EMT-TME Score and EMT-TME Risk Model with 126 Published CRC Risk Signatures.

Acknowledgements
Thank all authors for their contributions to this study.

Author contributions
Junli Zhang: Conceptualization, Methodology, Data Curation, Formal Analysis, Visualization, Writing - Original Draft Preparation, Investigation. Xinxin Jin: Data Curation, Software, Formal Analysis, Writing - Review & Editing. Yachao Hou: Investigation, Validation, Methodology, Writing - Review & Editing. Biao Gu: Resources, Supervision, Project Administration, Writing - Review & Editing. Hongwei Li: Data Curation, Validation, Resources. Li Yi: Software, Visualization, Data Curation. Wenjuan Wu: Conceptualization, Supervision, Project Administration, Funding Acquisition, Writing - Review & Editing. Shangshang Hu: Conceptualization, Supervision, Project Administration, Funding Acquisition, Writing - Review & Editing. All authors have reviewed and approved the final version of the manuscript.

Funding
This work is supported by Open Project of Anhui Provincial Key Laboratory of Tumor Evolution and Intelligent Diagnosis and Treatment (KFKT202411),. the Key Natural Science Project of Bengbu Medical University (2023byzd127), the Key Project of Natural Science Research Project of Colleges and Universities in Anhui Province(2022AH051485) and.the Key Project of Natural Science Research Project of Colleges and Universities in Anhui Province(2022AH051452).

Data availability
The datasets presented in this study can be found in online repositories. These can be found in the GEO database (https://www.ncbi.nlm.nih.gov/geo) and The Cancer Genome Atlas (TCGA) (https://portal.gdc.cancer.gov). The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author.

Declarations

Ethical approval
The studies involving human participants were reviewed and approved by the Ethics Committees and Institutional Review Boards of Nanjing First Hospital, affiliated with Nanjing Medical University. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Consent for publication
All authors have provided their consent for publication.

Conflict of interest
All authors declare that no conflict of interest exists.

Received: 29 January 2025 / Accepted: 5 May 2025
Published online: 24 May 2025

References

1. Siegel RL, Wagle NS, Cercek A, Smith RA, Jemal A. Colorectal cancer statistics, 2023. Cancer J Clin. 2023;73(3):233–54.
2. Cañellas-Socias A, Sancho E, Batlle E. Mechanisms of metastatic colorectal cancer. Nat Reviews Gastroenterol Hepatol. 2024;21(9):609–25.
3. Imodoye SO, Adedokun KA. EMT-induced immune evasion: connecting the Dots from mechanisms to therapy. Clin Experimental Med. 2023;23(8):4265–87.
4. Khan AQ, Hasan A, Mir SS, Rashid K, Uddin S, Steinhoff M. Exploiting transcription factors to target EMT and cancer stem cells for tumor modulation and therapy. Sem Cancer Biol. 2024;100:1–16.
5. Li X, Zhao S, Bian X, Zhang L, Lu L, Pei S, Dong L, Shi W, Huang L, Zhang X, Chen M, Chen X, Yin M. Signatures of EMT, immunosuppression, and inflammation in primary and recurrent human cutaneous squamous cell carcinoma at single-cell resolution. Theranostics. 2022;12(17):7532–49.
6. Huang Y, Hong W, Wei X. The molecular mechanisms and therapeutic strategies of EMT in tumor progression and metastasis. J Hematol Oncol. 2022;15(1):129.
7. Taki M, Abiko K, Ukita M, Murakami R, Yamanoi K, Yamaguchi K, Hamanishi J, Baba T, Matsumura N, Mandai M. Tumor immune microenvironment during Epithelial-Mesenchymal transition. Clin cancer Research: Official J Am Association Cancer Res. 2021;27(17):4669–79.
8. Van den Eynde A, Gehrcken L, Verhezen T, Lau HW, Hermans C, Lambrechts H, Flieswasser T, Quatannens D, Roex G, Zwaenepoel K, Marcq E, Joye P, De La Hoz C, Deben E, Gasparini C, Montay-Gruel A, Le Compte P, Lion M, Lardon E, Van Laere F, Siozopoulou S, Campillo-Davo V, De Waele D, Pauwels J, Jacobs P, Smits J, Van Audenaerde E. IL-15-secreting CAR natural killer cells directed toward the pan-cancer target CD70 eliminate both cancer cells and cancer-associated fibroblasts. J Hematol Oncol. 2024;17(1):8.
9. Wang H, Liang Y, Liu Z, Zhang R, Chao J, Wang M, Liu M, Qiao L, Xuan Z, Zhao H, Lu L. POSTN(+) cancer-associated fibroblasts determine the efficacy of immunotherapy in hepatocellular carcinoma. J Immunother Cancer. 2024;12(7).
10. Yang S, Zhang D, Sun Q, Nie H, Zhang Y, Wang X, Huang Y, Sun Y. Single-Cell and Spatial transcriptome profiling identifies the transcription factor BHLHE40 as a driver of EMT in metastatic colorectal Cancer. Cancer Res. 2024;84(13):2202–17.
11. Liu X, Qin J, Nie J, Gao R, Hu S, Sun H, Wang S, Pan Y. ANGPTL2 + cancer-associated fibroblasts and SPP1 + macrophages are metastasis accelerators of colorectal cancer. Front Immunol. 2023;14:1185208.
12. Qin J, Hu S, Lou J, Xu M, Gao R, Xiao Q, Chen Y, Ding M, Pan Y, Wang S. Selumetinib overcomes ITGA2-induced 5-fluorouracil resistance in colorectal cancer. Int Immunopharmacol. 2024;137:112487.
13. Marisa L, de Reyniès A, Duval A, Selves J, Gaub MP, Vescovo L, Etienne-Grimaldi MC, Schiappa R, Guenot D, Ayadi M, Kirzin S, Chazal M, Fléjou JF, Benchimol D, Berger A, Lagarde A, Pencreach E, Piard F, Elias D, Parc Y, Olschwang S, Milano G, Laurent-Puig P, Boige V. Gene expression classification of colon cancer into molecular subtypes: characterization, validation, and prognostic value. PLoS Med. 2013;10(5):e1001453.
14. Cherradi S, Martineau P, Gongora C, Del Rio M. Claudin gene expression profiles and clinical value in colorectal tumors classified according to their molecular subtype. Cancer Manage Res. 2019;11:1337–48.
15. Chen MS, Lo YH, Chen X, Williams CS, Donnelly JM, Criss ZK 2nd, Patel S, Butkus JM, Dubrulle J, Finegold MJ, Shroyer NF. Growth Factor-Independent 1 is a tumor suppressor gene in colorectal Cancer. Mol cancer Research: MCR. 2019;17(3):697–708.
16. Chen DT, Hernandez JM, Shibata D, McCarthy SM, Humphries LA, Clark W, Elahi A, Gruidl M, Coppola D, Yeatman T. Complementary strand MicroRNAs mediate acquisition of metastatic potential in colonic adenocarcinoma. J Gastrointest Surgery: Official J Soc Surg Aliment Tract. 2012;16(5):905–12. discussion 12– 3.
17. Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K, Baglaenko Y, Brenner M, Loh PR, Raychaudhuri S. Fast, sensitive and accurate integration of single-cell data with harmony. Nat Methods. 2019;16(12):1289–96.
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. Wu Y, Yang S, Ma J, Chen Z, Song G, Rao D, Cheng Y, Huang S, Liu Y, Jiang S, Liu J, Huang X, Wang X, Qiu S, Xu J, Xi R, Bai F, Zhou J, Fan J, Zhang X, Gao Q. Spatiotemporal immune landscape of colorectal Cancer liver metastasis at Single-Cell level. Cancer Discov. 2022;12(1):134–53.

20. Jin S, Guerrero-Juarez CF, Zhang L, Chang I, Ramos R, Kuan CH, Myung P, Plikus MV, Nie Q. Inference and analysis of cell-cell communication using cellchat. Nat Commun. 2021;12(1):1088.
21. Sun Y, Wu J, Zhang Q, Wang P, Zhang J, Yuan Y. Single-cell HdWGCNA reveals metastatic protective macrophages and development of deep learning model in uveal melanoma. J Translational Med. 2024;22(1):695.
22. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9:559.
23. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci USA. 2005;102(43):15545–50.
24. Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson Å, Kampf C, Sjöstedt E, Asplund A, Olsson I, Edlund K, Lundberg E, Navani S, Szigyarto CA, Odeberg J, Djureinovic D, Takanen JO, Hober S, Alm T, Edqvist PH, Berling H, Tegel H, Mulder J, Rockberg J, Nilsson P, Schwenk JM, Hamsten M, von Feilitzen K, Forsberg M, Persson L, Johansson F, Zwahlen M, von Heijne G, Nielsen J, Pontén F. Proteomics. Tissue-based map of the human proteome. Volume 347. New York, NY): Science; 2015. p. 1260419. 6220.
25. Hu S, Qin J, Ding M, Gao R, Xiao Q, Lou J, Chen Y, Wang S, Pan Y. Bulk integrated single-cell-spatial transcriptomics reveals the impact of preoperative chemotherapy on cancer-associated fibroblasts and tumor cells in colorectal cancer, and construction of related predictive models using machine learning. Biochim Et Biophys Acta Mol Basis Disease. 2025;1871(1):167535.
26. Chen B, Khodadoust MS, Liu CL, Newman AM, Alizadeh AA. Profiling tumor infiltrating immune cells with CIBERSORT. Methods in molecular biology. (Clifton NJ). 2018;1711:243–59.
27. Jiang P, Gu S, Pan D, Fu J, Sahu A, Hu X, Li Z, Traugh N, Bu X, Li B, Liu J, Freeman GJ, Brown MA, Wucherpfennig KW, Liu XS. Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. Nat Med. 2018;24(10):1550–8.
28. Necchi A, Joseph RW, Loriot Y, Hoffman-Censits J, Perez-Gracia JL, Petrylak DP, Derleth CL, Tayama D, Zhu Q, Ding B, Kaiser C, Rosenberg JE. Atezolizumab in platinum-treated locally advanced or metastatic urothelial carcinoma: post-progression outcomes from the phase II IMvigor210 study. Annals Oncology: Official J Eur Soc Med Oncol. 2017;28(12):3044–50.
29. Guinney J, Dienstmann R, Wang X, de Reyniès A, Schlicker A, Soneson C, Marisa L, Roepman P, Nyamundanda G, Angelino P, Bot BM, Morris JS, Simon IM, Gerster S, Fessler E, De Sousa EMF, Missiaglia E, Ramay H, Barras D, Homicsko K, Maru D, Manyam GC, Broom B, Boige V, Perez-Villamil B, Laderas T, Salazar R, Gray JW, Hanahan D, Tabernero J, Bernards R, Friend SH, Laurent-Puig P, Medema JP, Sadanandam A, Wessels L, Delorenzi M, Kopetz S, Vermeulen L, Tejpar S. The consensus molecular subtypes of colorectal cancer. Nat Med. 2015;21(11):1350–6.
30. Liu Z, Liu L, Weng S, Guo C, Dang Q, Xu H, Wang L, Lu T, Zhang Y, Sun Z, Han X. Machine learning-based integration develops an immune-derived LncRNA signature for improving outcomes in colorectal cancer. Nat Commun. 2022;13(1):816.
31. Shah R, Johnson KA, Lippert AEL, Kraus SG, Emmerich PB, Pasch CA, Zhang W, Matkowskyj KA, LeBeau AM, Deming DA. Cancer-Associated fibroblast proteins as potential targets against colorectal cancers. Cancers. 2024;16(18).
32. Zhao Y, Jia Y, Wang J, Chen X, Han J, Zhen S, Yin S, Lv W, Yu F, Wang J, Xu F, Zhao X, Liu L. circNOX4 activates an inflammatory fibroblast niche to promote tumor growth and metastasis in NSCLC via FAP/IL-6 axis. Mol Cancer. 2024;23(1):47.
33. Varveri A, Papadopoulou M, Papadovasilakis Z, Compeer EB, Legaki AI, Delis A, Damaskou V, Boon L, Papadogiorgaki S, Samiotaki M, Foukas PG, Eliopoulos AG, Hatzioannou A, Alissafi T, Dustin ML, Verginis P. Immunological synapse formation between T regulatory cells and cancer-associated fibroblasts promotes tumour development. Nat Commun. 2024;15(1):4988.
34. Cai J, Song L, Zhang F, Wu S, Zhu G, Zhang P, Chen S, Du J, Wang B, Cai Y, Yang Y, Wan J, Zhou J, Fan J, Dai Z. Targeting SRSF10 might inhibit M2 macrophage polarization and potentiate anti-PD-1 therapy in hepatocellular carcinoma. Cancer Commun (London England). 2024;44(11):1231–60.
35. Singh D, Khan MA, Siddique HR. Role of p53-miRNAs circuitry in immune surveillance and cancer development: A potential avenue for therapeutic intervention. Semin Cell Dev Biol. 2022;124:15–25.
36. De Angelis ML, Francescangeli F, Zeuner A. Breast Cancer stem cells as drivers of tumor chemoresistance, dormancy and relapse: new challenges and therapeutic opportunities. Cancers. 2019;11(10).
37. Singh D, Siddique HR. Epithelial-to-mesenchymal transition in cancer progression: unraveling the immunosuppressive module driving therapy resistance. Cancer Metastasis Rev. 2024;43(1):155–73.
38. Ruiz de Galarreta M, Bresnahan E, Molina-Sánchez P, Lindblad KE, Maier B, Sia D, Puigvehi M, Miguela V, Casanova-Acebes M, Dhainaut M, Villacorta-Martin C, Singhi AD, Moghe A, von Felden J, Tal Grinspan L, Wang S, Kamphorst AO, Monga SP, Brown BD, Villanueva A, Llovet JM, Merad M, Lujambio A. β-Catenin activation promotes immune escape and resistance to Anti-PD-1 therapy in hepatocellular carcinoma. Cancer Discov. 2019;9(8):1124–41.
39. Lin H, Fu L, Zhou X, Yu A, Chen Y, Liao W, Shu G, Zhang L, Tan L, Liang H, Wang Z, Deng Q, Wang J, Jin M, Chen Z, Wei J, Cao J, Chen W, Li X, Li P, Lu J, Luo J. LRP1 induces anti-PD-1 resistance by modulating the DLL4-NOTCH2-CCL2 axis and redirecting M2-like macrophage polarisation in bladder cancer. Cancer Lett. 2024;593:216807.
40. Li H, Song C, Zhang Y, Liu G, Mi H, Li Y, Chen Z, Ma X, Zhang P, Cheng L, Peng P, Zhu H, Chen Z, Dong M, Chen S, Meng H, Xiao Q, Li H, Wu Q, Wang B, Zhang S, Shu K, Wan F, Guo D, Zhou W, Zhou L, Mao F, Rich JN, Yu X. Transgelin Promotes Glioblastoma Stem Cell Hypoxic Responses and Maintenance Through p53 Acetylation. Advanced science (Weinheim, Baden-Wurttemberg, Germany). 2024;11(7):e2305620.
41. Zhou H, Li L, Xie W, Wu L, Lin Y, He X. TAGLN and High-mobility group AT-Hook 2 (HMGA2) complex regulates TGF-β-induced colorectal Cancer metastasis. OncoTargets Therapy. 2020;13:10489–98.
42. Zhu Z, Li J, Fa Z, Xu X, Wang Y, Zhou J, Xu Y. Functional gene signature offers a powerful tool for characterizing clinicopathological features and depicting tumor immune microenvironment of colorectal cancer. BMC Cancer. 2024;24(1):1199.
43. Park SS, Lee YK, Choi YW, Lim SB, Park SH, Kim HK, Shin JS, Kim YH, Lee DH, Kim JH, Park TJ. Cellular senescence is associated with the Spatial evolution toward a higher metastatic phenotype in colorectal cancer. Cell Rep. 2024;43(3):113912.
44. Blaheta RA, Han J, Oppermann E, Bechstein WO, Burkhard K, Haferkamp A, Rieger MA, Malkomes P. Transglutaminase 2 promotes epithelial-to-mesenchymal transition by regulating the expression of matrix metalloproteinase 7 in colorectal cancer cells via the MEK/ERK signaling pathway. Biochim Et Biophys Acta Mol Basis Disease. 2025;1871(1):167538.
45. Bill R, Wirapati P, Messemaker M, Roh W, Zitti B, Duval F, Kiss M, Park JC, Saal TM, Hoelzl J, Tarussio D, Benedetti F, Tissot S, Kandalaft L, Varrone M, Ciriello G, McKee TA, Monnier Y, Mermod M, Blaum EM, Gushterova I, Gonye ALK, Hacohen N, Getz G, Mempel TR, Klein AM, Weissleder R, Faquin WC, Sadow PM, Lin D, Pai SI, Sade-Feldman M, Pittet MJ. CXCL9:SPP1 macrophage Polarity identifies a network of cellular programs that control human cancers. Volume 381. New York, NY): Science; 2023. pp. 515–24. 6657.
46. Cheng MK, McKean J, Boisvert D, Tulip J, Mielke BW. Effects of photoradiation therapy on normal rat brain. Neurosurgery. 1984;15(6):804–10.
47. Bai J, Liu T, Tu B, Yuan M, Shu Z, Fan M, Huo S, Guo Y, Wang L, Wang H, Zhao Y. Autophagy loss impedes cancer-associated fibroblast activation via downregulating proline biosynthesis. Autophagy. 2023;19(2):632–43.
48. Chen Y, Yang S, Tavormina J, Tampe D, Zeisberg M, Wang H, Mahadevan KK, Wu CJ, Sugimoto H, Chang CC, Jenq RR, McAndrews KM, Kalluri R. Oncogenic collagen I homotrimers from cancer cells bind to α3β1 integrin and impact tumor Microbiome and immunity to promote pancreatic cancer. Cancer Cell. 2022;40(8):818–e349.
49. Huang Q, Liu L, Xiao D, Huang Z, Wang W, Zhai K, Fang X, Kim J, Liu J, Liang W, He J, Bao S. CD44(+) lung cancer stem cell-derived pericyte-like cells cause brain metastases through GPR124-enhanced trans-endothelial migration. Cancer Cell. 2023;41(9):1621–e368.
50. Zhu GQ, Tang Z, Huang R, Qu WF, Fang Y, Yang R, Tao CY, Gao J, Wu XL, Sun HX, Zhou YF, Song SS, Ding ZB, Dai Z, Zhou J, Ye D, Wu DJ, Liu WR, Fan J, Shi YH. CD36(+) cancer-associated fibroblasts provide immunosuppressive microenvironment for hepatocellular carcinoma via secretion of macrophage migration inhibitory factor. Cell Discovery. 2023;9(1):25.
51. Su G, Wang M, Qian J, Wang Y, Zhu Y, Wang N, Wang K, Wang Q, Wang Y, Li D, Yang L. Comprehensive analysis of a Platelet- and Coagulation-Related prognostic gene signature identifies CYP19A1 as a key tumorigenic driver of colorectal Cancer. Biomedicines. 2024;12(10).

52. Lei J, Fu J, Wang T, Guo Y, Gong M, Xia T, Shang S, Xu Y, Cheng L, Lin B. Molecular subtype identification and prognosis stratification by a Immunogenic cell death-related gene expression signature in colorectal cancer. Expert Rev Anticancer Ther. 2024;24(7):635–47.

Publisher's note