Supplemental materials for

# Novel TCF21$^{high}$ pericyte subpopulation promotes colorectal cancer metastasis by remodelling perivascular matrix

## Supplemental methods

### Cell lines and cell culture

Human CRC cell lines HCT116, DLD-1, RKO, SW480, SW620, Caco-2, HT29 and human microvascular endothelial cell-1 (HMEC-1) were purchased from the American Type Culture Collection (Manassas, VA). Mouse colon cancer cell line MC38 (Cat. BNCC337716) was from BeNa Culture Collection (Beijing, China). HCT116, DLD-1, RKO, SW480, SW620 and Caco-2 cells were cultured with DMEM. MC38 cells were maintained in RPMI-1640. DMEM and RPMI-1640 medium were supplemented with 10% FBS (Cat. FSP500, ExCell Bio, Shanghai, China) and 1% penicillin-streptomycin (PS). HMEC-1 cells were cultured in endothelial cell medium (ECM, Cat. 1001, Sciencell research laboratories, Corte Del Cedro Carlsbad, CA) supplemented with 5% FBS, 1% endothelial cell growth supplement (ECGS), and 1% PS. All cell lines were cultured at 37 °C in incubator with 5% $CO_2$. MC38, HCT116 and DLD-1 cells were infected with lentivirus harboring luciferase (Genechem, Shanghai, China) to generate the MC38-luc, HCT116-luc and DLD-1-luc cells, which were then selected with puromycin (2 μg/mL) for 2 days. All cell lines were authenticated to have no cross-contamination using a STR Multi-amplification Kit (Microreader$^{TM}$21 ID System) and tested negative for mycoplasma by the TransDetect® PCR Mycoplasma Detection Kit (Cat. FM311-01 Transgen, Beijing, China).

### Human samples and specimens



Li X, *et al*. *Gut* 2023; 72:710–721. doi: 10.1136/gutjnl-2022-327913

Human CRC surgical samples (12 cases, patients' information was listed in **Supplemental Table 5 and Supplemental Table 6**) and specimens (75 cases, patients' information was listed in **Supplemental Table 1**) were obtained from the First Affiliated Hospital of Jinan University (Guangzhou, China).

## Mice

Male C57BL/6JGpt mice, male BALB/c nude mice (4-6 weeks, 20-22 g), Rosa26-CAG-LSL-Cas9-tdTomato mice (B6/JGpt-*Rosa26*$^{tm1(CAG-LSL-Cas9-tdTomato)}$/Gpt; Cat. T002249), *Cspg4*-CreERT2 mice (B6/JGpt-*Cspg4*$^{em1Cin(CreERT2-P2A)/Gpt}$; T006187), and *Tcf21*-flox mice (B6/JGpt-*Tcf21*$^{em1Cflox}$/Gpt; T013083) were obtained from GemPharmatech Co., Ltd (Nanjing, Jiangsu, China). Pericyte lineage tracing mice ($PC^{lin}$) were generated by crossing mice carrying a tamoxifen-inducible Cre recombinase driven by the pericyte-specific *Cspg4* promoter ($Tg^{Cspg4\text{-}CreERT2}$) with mice carrying a Cre-responsive reporter gene (tandem dimer Tomato (*tdT*)) inserted at the *ROSA26* locus ($ROSA^{tdT/+}$). The $PC^{lin}$ mice were further crossed with mice harboring both *Tcf21* alleles flanked by LoxP sites ($Tcf21^{flox/flox}$) to generate tamoxifen-inducible *Cspg4*-driven pericyte-specific *Tcf21* knockout mice ($PC^{lin\text{-}KO}$). All mice were maintained in a specific pathogen-free (SPF) facility. Mouse genotyping was detected by PCR (The primer sequences were listed in **Supplemental Table 7**). Cre activity was induced in tumor-bearing mice (6-7 weeks, weight 22-25 g) via oral gavage every other day for 3 times (10 mg/kg of tamoxifen in peanut oil). The animal experiments were complied with the ARRIVE Guidelines 2.0: updated guidelines for reporting animal research[1].

## Isolation and culture of TPCs

TPCs were isolated from CRC patients though a microdissection combined with pericyte medium-based approach that developed by our lab. Briefly, fresh surgical tumor specimens were obtained from CRC patients with or without liver metastasis. Information of the CRC patients was listed in **Supplemental Table 5 and Supplemental Table 6.** Tumor tissues were kept in serum-free DMEM containing PS and placed on the ice, and then washed with pre-chilled PBS in a sterile hood to remove the blood, adipose tissues. Tumor vessels were separated from perivascular adipose tissues under a stereomicroscope (Olympus, SZX7). The acquired tumor vessels were cultured in Pericyte Medium (PM, Cat. 1201, Corte Del Cedro Carlsbad, CA, USA, Sciencell research laboratories) with 5% FBS, 1% PGS and 1% PS at 37 °C with 5% $CO_2$. TPCs were migrated from the tumor vessels within 14 days, which were then disassociated by trypsin once the confluence reaches 80%. The purity of the isolated TPCs were authenticated by STR Multi-amplification Kit (**Supplemental Table 8**).

**Construction of single cell cDNA libraries**

For single cell cDNA libraries construction, passage 1 TPCs derived from four CRC patients (patient information was listed in **Supplemental Table 9**) were prepared and analyzed by a 10×Genomics GemCode Single-cell instrument, generating single-cell Gel Bead-In-EMlusion (GEMs). The libraries were generated and sequenced by Chromium Next GEM Single Cell 3' Reagent Kits v3.1 and Illumina HiSeq 4000 by Genedenovo Biotechnology Co., Ltd (Guangzhou, China) with a custom paired-end sequencing mode 26 bp (read 1) × 98 bp (read 2).

**Bioinformatic analysis of scRNA-seq**

Reads uniquely mapped to the transcriptome and intersecting an exon at least 50% were considered for UMI counting. Before quantification, the UMI sequences would be corrected for sequencing errors, and valid barcodes were identified based on the EmptyDrops method. The cells by gene matrices were produced via UMI counting and cell barcodes calling. The cells by gene matrices for each sample were individually imported to Seurat version 3.1.1 for downstream analysis. After removing the unwanted cells from the dataset, data normalization and batch effect correction, the integrated expression of matrix was then scaled and performed on principal component analysis (PCA) for dimensional reduction, those had a strong enrichment of low *P*-value genes for downstream clustering were identified as significant PCs. Seurat implemented a graph-based clustering approach. Distances between the cells were calculated based on previously identified PCs. For visualization of clusters, t-distributed Stochastic Neighbor Embedding (t-SNE) were generated using the same PCs. Expression value of each gene in given clusters were compared against the rest of cells using Wilcoxon rank sum test. Significantly upregulated genes were identified using several criteria. First, genes had to be at least 1.28-fold overexpressed in the target cluster. Second, genes had to be expressed in more than 25% of the cells belonging to the target cluster. Third, *P* value is less than 0.05.

The Gene ontology (GO) enrichment analysis was performed with the GO database (http://www.geneontology.org/). GO has three ontologies: molecular function, cellular component, and biological process. The calculated *P*-values were false discovery rate (FDR)-corrected, taking FDR $\leq$ 0.05 as a threshold. GO terms meeting this criterion were defined as significantly enriched GO terms in differentially expressed genes.

Analysis of transcription factor network inference was performed with the SCENIC R package[2]. In brief, log-normalized expression matrix generated using Seurat was used as input, and the pipeline was implanted in three steps. First, gene co-expression network was established via GENIE3[3]. Second, each module was pruned based on a regulatory motif near a transcription start site via RcisTarget. Precisely, the networks were retained if the transcription factor (TF)-binding motif was enriched among its targets, while target genes without direct TF-binding motifs were removed. The retained networks were called regulons. Third, the activity of each regulon in each single cell was scored (AUC score) using AUCell R package. Gene regulatory network (GRN) plots of all regulons were done using the cytoscape software[4].

**Analysis of the public datasets**

scRNA-seq data of colon (GEO accession GSM3140596, GSM3140595, GSM3140594, and GSM3140593)[5] and intestine (GEO accession GSM4159165 and GSM4159164)[6] acquired using 10× Chromium protocol were download, and the sequencing reads were realigned, and cell clustering was performed as described above.

**Construction of MC38-luc-LM3, HCT116-luc-LM3 cells and DLD-1-luc-LM3 cells**

To establish highly metastatic MC38-luc-LM3 cells, MC38-luc cells ($1\times10^5$) suspended in 100 μL of Matrigel (Cat. 354248, Corning, NY) were injected into the spleen of male C57BL/6JGpt mice anesthetized with isoflurane inhalation. Liver metastasis was detected by bioluminescence imaging. MC38-luc-LM1 cells from the metastatic foci were isolated by mouse tumor dissociation kit (Cat. 130-096-730, Miltenyi Biotec, Bergisch Gladbach, Germany) and cultured in complete RPMI-1640 medium and

selected by puromycin (2 μg/mL). MC38-luc-LM1 cells were inoculated into the spleen of male C57BL/6JGpt mice, and MC38-luc-LM2 cells were obtained from the liver metastatic foci. Tumor cells isolated from the third round of liver metastatic foci were termed MC38-luc-LM3 cells, which were employed for the subsequent experiments. The HCT116-luc-LM3 and DLD-1-luc-LM3 cells were isolated by human tumor dissociation kit (Cat. 130-095-929, Miltenyi Biotec) and acquired with BALB/C nude mice by the same pattern of MC38-luc-LM3. The cellular morphology, nucleus size, cell size, cell migration, proliferation, EpCAM expression, stemness and EMT were assessed to evaluate the phenotypical/biological differences between the parental cells and the LM3 cells (**Supplemental Figure 20**). The origin of all LM3 cells was further validated by short tandem repeat (STR) (**Supplemental table 10-12**) and luciferase activity (**Supplemental Figure 21**).

**Flow cytometry**

Cells were collected, re-suspended in flow cytometry staining buffer, and distributed into 1.5 mL EP tubes. Following fixation with 4% paraformaldehyde on ice and permeabilization with 0.1% Triton X-100 in PBS for 5 min, the cells were incubated with anti-TCF21 antibody (Cat. AB_182134, Abcam) or anti-MATN2 antibody (Cat. AF3044, R&D system,) for 1 h on ice. Then, cells were washed with PBS twice and incubated with Donkey anti-Rabbit IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor 546 (Cat. AB_2534016, Invitrogen, Carlsbad, CA, USA) or Donkey anti-Goat IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor 488 (Cat. AB_2534102, Invitrogen) for 1 h and analyzed by flow cytometry (BD Biosciences, San Jose, CA). The size of HCT116-luc, DLD-1-luc, MC38-luc, HCT116-luc-LM3, DLD-1-luc-LM3 and MC38-luc-LM3 cells was directly evaluated by forward

scatter using flow cytometry

**Animal studies**

MC38-luc-LM3 cells ($1\times10^5$) suspended in 100 μL of Matrigel were orthotopically injected into the cecum wall of $PC^{lin}$ mice and $PC^{lin\text{-}KO}$ mice anesthetized with isoflurane inhalation. At the end of the experiment, tumors were collected and subjected to immunohistochemistry and immunofluorescence analysis. The livers were harvested, photographed, and prepared for H&E staining. For co-injection assays, HCT116-luc-LM3, DLD-1-luc-LM3 cells, $TPC_{NM}$ transfected with lentivirus harboring negative control shNC ($TPC_{NM}^{shNC}$) or shITGA5 ($TPC_{NM}^{shITGA5}$), $TPC_{LM}$ transfected with Vector ($TPC_{LM}^{Vector}$) or lentivirus expressing ITGA5 ($TPC_{LM}^{ITGA5}$) were collected. HCT116-luc-LM3 or DLD-1-luc-LM3 cells ($4\times10^5$) were premixed with TPCs ($1.6\times10^6$) in 100 μL of Matrigel, which was then injected into the cecum wall of BALB/C nude mice. At the end of the experiment, the mice were sacrificed with $CO_2$ and the metastatic foci in mouse liver were analyzed by H&E staining. Orthotopic tumor tissues were obtained for Masson staining, immunohistochemical staining, immunofluorescence, and transmission electron microscope analysis.

***In vivo* cell tracking**

For the whole animal imaging *in vivo*, mice were intraperitoneal (i.p.) injected with 3 mg of D-luciferin (Cat. 40901ES01, Yeason Biotechnology, Shanghai, China) dissolved in 200 μL saline and were anesthetized by isoflurane after injection for 5 min. Luminescence signals were collected with Xenogen IVIS 200 (Alameda, CA, USA) and analyzed by the Xenogen Living Image software (Alameda, CA, USA).

**Vessel permeability assay**

$PC^{lin}$ and $PC^{lin\text{-}KO}$ mice bearing MC38 orthotopic xenografts were intravenously (i.v.) injected with 1 mg of FITC-labeled Dextran-40 kDa (Cat. D1845, Thermo Scientific) for 10 min. Then, the mice were perfused with 4% PFA and tumors were obtained and then frozen. Tumor tissues were sectioned, and tumor vessels were stained for CD31 (RRID: AB_2161028, RD, Minneapolis, MN, USA) followed by Donkey anti-Goat IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor™ 647 (RRID: AB_2534102, Invitrogen, Carlsbad, CA, USA), and the double staining of FITC-labeled Dextran over CD31-positive vessels indicates vessel permeability.

**Isolation and identification of circulating tumor cells (CTCs)**

CTC isolation was performed according to a previous study[7]. Briefly, blood (500 μL) was collected from each of $PC^{lin}$ and $PC^{lin\text{-}KO}$ mice bearing MC38-luc-LM3 allografts by cardiac puncture and immediately released into heparin-coated tube to avoid coagulation. The red blood cells were removed by blood red cell lysing reagent before cells were seeded in a 12-well plate and cultured with DMEM supplemented with 20% FBS. The adherent tumor cells were identified and counted within 12 h prior to no tumor cell growth. Adherent cells were stained with cancer cell-associated surface marker EpCAM and leukocyte marker CD45 and identified by confocal microscopy as described previously[8, 9]. Cells positive for EpCAM but not CD45 were scored as CTCs and subsequently subjected to manual counting, and the CTC counts were presented as CTCs per milliliter of whole blood.

**Isolation efficiency of CTCs**

To determine the isolation efficiency of CTCs[10-12], 500 MC38-luc-LM3 cells were

spiked into 500 μL of blood collected from the healthy C57BL/6JGpt mice by cardiac puncture. The spiked blood was then treated with blood red cell lysing reagent and the remaining cells were seeded on a 12-well plate and cultured with DMEM containing 20% FBS. The adherent tumor cells were stained with EpCAM and CD45 and the $EpCAM^{+}CD45^{-}$ CTCs were identified by confocal microscopy. The number of $EpCAM^{+}CD45^{-}$ CTCs was counted within 12 h at a time of no tumor cell growth. The efficiency of CTC recovery was calculated using the following equation: Cell recovery (%) = counts of isolated MC38-luc-LM3 cells/500 × 100%.

**Chromatin immunoprecipitation (ChIP) and ChIP-Seq**

ChIP assay was performed according to the manufacture manual of SimpleChIP® Enzymatic Chromatin IP Kit (Cat. 9003, Cell Signaling Technology, MA, USA). Briefly, TCF21-overexpressing TPCs ($TPC_{NM}^{TCF21}$) were washed twice in clod PBS buffer and cross-linked with 1% formaldehyde for 10 min at room temperature and then stopped by addition of glycine (125 mM). Afterwards, samples were lysed, and chromatins were obtained on ice. Chromatins were then sonicated to get soluble sheared chromatin (average DNA length of 150-900 bp). Then, 20 μL of chromatin was saved as input and 100 μL of chromatin was harvested for immunoprecipitation by anti-TCF21 antibody (RRID: AB_10601215, Sigma, Shanghai， China), and anti-IgG was served as the negative control. 10 μg of anti-TCF21 was used in the immunoprecipitation reactions at 4 °C overnight. Then 30 μL of protein A beads was added and the samples were further incubated for 2 h. After reverse cross-linking and DNA purification, immunoprecipitated DNA was quantified by real-time PCR. Immunoprecipitated DNA was used to construct sequencing libraries following the protocol provided by the NEXTflex® ChIP-Seq kit（Cat. NOVA-5143-02, BioScientific, TX , USA） and

sequenced on Illumina Xten with PE 150 method (LC-Bio Technology CO., Ltd., Hangzhou, China). For data analysis, Trimmomatic (version 0.38) was used to filter out low-quality read. MACS2 software (version 2.1.1.20160309) was used to call peaks by default parameters (bandwidth, 300 bp; model fold, 5, 50; q value, 0.05). If the summit of a peak located closest to the TSS of one gene, the peak will be assigned to that gene. GO enrichment analysis was performed using the EasyGO gene ontology enrichment analysis tool (http://bioinformatics.cau.edu.cn/easygo). The GO term enrichment was calculated using hypergeometric distribution with a *P* value cutoff of 0.01. *P* values obtained by Fisher's exact test were adjusted with FDR for multiple comparisons to detect overrepresented GO terms.

**RT-qPCR assay**

Total RNA was collected by E.Z.N.A.® Total RNA Kit I (Cat. R6834-02, Omega Bio-Tek, Norcross, GA, USA). The purity and concentration of RNA was examined by Nanodrop Lite micro spectrophotometer. RNA (2 μg) was reversely transcribed to cDNA with All-in-One cDNA Synthesis SuperMix (Cat. B24408-1000, Bimake, Houston, TX, USA). Reverse transcription quantitative PCR (RT-qPCR) was performed in triplicate using 2× SYBR Green qPCR Master Mix (Cat. B21202, Bimake). Samples were loaded into a Roche LightCycler 480 II real-time polymerase chain reaction detection system (Roche，Basel, Switzerland) and the data is analyzed by $2^{-\triangle\triangle Ct}$ method. The primer sequences were listed in **Supplemental Table 13.**

**Cell infection and transfection**

TPCs derived from CRC patients with non-metastasis ($TPC_{NM}$) were infected with lentivirus harboring *TCF21* overexpression plasmid for 48 h and selected by puromycin

(2 mg/mL). Detailed information of *TCF21* overexpression plasmid and vector was listed as follows: TCF21 (NM_198392) Human Untagged Clone (Cat. RC220002, Origene. Rockwell, MD, USA), Cloning vector pCMV6-Entry (Cat. PS100001, Origene). For TCF21- or MATN2-knockdown experiments, TPCs derived from CRC patients with liver metastasis ($TPC_{LM}$) were transfected with siRNA for 48 h followed by subsequent analysis. For ITGA2- and ITGB1-knockdown experiments, $TPC_{NM}$ were transfected with siRNA for 48 h followed by subsequent analysis. Transfection was performed with Lipofectamine™ 3000 (Cat. L3000015, Invitrogen, Carlsbad, CA, USA) according to the manufacturer instructions, and the siRNA sequences were listed in **Supplemental Table 14.** For MATN2-overexpression experiments, $TPC_{NM}$ were transfected with pCMV6-*MATN2*-overexpressing plasmid (Cat. RC203833, Origene) or pCMV6-Entry as empty vector for 48 h (Cat. PS100001, Origene). For ITGA5 overexpression experiments, $TPC_{LM}$ were infected with lentivirus harboring *ITGA5* or its corresponding Vector (pGC-FU-3FLAG-CBh-gcGFP-IRES-puromycin) (Genechem, Shanghai, China). For ITGA5-knockdown experiments, $TPC_{NM}$ were infected with lentivirus harboring shITGA5 or pFU-GW-016 as Vector (Genechem, Shanghai, China).

**Y15 and SGI1027 treatment**

Y15 and SGI1027 were purchased from Selleck (Shanghai, China) and dissolved in DMSO. Integrin α5-overexpressing TPCs ($2\times10^5$) were seeded into 6-well plates and cultured overnight. The next day, cells were treated with Y15 (5 μM) or SGI1027 (2.5 μM) for 24 h, and then cells were applied for Western blotting assay and bisulfite sequencing.

**Western blotting assay**

Cells were lysed in RIPA lysis buffer on ice for 30 min. Total protein concentration was measured with Pierce™ BCA Protein Assay Kit (Cat. 23225, Thermo Scientific, Waltham, MA, USA). Equal amounts of protein (20 μg) were separated in SDS-PAGE gel (Cat. G2004, Solarbio, Beijing, China) and transferred onto polyvinylidene fluoride (PVDF) membranes (Cat. IPVH00010, Millipore, Boston, MA, USA). Following blocking with 5% BSA, the membranes were incubated with indicated antibodies. The blots were detected by Amersham Imager 600 (GE, Boston, MA, USA). The antibodies were listed in **Supplemental Table 15**.

**Immunofluorescence analysis**

Tissue slices were deparaffinized and incubated with 1×Tris-EDTA (pH 9.0) and 0.05% Tween for 3 min for antigen retrieval. After that, tumor sections were permeabilized in 0.1% Triton™ X-100, blocked with QuickBlock™ immunostaining blocking solution (Cat. ST797, Beyotime, Shanghai, China) and incubated with the corresponding primary antibody overnight at 4 °C. Then, the sections were incubated with the corresponding secondary antibody for 1 h at room temperature. For nucleus staining, sections were incubated with 1 μg/mL DAPI (Cat. MBD0015, Sigma) for 15 min. The slides were photographed with a Zeiss LSM 800 confocal microscope and analyzed with Image J software (RRID: SCR_003070, Rawak Software Inc., Stuttgart, Germany). The primary and secondary antibodies were listed in **Supplemental Table 16.** For phalloidin immunofluorescence assay, HCT116-luc, DLD-1-luc, MC38-luc, HCT116-luc-LM3, DLD-1-luc-LM3 or MC38-luc-LM3 cells were plated on the glass bottom cell culture dish and incubated with DMEM for 24 h. The next day, cells were fixed, permeabilized with 0.1% Triton™ X-100 and then incubated with Alexa Fluor™ 594-

phalloidin (Cat. A12381, Thermo) for 1 h. Cell nucleus were stained with 1 μg/mL DAPI for 15 min. The cytoskeleton elements were photographed with a Zeiss LSM 800 confocal microscope.

**H&E staining, immunohistochemistry, and Masson staining**

Fixed tissues were embedded in paraffin and sectioned (5 μm). Following deparaffinized, the sections were subjected to antigen retrieval procedures with an EDTA antigen retrieval solution (Cat. P0086, Beyotime, Shanghai, China). Then, the slides were incubated with hematoxylin followed by counterstaining with eosin. For immunohistochemistry assay, tumor sections were incubated with primary antibodies overnight at 4 °C followed by incubation with HRP-conjugated secondary antibodies. The primary and secondary antibodies were listed in **Supplemental Table 17.** Protein expression in tumor sections was detected using a DAB kit (Cat. G1212, Servicebio, Wuhan, Hubei, China), followed by counterstaining with hematoxylin (Cat. G1004, Servicebio, Wuhan, Hubei, China). Images were acquired with an Olympus BX 53 microscope and analyzed with Image J software. For Masson staining, tissue sections were prepared with Masson Tricolor Staining Solution (Fast Green Method) kit (Cat. G1343, Solarbio, Beijing, China). Images were acquired with Olympus BX 53 microscope and analyzed with Image J software.

**RNA sequencing analysis**

Total RNA was isolated and purified by TRIzol reagent (Cat. 15596018, Invitrogen, Carlsbad, CA, USA) following the manufacturer manual. The RNA concentration and integrity were evaluated by NanoDrop ND-1000 (NanoDrop, Wilmington, DE, USA) and Bioanalyzer 2100 (Agilent, CA, USA). Then, poly (A) RNA was purified from 1μg

total RNA by Dynabeads Oligo (dT) 25-61005 (Thermo Fisher, CA, USA) using two rounds of purification. Then the poly(A) RNA was cut into pieces using Magnesium RNA Fragmentation Module (Cat. e6150, NEB, NY, USA) under 94 °C for 5-7 min. Then the fragmented RNA pieces were reversely transcribed into cDNA by SuperScript™ II Reverse Transcriptase (Cat.18064022, Invitrogen, USA) and sequenced with illumina Novaseq™ 6000 (LC-Bio Technology CO., Ltd., Hangzhou, China). Then, StringTie and edgeR were used to evaluate the expression levels of all transcripts. The differentially expressed mRNAs and genes were picked with log2 (fold change) >1 or log2 (fold change) <-1 and with statistical significance (*P* value < 0.05) by R package-edgeR. The volcano plot revealed the distributions of log2 fold change and *P* values for the differentially expressed genes. The GO terms (http://www.geneontology. org) of these differentially expressed genes were annotated.

**Migration and invasion assay**

Migration assay was performed with 24-well Boyden chambers (Corning, NY, USA) containing inserts of polycarbonate membranes with 8 μm-pores. Cells suspended with 100 μL of serum-free medium were seeded in the upper compartment ($3\times10^4$ HCT116 cells or $2\times10^4$ TPCs). The bottom chamber was filled with different chemoattractants. For invasion assay, the upper chamber was pre-coated with 30 μL of Matrigel (diluted at 3:1 using PBS) and incubated for 30 min. Then, PKH67-labeled HCT116 cells or DLD-1 cells ($5\times10^3$) mixed with TPCs ($2.5\times10^4$) were seeded into the upper chamber. The bottom chamber was filled with PM and DMEM (5:1). Following incubation for 48 h, the upper chamber was fixed with 4% paraformaldehyde for 30 min and then the cells were stained with 0.1% crystal violet. The non-migrated cells on the upper side of the membrane were removed with a cotton swab. The cells remaining on the lower

surface were photographed under an inverted microscope and analyzed with Image Pro Plus 6 software.

**Collagen gel contraction assay**

This experiment was performed with Cell Contraction Assay kit (Cat. CBA-5020, CELL BIOLABS, San Diego, CA, USA). TCF21-overexpressing or -knockdown TPCs were harvested and resuspended in PM at $5 \times 10^6$ cells/mL. Cold collagen gel working solution was prepared according to the instructions and mixed with the cell suspension at a ratio of 4: 1. 0.5 mL of the cell-collagen mixture per well was added in a 24-well plate. After incubating 1 h at 37 ℃, 1.0 mL culture medium was added into the collagen gel. Cultures were incubated for two days, and the collagen gels were gently released from the sides of the culture dishes with a sterile spatula. The collagen gel size (contraction index) was measured at 0, 6, 12, 24 and 48 h and quantified with Image J.

**Cell proliferation assay**

Cells ($5\times10^3$) were cultured overnight in 96-well plates. The next day, cells were treated with or without the culture supernatant of TPCs (48-h culture medium) and cell proliferation was determined by BeyoClick™ EdU Cell Proliferation kit (Cat. C0071S, Beyotime, Shanghai, China) and analyzed with Image Pro Plus 6 software.

**Adhesion assay**

TPCs with the overexpression or knockdown of TCF21 were collected, washed, and stained with PKH67 (Cat. MINI67, Sigma). PKH67-labeled TPCs ($2\times10^4$) were seeded in a 96-well plate for 2 h. Then, the media was removed, and cells were washed with



Li X, *et al. Gut* 2023; 72:710–721. doi: 10.1136/gutjnl-2022-327913

PBS twice to remove the non-adherent cells. Images were acquired with a fluorescent microscope (ZEISS) and analyzed with Image Pro Plus 6 software.

**Tube formation assay**

Tube formation was performed with a 96-well plate. Matrigel was first coated in the plates at 37 °C for 30 min. Then, HMEC-1 cells ($2\times10^4$) supplemented with 100 μL ECM were seeded in the Matrigel-coated plates. After 2-h incubation, ECM was replaced, and HMEC-1 cells were incubated with the conditioned medium of TPCs. The capillary tubes were photographed under an inverted light microscopy, and the number of tubes was analyzed by Image Pro Plus 6 software.

**Transmission electron microscopy analysis**

Tumor tissues were acquired and fixed in 2.5% glutaraldehyde (Cat. PH9003, Maya Reagent, Zhejiang, China). All samples were post-fixed in 1% osmium tetroxide (Cat. 23311-10, Polysciences, USA), dehydrated in graded concentration of alcohols, and then embedded in low-viscosity resin. The embedded tissues were sectioned and stained with saturated uranyl-acetate and Sato's lead-citrate. Sections were imaged using JEM1200EX transmission electron microscope equipped with BioScan600W digital camera (JEOL, Tokyo, Japan).

**Second harmonic generation and two-photon excited fluorescence (SHG/TPEF)**

Tumor tissues were acquired and fixed in 4% paraformaldehyde overnight followed by washing with PBS twice and sectioned (250 μm) using Vibration slice (Leica, VT1000S). Following blocking with 5% BSA solution for 1 h, the sections were incubated with anti-CD31 antibody at 4 °C overnight. The slides were then incubated

399 with Donkey anti-Goat IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa
400 Fluor™ 488 (RRID: AB_2534102, Invitrogen) and immersed in PBS for SHG/TPEF
401 microscopy (Nikon, Tokyo, Japan). TPEF was utilized for visualization of tumor
402 vessels stained by CD31 (red) and SHG was used to visualize collagen structure (green)
403 at 790 nm excitation light.
404
405 **Atomic Force Microscopy (AFM) measurement**
406 For collagen organization, coatings of isotropic collagen I and AFM detection were
407 performed according to previous report[13]. *In vitro* fibrillogenesis of collagen-I (Cat. 08-
408 115, Merck; 4.4 mg/mL) was initiated by diluting the collagen I solution tenfold in
409 fibrillogenic buffer (50 mM glycine, 200 mM KCl, pH 9.0). After mixture for 10 min
410 at room temperature, 60 μL of the diluted solution were added to the 22 mm-silicide
411 coverslips and incubated overnight at 37 °C. The next day, coverslips were washed with
412 PBS twice and plated in 12-well plate. Then, TPCs ($1\times10^4$) were seeded on the Collagen
413 I-coated coverslip and cultured for 5 days. AFM Imaging was performed with a
414 NanoWizard II AFM (JPK-Instruments, Berlin, Germany) mounted on an inverted
415 microscope (Axiovert 200, Zeiss, Jena, Germany). Scanning of samples was performed
416 at a scan rate of 0.25 Hz and five fields were recorded for each sample. For perivascular
417 stiffness measurements, AFM was performed according to modified published
418 procedures[14]. Tissues were acquired and embedded within OCT. Then, the frozen tissue
419 blocks were sectioned at a thickness of 20 μm and immersed in proteinase inhibitor-
420 contained PBS at room temperature. The next day, tumor sections were applied for
421 AFM quantification of Young's modulus (Bruker, USA). Briefly, silicon nitride
422 cantilevers with a spring constant of 0.15 N/m were attached by a borosilicate glass
423 spherical tip with a diameter of 5 μm. Cantilevers were tapping on the perivascular

region of tumor sections and five 15 μm × 15 μm AFM stiffness map (16 × 16 raster series) were acquired for each sample. The Young's modulus of the perivascular region in each section were determined by Hertz model. Tissue samples were assumed to be incompressible and a Poisson's ratio of 0.5 was used in the calculation of the Young's modulus.

**DNA extraction and bisulfite sequencing**

DNA was extracted using the Genomic DNA Purification Kit (Cat. A1120, Promega, WI, USA), which was followed by treatment with sodium bisulfite (Zymo Research, CA, USA). The converted DNA was purified and amplified for sequencing by Biossci Biotechnology Co. Ltd (Wuhan, Hubei, China). Primers of *TCF21* promoter bisulfite-modified regions were: Forward primer (5′-3′): TTTTTGATGTTTTGAAAATGATT - AGG; Reverse primer (5′-3′): CAACCACCTTC TCCCAACTATAA.

**Organotypic culture system**

Organotypic culture system was constructed with 12-well Boyden chambers (Corning, NY, USA) containing inserts of polycarbonate membranes with 0.4 μm-pores. HMEC-1 cells ($2\times10^4$) supplemented with 100 μL ECM were seeded in the upper chamber and incubated overnight. The next day, ECM was removed, and TPCs ($5 \times 10^5$) were embedded in 1.0 mL collagen I and plated in the chamber. The lower chamber was filled with complete PM and ECM (PM: ECM, 5:1). Following incubation at 37 °C, 5% $CO_2$ for 5 days, the mechanical properties of the matrix, complex modulus (G*), were determined with a rheometer (Malvern Kinexus pro$^+$, USA) according to previous report[15]. The elastic modulus (E) was determined from G* by assuming a poison's ratio (ν) of 0.5 with the expression $E=2G^*(1+\nu)$ to allow comparison to other published

work. For invasion assay, TCF21-overexpressing TPCs ($5 \times 10^5$) were embedded in a matrix mixture of 0.25 mL collagen I and 0.75 mL Matrigel and plated in the chamber. Following incubation for 5 days, PKH67-labeled HCT116 cells ($3 \times 10^4$) supplemented with 100 μL complete DMEM were plated on the top of matrix and further cultured for 1 day. At the end of experiment, the whole matrix was fixed in 4% overnight and subjected to immunofluorescence analysis. The migrated HCT116 cells were detected by staining of EpCAM. The invaded cells were observed under a Zeiss LSM 800 confocal microscope and analyzed with Image Pro Plus 6 software.

**Statistical analysis**

The statistical values were calculated with GraphPad Prism 8.0 (GraphPad Software, Inc., San Diego, CA). Differences between two groups were evaluated with two-tailed unpaired *t*-test or Mann Whitney test. Differences among three groups or more were evaluated using one-way ANOVA followed by Tukey's post hoc test. Survival curves were plotted using the Kaplan Meier method and compared using the log-rank test. The receiver operating characteristic (ROC) curves were performed and the area under ROC curve (AUC) was calculated by logistic regression model to evaluate the diagnostic accuracy. Comparisons of variables were performed using Fisher's exact test or chi-squared test based on their categorical data. Multivariable logistic regression was used to analyze the predictors of CRC metastasis. $P < 0.05$ was considered as significant difference.

**References**

1. Percie du Sert N, Hurst V, Ahluwalia A, *et al*. The ARRIVE guidelines 2.0: Updated guidelines for reporting animal research. *PLoS Biol* 2020; 18: e3000410.

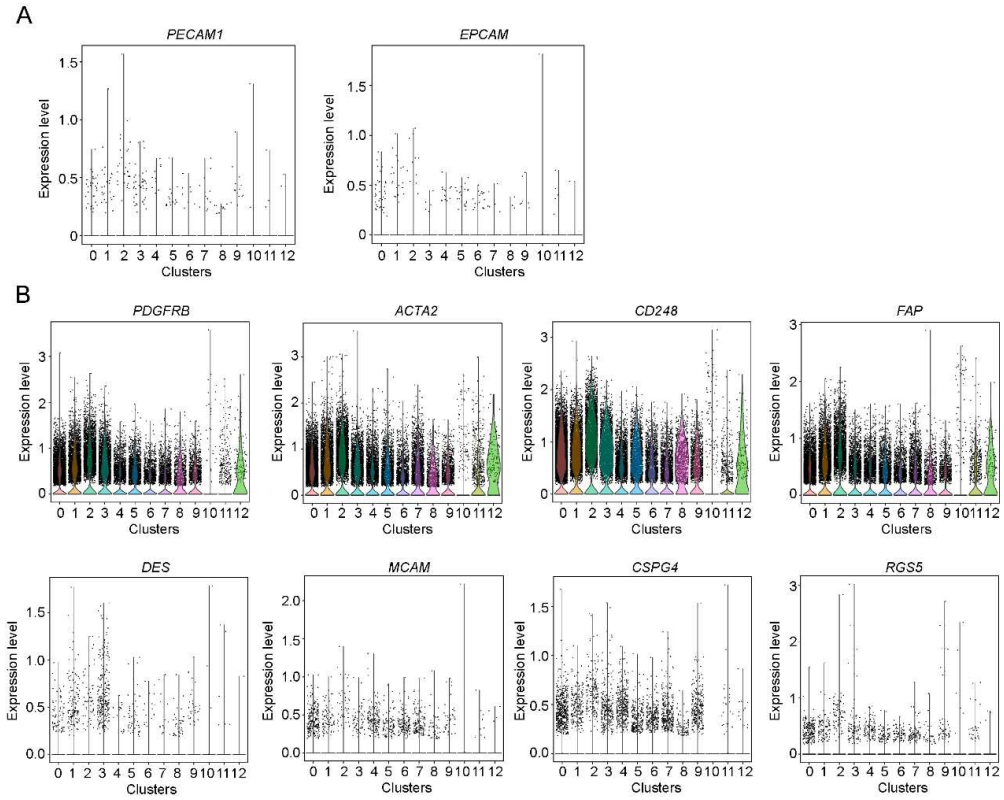
2. Aibar S, González-Blas CB, Moerman T, *et al*. SCENIC: single-cell regulatory

network inference and clustering. *Nat Methods* 2017; 14: 1083-6.
3. Huynh-Thu VA, Irrthum A, Wehenkel L, *et al*. Inferring regulatory networks from expression data using tree-based methods. *PLoS One* 2010; 5: e12776.
4. Shannon P, Markiel A, Ozier O, *et al*. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003; 13: 2498-504.
[dataset] [5] Kinchen J, Chen HH, Parikh K, et al. Data from: Composition of the Colonic Mesenchyme and the Nature of its Plasticity in Inflammatory Bowel Disease. *Gene Expression Omnibus*, Sep 27, 2018. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc= GSE114374.
[dataset] [6] Rao M, Oh K, Moffitt R, *et al*. Single-Cell RNA-Seq Analysis of Metastasis-Primary Tumor Dissimilarity in a Patient with Gastrointestinal Neuroendocrine Cancer. *Gene Expression Omnibus*, Feb 01, 2020. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi.
7. Harney AS, Arwert EN, Entenberg D, *et al*. Real-Time Imaging Reveals Local, Transient Vascular Permeability, and Tumor Cell Intravasation Stimulated by $TIE2^{hi}$ Macrophage-Derived VEGFA. *Cancer Discov* 2015; 5: 932-43.
8. Szczerba BM, Castro-Giner F, Vetter M, *et al*. Neutrophils escort circulating tumour cells to enable cell cycle progression. *Nature* 2019; 566: 553-7.
9. Gkountela S, Castro-Giner F, Szczerba BM, *et al*. Circulating Tumor Cell Clustering Shapes DNA Methylation to Enable Metastasis Seeding. *Cell* 2019; 176: 98-112.e14.
10. Nagrath S, Sequist LV, Maheswaran S, *et al*. Isolation of rare circulating tumour cells in cancer patients by microchip technology. *Nature* 2007; 450: 1235-9.
11. Ozkumur E, Shah AM, Ciciliano JC, *et al*. Inertial focusing for tumor antigen-dependent and -independent sorting of rare circulating tumor cells. *Sci Transl Med* 2013; 5: 179ra47.
12. Ramani VC, Lemaire CA, Triboulet M, *et al*. Investigating circulating tumor cells and distant metastases in patient-derived orthotopic xenograft models of triple-negative breast cancer. *Breast Cancer Res* 2019; 21: 98.
13. Kirmse R, Otto H, Ludwig T. Interdependency of cell adhesion, force generation and extracellular proteolysis in matrix remodeling. *J Cell Sci* 2011; 124: 1857-66.
14. Laklai H, Miroshnikova YA, Pickup MW, *et al*. Genotype tunes pancreatic ductal adenocarcinoma tissue tension to induce matricellular fibrosis and tumor progression.

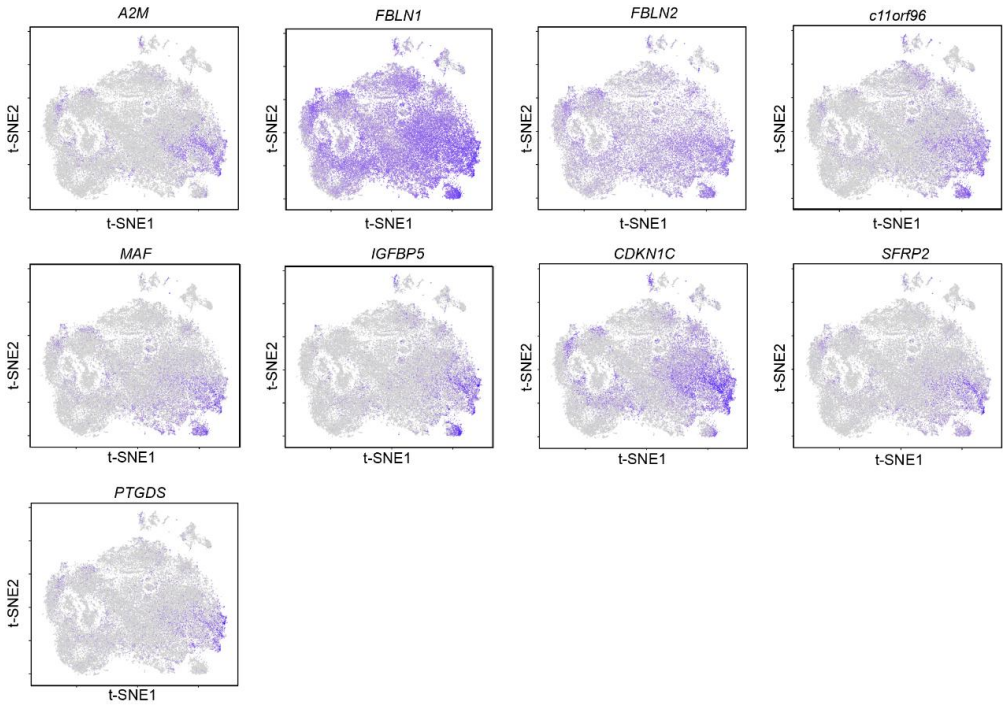
*Nat Med* 2016; 22: 497-505.

15. Li H, Wijekoon A, Leipzig ND. 3D differentiation of neural stem cells in macroporous photopolymerizable hydrogel scaffolds. *PLoS One* 2012; 7: e48824.

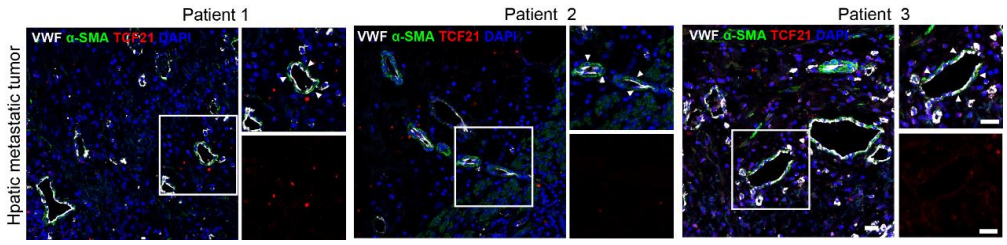
## Supplemental figures and figure legends



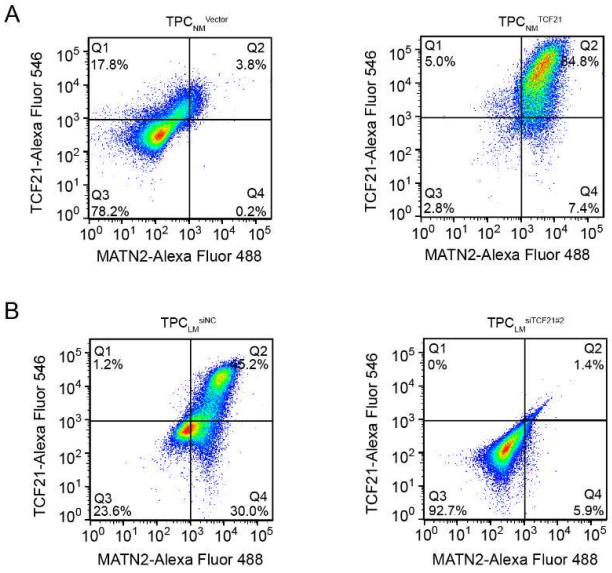

**Supplemental Figure 1. Transcriptomic characterization of 13 subsets of TPCs. (A)** Gene expression profiles of *PECAM1* and *EPCAM* in distinct subsets of TPCs. **(B)** Gene expression profiles of *PDGFRB*, *ACTA2*, *CD248*, *FAP*, *DES*, *MCAM*, *CSPG4,* and *RGS5* in distinct subsets of TPCs.



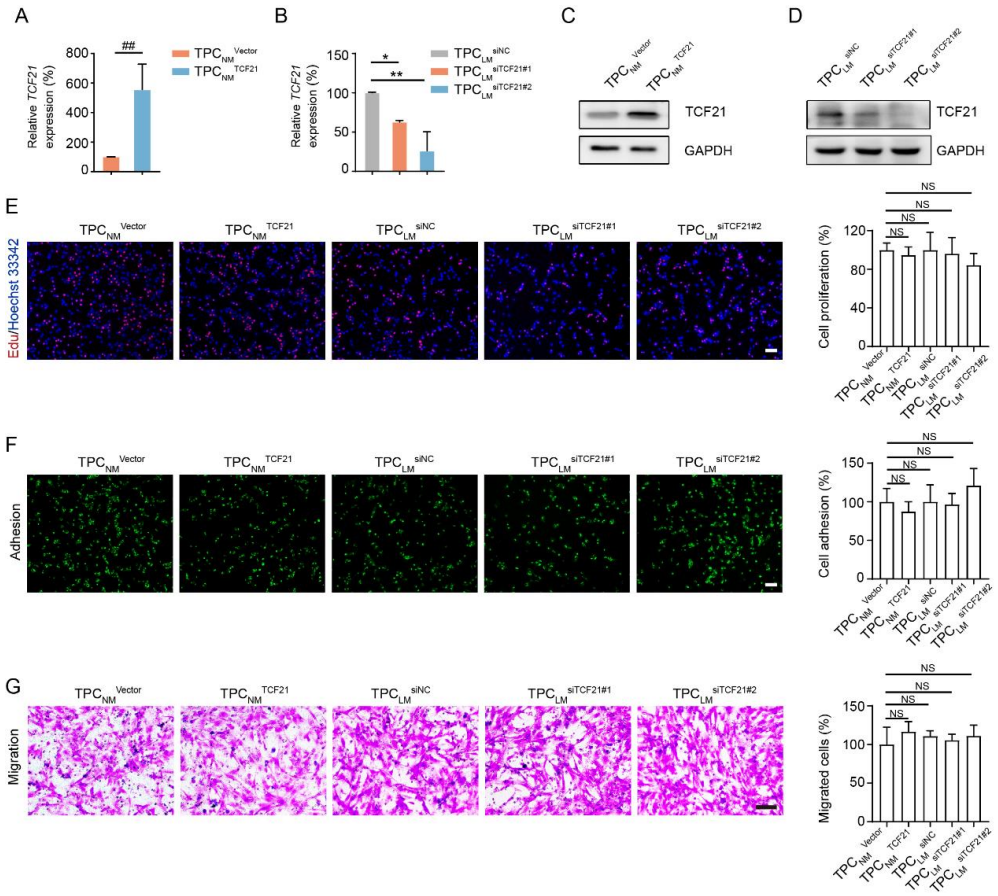

**Supplemental Figure 2. Distribution of the matrix-pericyte-related genes in all subsets of TPCs.** t-SNE visualization of gene distribution of *A2M*, *FBLN1*, *FBLN2*, *c11orf96*, *MAF*, *IGFBP5*, *CDKN1C*, *SFRP2* and *PTGDS* in all subsets of TPCs.



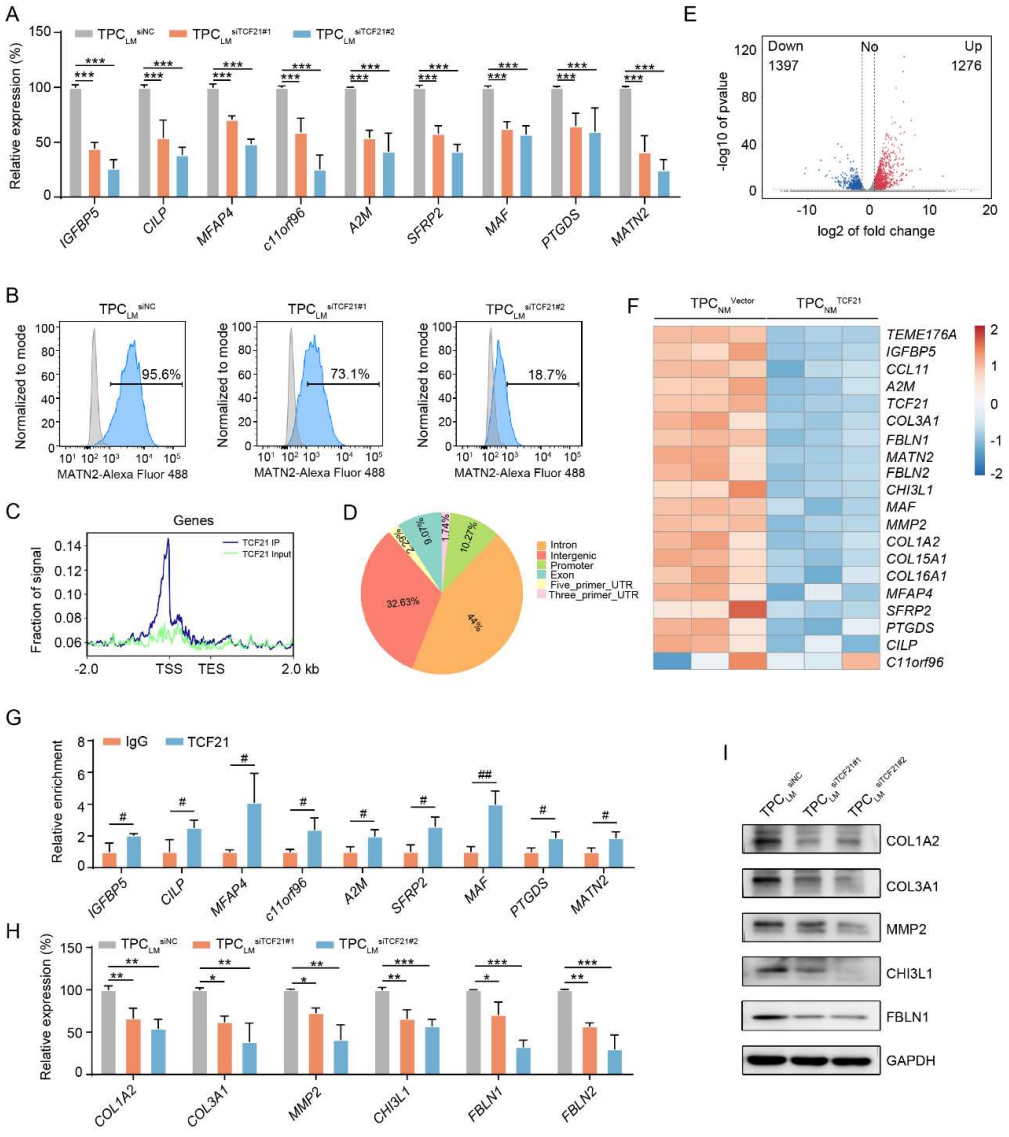

**Supplemental Figure 3. Determination of TCF21 in TPCs in the hepatic metastatic tumors from CRC patients.** Representative images of TCF21 staining (red) in TPCs ($\alpha SMA^+$, green) in the hepatic metastatic tumors from CRC patients (n = 20). Scale bar, 20 μm.




**Supplemental Figure 4. The expression of TCF21 is positively correlated with MATN2 in TPCs. (A)** FCM analysis of the $TCF21^{+}MATN2^{+}$ TPCs in $TPC_{NM}$ infected with *TCF21* lentivirus ($TPC_{NM}^{TCF21}$) or Vector ($TPC_{NM}^{Vector}$) (n = 3). **(B)** FCM analysis of the $TCF21^{+}MATN2^{+}$ TPCs in $TPC_{LM}$ transfected with siRNA targeting TCF21 ($TPC_{LM}^{siTCF21}$) or negative control ($TPC_{LM}^{siNC}$) (n = 3).

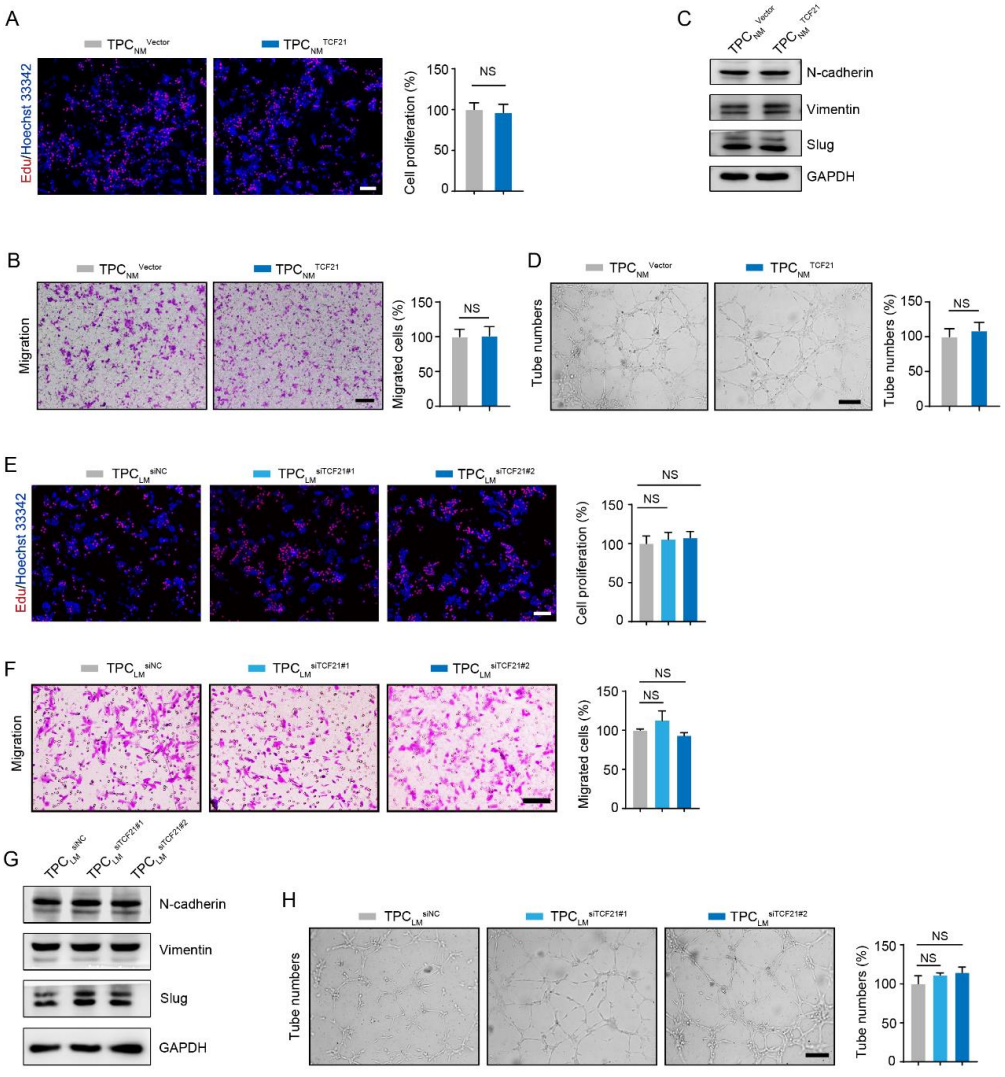



**Supplemental Figure 5. Effects of TCF21 on the proliferation, adhesion, and migration of TPCs. (A, B)** RT-qPCR analysis of *TCF21* mRNA levels in TCF21-overexpressing **(A)** or -knockdown **(B)** TPCs (n = 3). **(C, D)** Western blotting analysis of TCF21 in TCF21-overexpressing **(C)** or -knockdown (**D)** TPCs (n = 3). **(E)** Representative images and quantification of cell proliferation in TCF21-overexpressing and -knockdown TPCs (n = 3). Scale bar, 100 μm. **(F)** Representative images and quantification of TPC adhesion (n = 3). Scale bar, 100 μm. **(G)** Transwell assay for the migration of TPCs. Quantification of the migrated TPCs is shown (n = 3). Scale bar, 200 μm. Data are presented as mean ± SEM. NS, not significant. $^{\#\#}P < 0.01$ by two-tailed unpaired *t*-test; NS, $^{*}P < 0.05$, $^{**}P < 0.01$ by one-way ANOVA followed by Tukey's post hoc test.

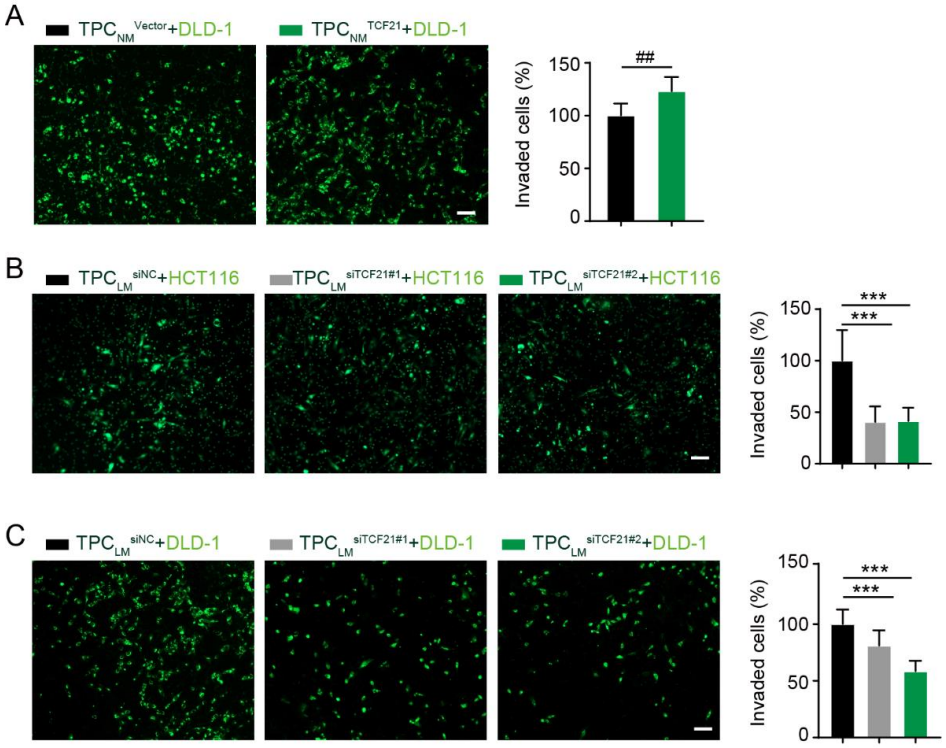



**Supplemental Figure 6. TCF21 stimulates the generation of matrix-pericytes and induces ECM remodeling. (A)** RT-qPCR analysis of matrix-pericyte-specific genes in $TPC_{LM}$ transfected with siRNA targeting TCF21 ($TPC_{LM}^{siTCF21}$) or negative control ($TPC_{LM}^{siNC}$) (n = 3). **(B)** FCM analysis of MATN2 expression in TCF21-knockdown TPCs (n = 3). **(C)** ChIP-seq summary plot of TCF21 enrichment across the indicated genomic distance in TCF21-overexpressing TPCs (n = 3). **(D)** The distribution of TCF21 peaks on gene elements (n = 3). **(E)** Volcano Plot of TCF21 regulated genes (n

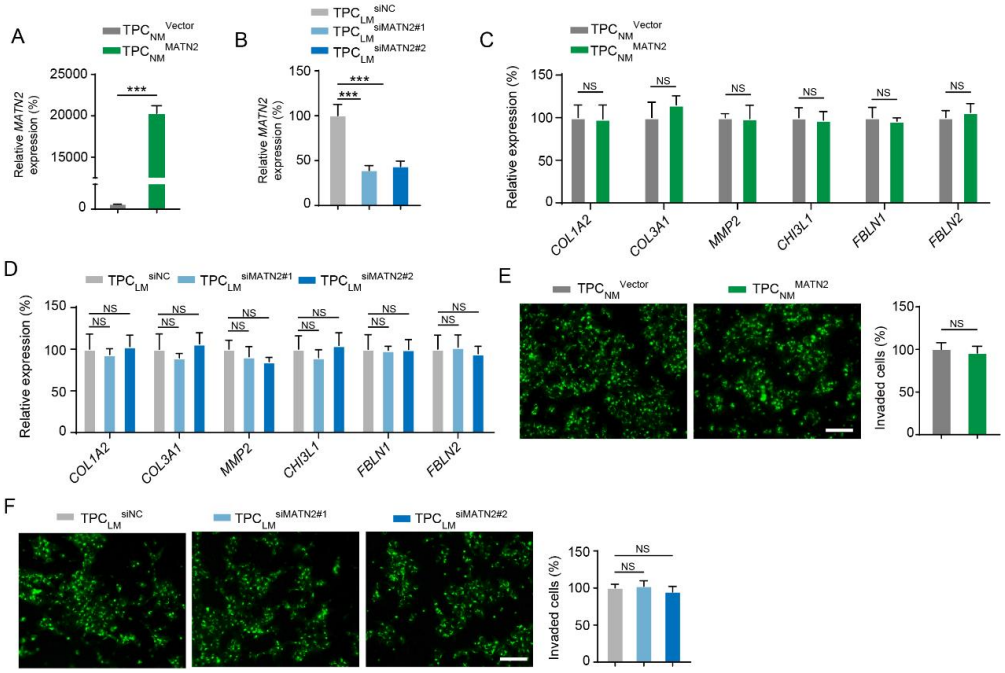
= 3); Red dots represent the up-regulated genes and blue dots represent the down-regulated genes. **(F)** Heat maps of the differentially expressed genes between $TPC_{NM}^{Vector}$ and $TPC_{NM}^{TCF21}$ (n = 3). **(G)** ChIP-qPCR analysis of TCF21 binding at the promoter of indicated genes in $TPC_{NM}^{TCF21}$ (n = 3). **(H)** RT-qPCR analysis of differentially expressed genes in $TPC_{LM}^{siNC}$ and $TPC_{LM}^{siTCF21}$. (**I**) Western blotting analysis of COL1A2, COL3A1, MMP2, CHI3L1, and FBLN1 in $TPC_{LM}^{siNC}$ and $TPC_{LM}^{siTCF21}$ (n = 3). Data are presented as mean ± SEM, $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$ by one-way ANOVA followed by Tukey's post hoc test, $^{\#}P < 0.05$, $^{\#\#}P < 0.01$ by two-tailed unpaired *t*-test.



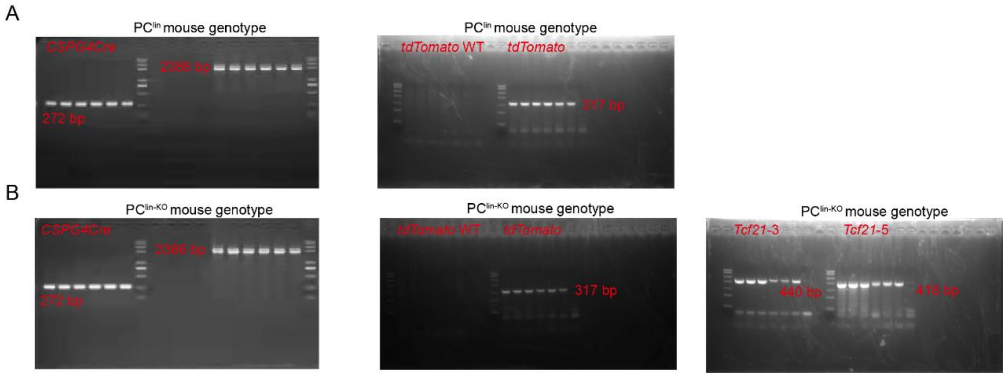
**Supplemental Figure 7. TCF21 in TPCs has negligible effects on cell migration and angiogenesis. (A)** EdU assay for the proliferation of HCT116 cells primed with conditioned medium from $TPC_{NM}^{Vector}$ and $TPC_{NM}^{TCF21}$ for 48 h (n = 3)**.** Scale bar, 100 μm. **(B)** Transwell assay for cell migration of HCT116 cells. HCT116 cells were seeded on the upper chamber of the transwell and the bottom compartment was filled with conditioned medium of $TPC_{NM}^{Vector}$ and $TPC_{NM}^{TCF21}$. After 48 h, the migrated cells were imaged and counted (n = 3). Scale bar, 100 μm. **(C)** Western blotting analysis of EMT markers in HCT116 cells primed with conditioned medium of $TPC_{NM}^{Vector}$ and $TPC_{NM}^{TCF21}$ (n = 3). **(D)** Representative images and quantification of tube numbers formed by HMEC-1 cells. HMEC-1 cells suspended with ECM were seeded on the Matrigel coated 96-well plated. After 2 h, ECM were replaced with the conditioned medium of $TPC_{NM}^{Vector}$ and $TPC_{NM}^{TCF21}$, and the number of formed tubes was calculated 2 hours later (n = 3). Scale bar, 100 μm. **(E)** EdU assay for the proliferation of HCT116 cells primed with conditioned medium of $TPC_{LM}^{siNC}$ and $TPC_{LM}^{siTCF21}$ (n = 3)**.** Scale bar, 100 μm. **(F)** Representative images and quantification of migrated HCT116 cells. HCT116 cells were seeded on the upper chamber of the transwell and the bottom chamber was filled with the conditioned medium of $TPC_{LM}^{siNC}$ and $TPC_{LM}^{siTCF21}$. After 48 h, the migrated cells were imaged and counted (n = 3). Scale bar, 100 μm. **(G)** Western blotting analysis of EMT markers in HCT116 cells primed with conditioned medium of $TPC_{LM}^{siNC}$ and $TPC_{LM}^{siTCF21}$ (n = 3). **(H)** Tube formation assay for HMEC-1 cells treated with conditioned medium of $TPC_{LM}^{siNC}$ and $TPC_{LM}^{siTCF21}$ as indicated in **(D)** (n = 3). Scale bar, 100 μm. Data are presented as mean ± SEM, NS, not significant. Two-tailed unpaired *t*-test (**A**, **B**, **D**), one-way ANOVA followed by Tukey's post hoc test (**E**, **F**, **H**).



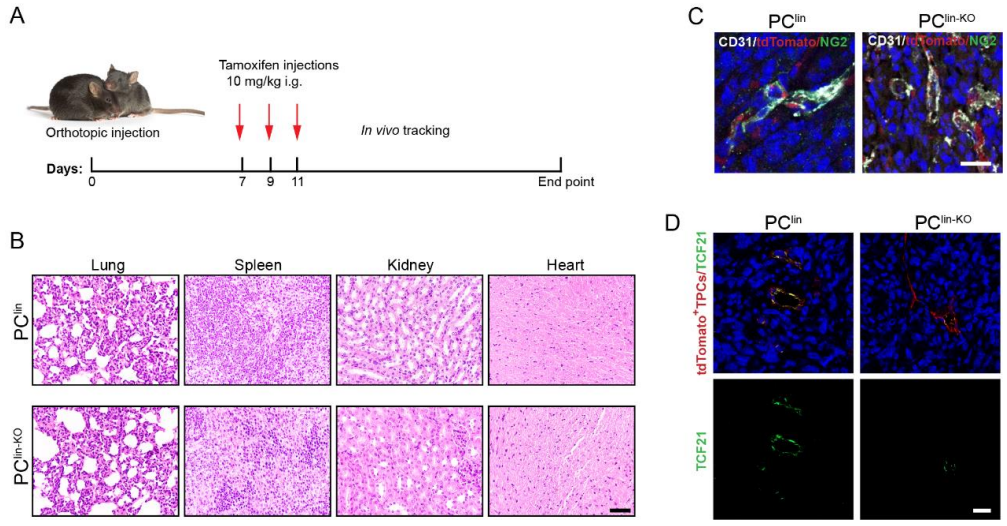

**Supplemental Figure 8. TCF21 in TPCs promotes invasion of CRC cells. (A)** Representative images and quantification of invaded DLD-1 cells (green). DLD-1 cells mixed with $TPC_{NM}^{Vector}$ or $TPC_{NM}^{TCF21}$ were seeded into the Matrigel-coated transwell. The invaded DLD-1 cells were photographed and counted after 48 h (n = 3). Scale bar, 100 μm. **(B, C)** Transwell assay for invasion of HCT116 cells and DLD-1 cells. HCT116 cells **(B)** or DLD-1 cells **(C)** were pre-mixed with $TPC_{LM}^{siNC}$ or $TPC_{LM}^{siTCF21}$ and subjected to invasion assay as indicated in **(A)** (n = 3). Scale bar, 100 μm. Data are presented as mean ± SEM. $^{\#\#}P < 0.01$ by two-tailed unpaired *t*-test, $^{***}P < 0.001$ one-way ANOVA followed by Tukey's post hoc test




**Supplemental Figure 9. MATN2 has negligible effects on ECM remodeling and CRC metastasis. (A, B)** RT-qPCR analysis of *MATN2* in MATN2-overexpressing **(A)** or -knockdown (**B)** TPCs (n = 3). **(C)** RT-qPCR analysis of the ECM-related genes in $TPC_{NM}^{Vector}$ and $TPC_{NM}^{MATN2}$ (n = 3). **(D)** RT-qPCR analysis of the indicated genes in $TPC_{LM}^{siNC}$ and $TPC_{LM}^{siMATN2}$ (n = 3). **(E, F)** Representative images and quantification of the invaded DLD-1 cells (green). DLD-1 cells pre-mixed with MATN2-overexpressing **(E)** or -knockdown **(F)** TPCs were seeded into the Matrigel-coated transwell. The invaded DLD-1 cells were photographed and counted after 48 h (n = 3). Scale bar, 100 μm. Data are presented as mean ± SEM, NS, not significant. $^{***}P$ <0.001. Two-tailed unpaired *t*-test (**A**, **C**, **E**), one-way ANOVA followed by Tukey's post hoc test (**B**, **D**, **F**).

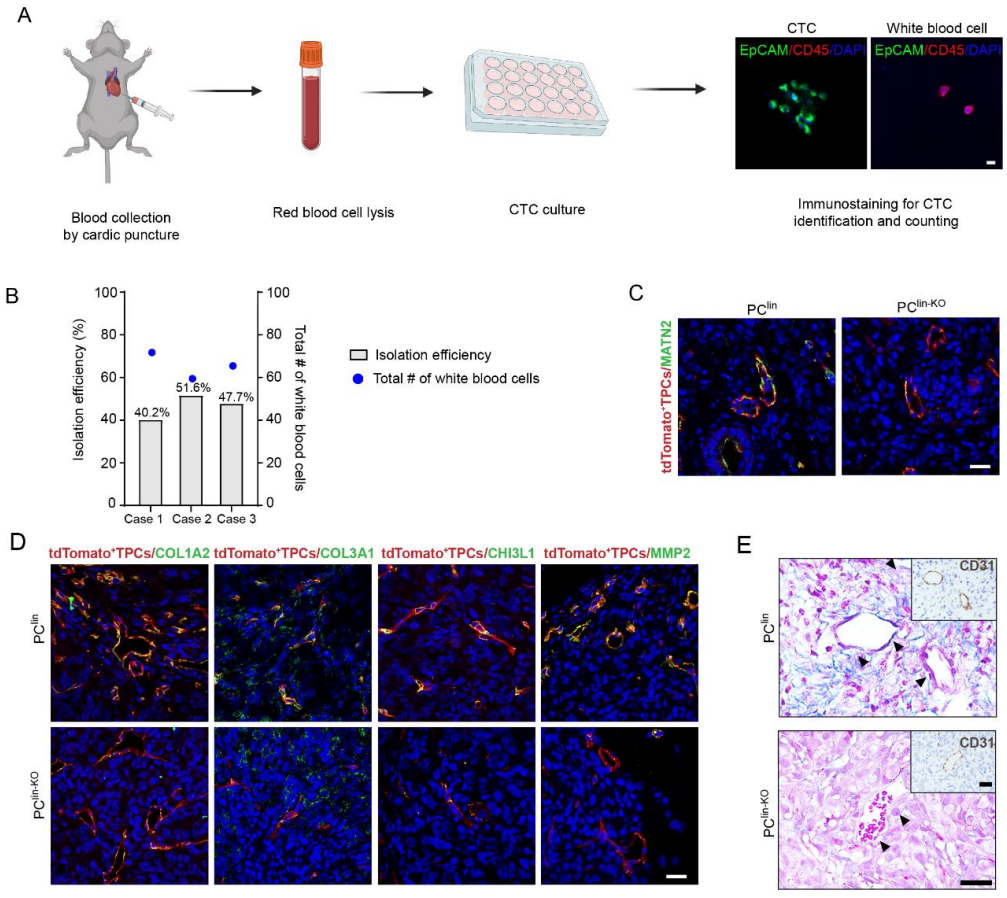



**Supplemental Figure 10. Mouse genotyping. (A)** PCR analysis of the genotype of $PC^{lin}$ mouse (n = 6). **(B)** PCR analysis of the genotype of $PC^{lin\text{-}KO}$ mice (n = 6). All mice were analyzed by PCR genotyping. PCR analysis of *Cspg4Cre* showed two bands identifying homozygous *Cspg4Cre* at 2386 bp (knock in) and 272 bp (wild type); PCR analysis of *tdTomato* showed homozygous *tdTomato* with a band at 317 bp (knock in) and no signal at 479 bp (wild type); PCR analysis of *Tcf21* indicated homozygous $Tcf21^{flox/flox}$ at 418 bp and 440bp.




**Supplemental Figure 11. Characterization of $PC^{lin}$ and $PC^{lin\text{-}KO}$ mice. (A)** Schematic diagram describing the experimental design of *in vivo* experiments. $PC^{lin}$ mice and $PC^{lin\text{-}KO}$ mice were orthotopically injected with MC38-luc-LM3 cells. After 7

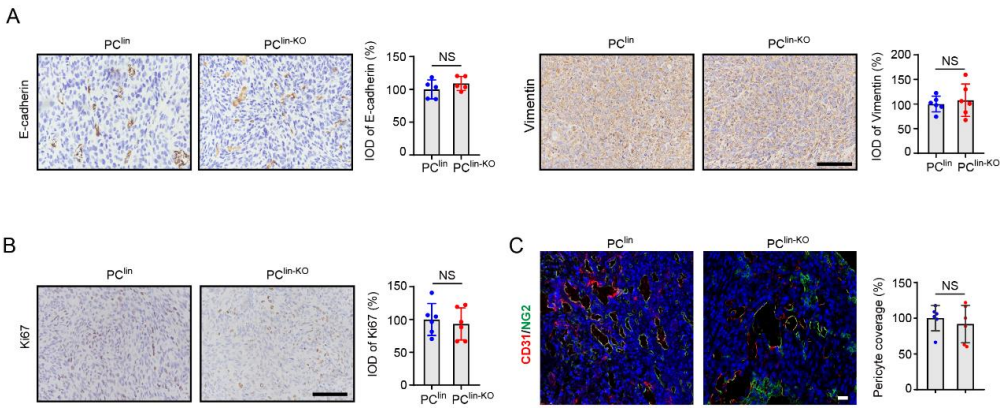
631 days, all mice were treated with tamoxifen (10 mg/kg) through intragastric
632 administration every other day for three times. *In vivo* tracking was performed to detect
633 tumor liver metastasis. **(B)** Representative images of H&E staining of lung, spleen,
634 kidney, heart, and liver derived from $PC^{lin}$ and $PC^{lin\text{-}KO}$ mice (n = 6). Scale bar, 50 μm.
635 **(C)** Immunofluorescence analysis of TPCs (tdTomato) in tumor sections by NG2 (green)
636 staining (n = 6). Scale bar, 20 μm. **(D)** Immunofluorescence analysis of TCF21 (green)
637 in TPCs (tdTomato) from $PC^{lin}$ mice and $PC^{lin\text{-}KO}$ mice (n = 6). Scale bar, 20 μm.
638



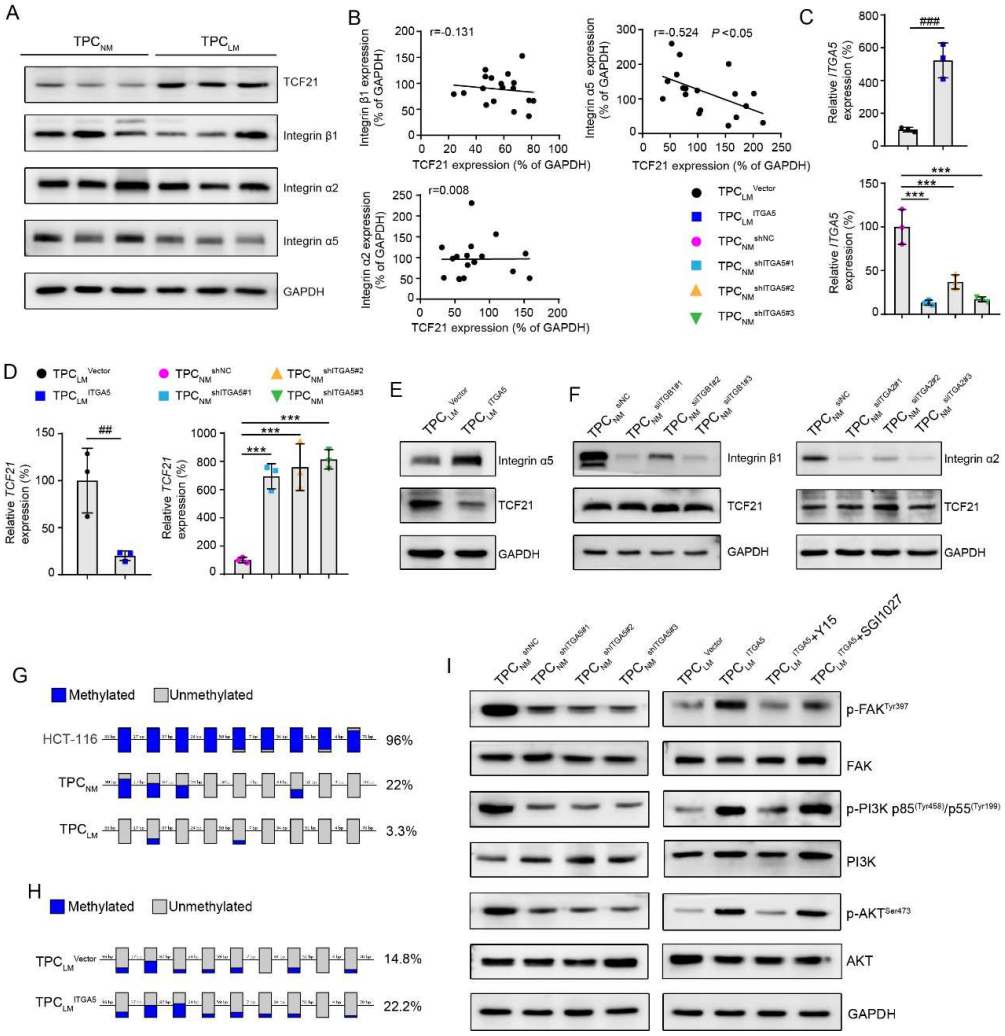
639

640 **Supplemental Figure 12. Pericyte-specific knockout of *Tcf21* inhibits perivascular**
641 **ECM remodeling and CRCLM. (A)** Schematic diagram of the isolation and
642 identification of CTCs. Scale bar, 20 μm. **(B)** Isolation efficiency of CTCs in the blood

spiked with MC38-luc-LM3 cells (n = 3). **(C)** Immunofluorescence staining for MATN2 (green) in TPCs (tdTomato) in primary tumor sections from MC38 allografts (n = 6). Scale bar, 20 μm. **(D)** Immunofluorescence staining for COL3A1, MMP2, COL1A1 and CHI3L1 (green) in TPCs (tdTomato) from primary tumor sections of MC38 allografts (n = 6). Scale bar, 20 μm. **(E)** Masson staining for perivascular collagen in primary tumor sections from MC38 allografts. Tumor vessels were labeled with CD31 (n = 6). Black arrows indicate the perivascular collagen fibers. Scale bar, 50 μm.

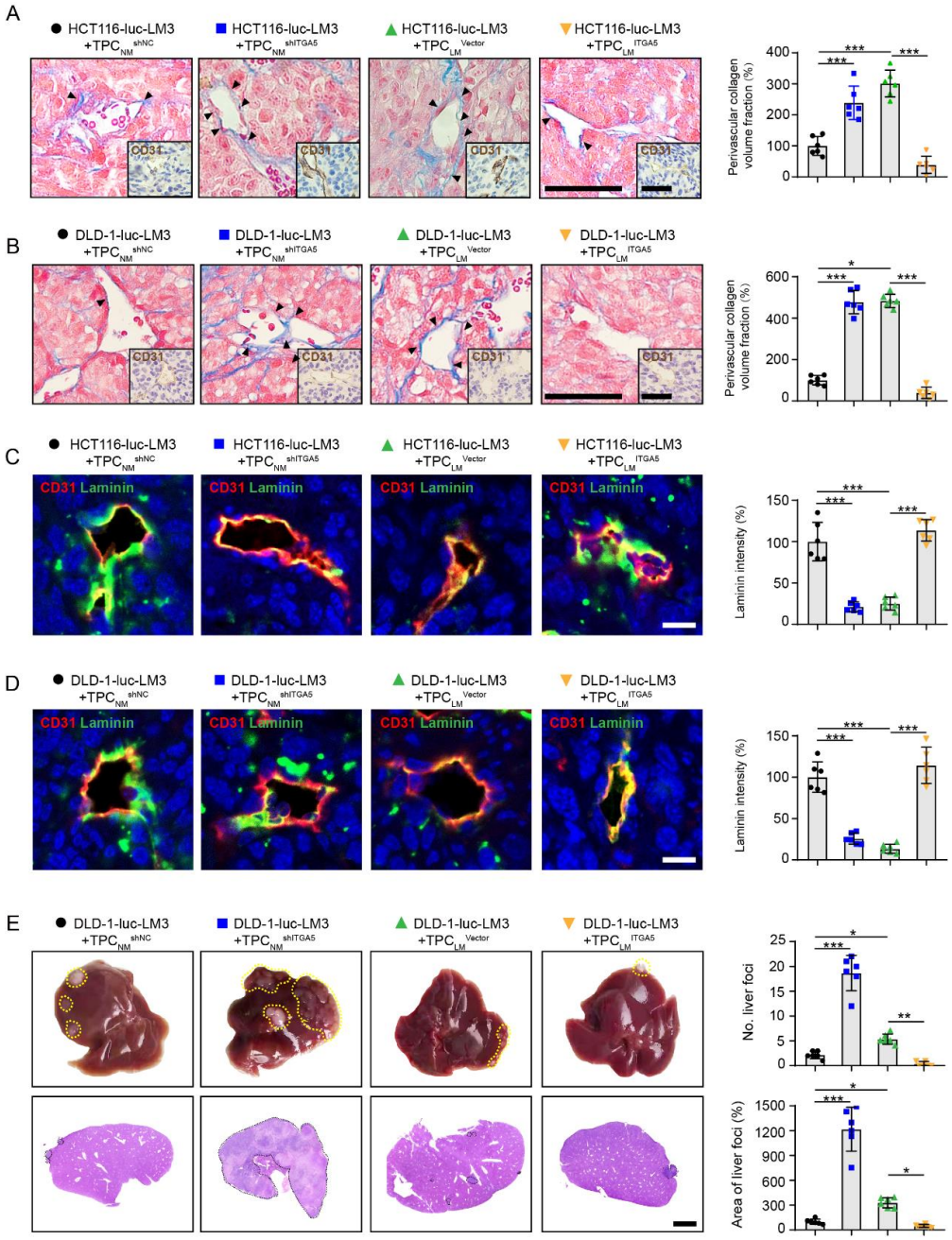



**Supplemental Figure 13. Pericyte-specific knockout of *Tcf21* has negligible effects on EMT and proliferation of CRC cells. (A)** Immunohistochemical staining and quantification of E-cadherin and Vimentin in orthotopic MC38 tumor tissues (n = 6). Scale bar, 50 μm. **(B)** Representative images and quantification of Ki67 staining in orthotopic MC38 tumor sections (n = 6). Scale bar, 50 μm. **(C)** Immunofluorescence staining and quantification of pericyte coverage as indicated by CD31 (red) and NG2 (green) in tumor sections (n = 6). Scale bar, 20 μm. Data are presented as mean ± SEM. NS, not significant. NS by two-tailed unpaired *t*-test.



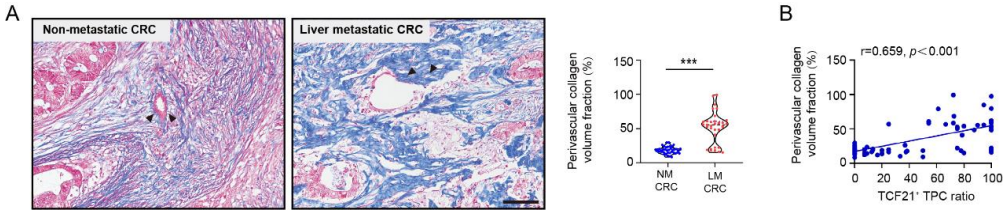

**Supplemental Figure 14. Effects of integrin α2 and integrin β1 on TCF21 in TPCs.** **(A)** Western blotting analysis of TCF21, integrin β1, integrin α2, and integrin α5 in $TPC_{NM}$ and $TPC_{LM}$ (n = 3). **(B)** Pearson's correlation analysis of the expression of integrin β1, integrin α2 and integrin α5 with TCF21 in TPCs (n = 3). **(C)** RT-qPCR analysis of the levels of *ITGA5* in the integrin α5-knockdown or -overexpressing TPCs (n = 3). **(D)** RT-qPCR analysis of the levels of *TCF21* in integrin α5-knockdown or -overexpressing TPCs (n = 3). **(E)** Western blotting analysis of integrin α5 and TCF21 in integrin α5-overexpressing $TPC_{LM}$ (n = 3). **(F)** Western blotting analysis of integrin α2, integrin β1 and TCF21 in integrin α2 or integrin β1-knockdown $TPC_{NM}$ (n = 3). **(G)**

Bisulfite DNA sequencing analysis of *TCF21* promoter region in TPCs or HCT116 cells. Blue and gray circles represent methylated and unmethylated CpGs, respectively. The percentage of total methylated CpGs is given on right of each dataset (n = 3). **(H)** Bisulfite DNA sequencing analysis of *TCF21* promoter region in integrin α5-overexpressing $TPC_{LM}$ (n = 3). **(I)** Western blotting analysis of FAK/PI3K/AKT signaling axis in integrin α5-knockdown or -overexpressing TPCs with or without FAK inhibitor (Y15) or DNMT1 inhibitor (SGI1027) treatment (n = 3). Data are presented as mean ± SEM, $^{\#\#}P$ <0.01, $^{\#\#\#}P$ <0.001 by two-tailed unpaired *t*-test, $^{***}P$ <0.001 by one-way ANOVA followed by Tukey's post hoc test.

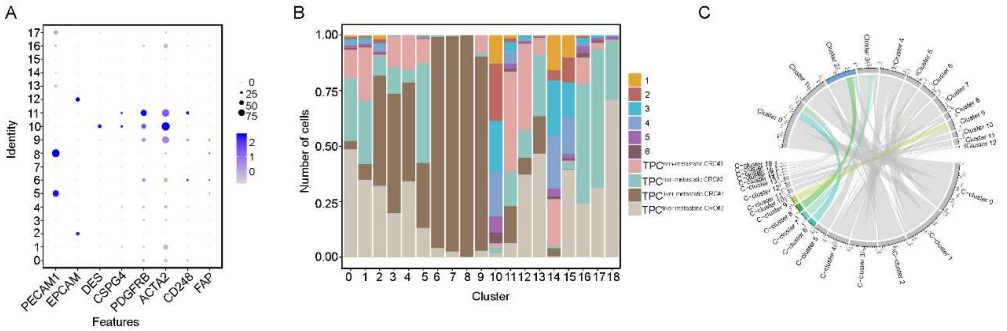



**Supplemental Figure 15. Loss of integrin α5 in TPCs promotes perivascular ECM remodeling and CRCLM. (A, B)** Representative images of Masson and CD31 staining in primary tumor sections (n = 6). Black arrows indicate the perivascular collagen fibers. Scale bar, 50 μm. The quantification of perivascular collagen volume fraction was shown in the right. **(C, D)** Immunofluorescence staining and quantification of laminin

(green) around the CD31$^+$ tumor vessels (red) in HCT116-luc-LM3 xenografts **(C)** and DLD-1-luc-LM3 xenografts **(D)** (n = 6). Scale bar, 20 μm. **(E)** Representative images and H&E staining of liver metastatic foci (n = 6). Yellow and black dotted lines indicate the metastatic loci. Scale bar, 2 mm. Data are presented as mean ± SEM. $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$ by one-way ANOVA followed by Tukey's post hoc test.

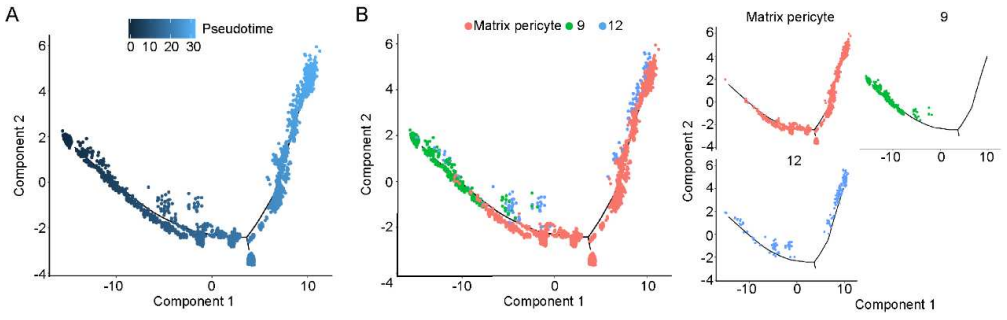



**Supplemental Figure 16. TCF21 in TPCs is associated with perivascular ECM deposition.** (A) Masson staining and quantification of perivascular collagen in tumors derived from CRC patients with non-metastasis or liver metastasis (n = 75). Scale bar, 50 μm. $^{***}P < 0.001$ by two-tailed Mann-Whitney test. (B) Pearson's correlation analysis of perivascular collagen volume fraction and TCF21$^+$ TPC ratio (n = 75). NM CRC, non-metastatic colorectal cancer, LM CRC, liver metastatic colorectal cancer.



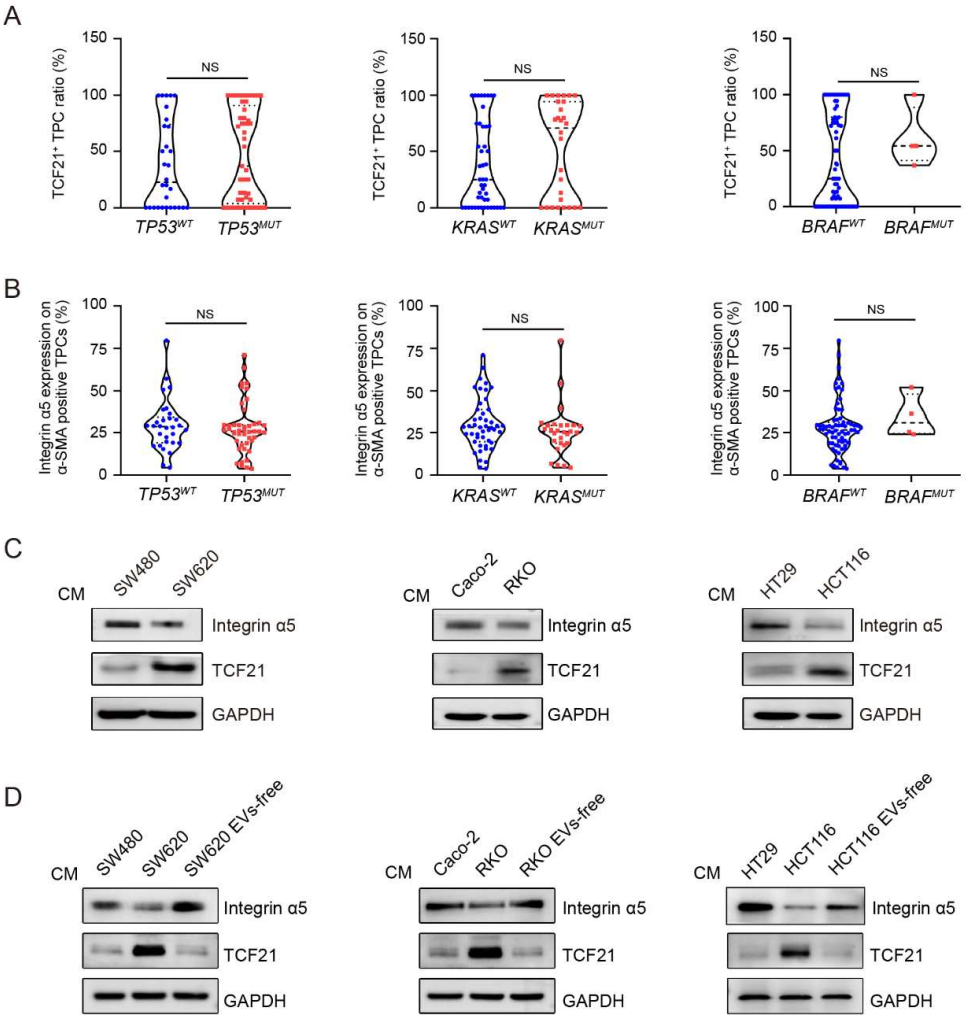

**Supplemental Figure 17. Comparison of the scRNA-seq data derived from TPCs and the previous published data. (A)** Dot plots for gene expressions in pericytes

acquired from previous studies. Raw data of pericyte scRNA-seq were collected from six samples derived from previous studies (GEO accession GSM3140596, GSM3140595, GSM3140594, and GSM3140593) and (GEO accession GSM4159165 and GSM4159164). Total cells derived from previous studies were classified into 18 clusters named L-Cluster 0 to L-Cluster 17. Among them, L-Cluster 3, L-Cluster 9, L-Cluster 10, and L-Cluster 11 were subjected for further analysis as these four populations were positive for *DES*, *CSPG4*, *PDGFRB*, *ACTA2*, *CD248* and *FAP* (pericyte markers), but negative for *PECAM1* (endothelial cell marker) and *EPCAM* (epithelial cell marker). **(B)** Analysis of the data of the four populations (L-Cluster 3, L-Cluster 9, L-Cluster 10, and L-Cluster 11) derived from six samples (Sample 1-6) in **(A)** and our scRNA-seq data (Cluster 0-12). The combined pericytes were categorized into 19 subpopulations, termed C-Cluster 0 to C-Cluster 18. Among them, C-cluster 0-5, C-cluster 10-18 were presented both in our data and the extended data (Sample1-6); however, C-cluster 6-9 were specifically revealed in our data, indicating that the existing pericyte clusters originated from the previous research[5, 6] were included in our scRNA-seq data and we discovered four new subsets. **(C)** Comparative analysis of pericyte cluster derived from **(B)** (C-cluster 0-18) with our data (Cluster 0-12). Among them, C-Cluster 8 was included in Cluster 2 (matrix pericytes).



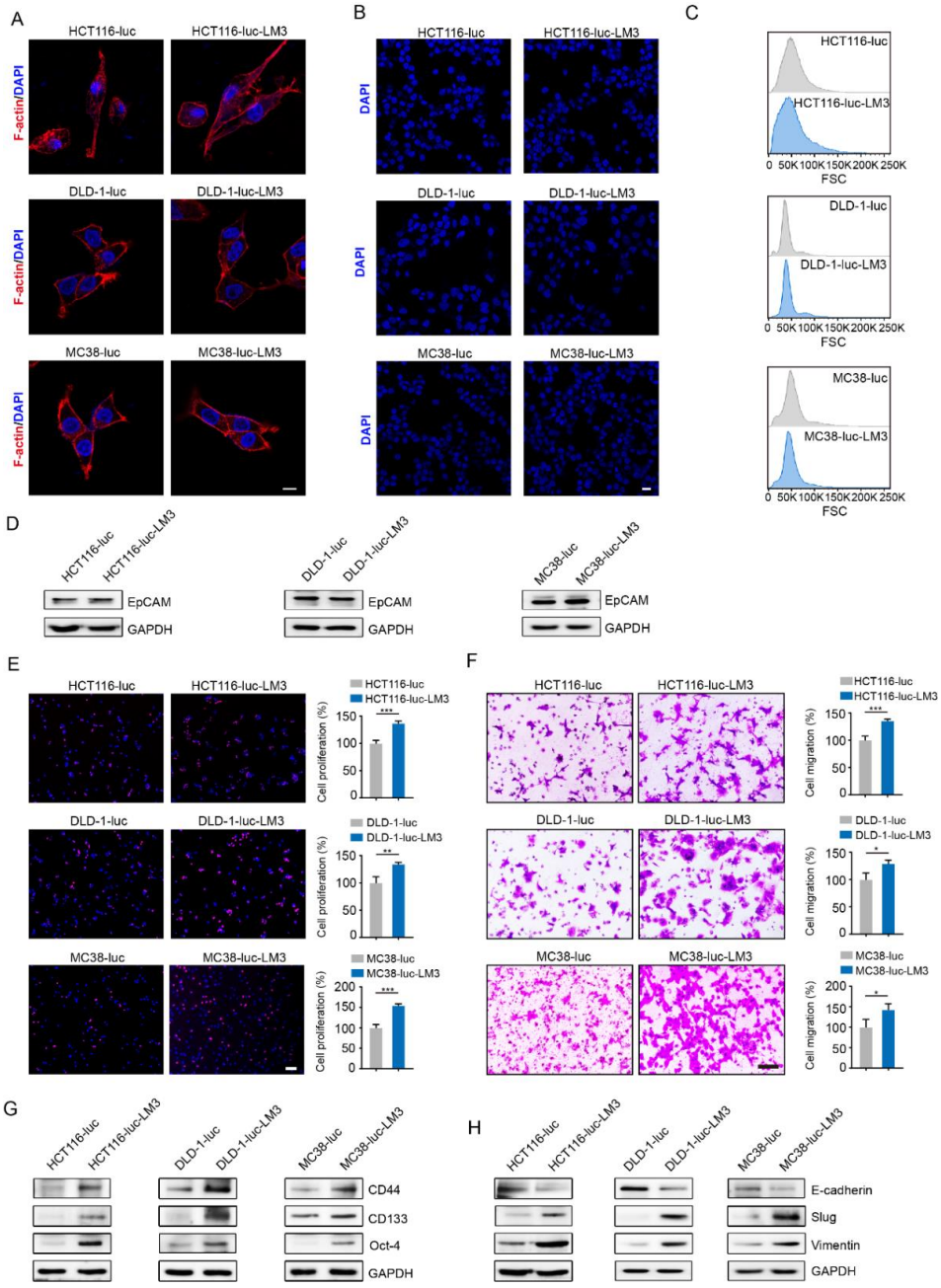

**Supplemental Figure 18. Pseudo-time trajectory for dynamic changes in matrix-**

**pericytes. (A)** Trajectory analysis plot for matrix-pericytes. Cells are ordered in pseudo-time colored in a gradient from dark blue to light blue **(B)** The trajectory of the differentiation state of matrix-pericytes. Matrix-pericytes were originated from Cluster 9 and evolved into Cluster 12.




**Supplemental Figure 19. Effects of metastatic CRC cells on the expressions of integrin α5 and TCF21 in TPCs. (A)** Quantification of TCF21$^+$ TPC ratio in CRC patients with or without *TP53*, *BRAF* and *KRAS* mutation (n = 75). **(B)** Quantification of integrin α5 expression in TPCs derived from CRC patients with or without *TP53*, *BRAF* and *KRAS* mutation (n = 75). **(C)** Western blotting analysis of integrin α5 and

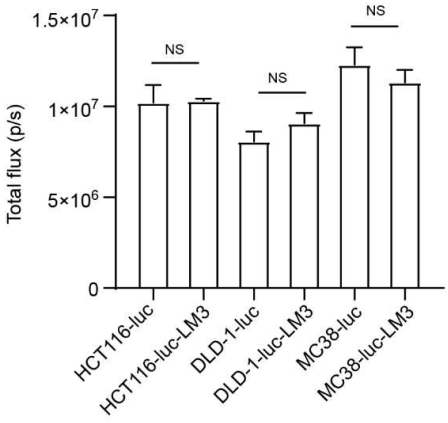
TCF21 in TPCs primed with CM from weakly-metastatic (SW480, Caco-2, HT29) and highly-metastatic (SW620, HCT116, RKO) CRC cells (n = 3). **(D)** Western blotting analysis of integrin α5 and TCF21 in TPCs primed with or without the EVs-free CM of highly-metastatic CRC cells (n = 3). The EVs-free CM of highly-metastatic CRC cells were generated by centrifugation at $1\times10^5$ g to remove the EVs. EVs, extracellular vesicles; CM, conditioned medium. Each sample on the violin plots represents individual patient data. NS, not significant. NS by two-tailed Mann-Whitney test. *WT, wildtype*; *MUT, mutant*.




744

745 **Supplemental Figure 20. Comparison of parental cells with LM3 cells. (A)**
746 Immunofluorescence analysis of cellular morphology (n = 3). Phalloidin-rhodamine
747 was used to identify F-actin. Scale bar, 10 μm. **(B)** Immunofluorescence analysis of
748 nucleus size (n = 3). Cell nucleus was measured after DAPI staining of fixed cells. Scale
749 bar, 20 μm. **(C)** FCM analysis of the cell size of parental cells and LM3 cells by Forward

scatter (n = 3). **(D)** Western blotting analysis of EpCAM in parental cells and LM3 cells (n = 3). **(E)** EdU assay for the proliferation of parental cells and LM3 cells (n = 3). Scale bar, 100 μm. **(F)** Transwell assay for the migration of parental cells and LM3 cells (n = 3). Scale bar, 100 μm. **(G)** Western blotting analysis of CD44, CD133 and Oct-4 in parental cells and LM3 cells (n = 3). **(H)** Western blotting analysis of E-cadherin, slug and Vimentin in parental cells and LM3 cells (n = 3). Data are presented as mean ± SEM. $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$ by two-tailed unpaired *t*-test.




**Supplemental Figure 21. Comparison of luciferase activities between parental cells and LM3 cells.** Bioluminescence detection of parental cells and LM3 cells. Data are presented as mean ± SEM. NS, not significant. NS by two-tailed unpaired *t*-test.

769 **Supplemental Table 1. C**linical characteristics of CRC specimens.

770

| Character | Overall population |
|---|---|
| **Gender, number of patients (%)** | |
| Female | 33 (44.0) |
| Male | 42 (56.0) |
| **Age, median (range)** | 69 (34 – 91) |
| **Site of primary tumor, number of patients (%)** | |
| Right hemicolon | 17 (22.7) |
| Left hemicolon | 58 (77.3) |
| **Histological grade, number of patients (%)** | |
| High/Moderate | 47 (62.7) |
| Low | 28 (37.3) |
| **Size (cm), median (range)** | 4.3 (2 – 8.6) |
| **TNM stage, number of patients (%)** | |
| I-II | 41 (54.7) |
| IV | 34 (45.3) |
| **Liver metastasis (%)** | |
| No | 41 (54.7) |
| Yes | 34 (45.3) |
| ***TP53*** mutation (%) | |
| No | 30 (40) |
| Yes | 45 (60) |
| ***KRAS*** mutation (%) | |
| No | 47 (62.7) |
| Yes | 28 (37.3) |
| ***BRAF*** mutation (%) | |
| No | 71 (94.7) |
| Yes | 4 (5.3) |

771

**Supplemental Table 2.** Correlation analysis between the MATN2$^+$ TPC ratio and the clinicopathologic data.

| Character | MATN2$^+$ TPC ratio (≤30%) | MATN2$^+$ TPC ratio (>30%) | *P* value |
|---|---|---|---|
| **Gender** | | | |
| Female | 22 (66.7%) | 11 (33.3%) | 0.826 |
| Male | 29 (69.0%) | 13 (31.0%) | |
| **Age** | | | |
| < 60 | 15 (75.0%) | 5 (15.0%) | 0.433 |
| ≥60 | 36 (65.5%) | 19 (34.5%) | |
| **Location** | | | |
| Right hemicolon | 13 (76.5%) | 4 (23.5%) | 0.395 |
| Left hemicolon | 38 (65.5%) | 20 (34.5%) | |
| **Differentiation** | | | |
| High/Moderate | 33 (70.2%) | 14 (29.8%) | 0.595 |
| Low | 18 (64.3%) | 10 (35.7%) | |
| **Size** | | | |
| < 5 cm | 35 (79.5%) | 9 (20.5%) | 0.011 |
| ≥5 cm | 16 (51.6%) | 15 (48.4%) | |
| **TNM stage** | | | |
| I-II | 40 (97.6%) | 1 (2.4%) | < 0.001 |
| IV | 11 (32.4%) | 23 (67.6%) | |
| **Liver metastasis** | | | |
| No | 40 (97.6%) | 1 (2.4%) | < 0.001 |
| Yes | 11 (32.4%) | 23 (67.6%) | |
| ***TP53*** mutation | | | |
| No | 23 (76.7%) | 7 (23.3%) | 0.189 |

| | | | |
|---|---|---|---|
| Yes | 28 (62.2%) | 17 (37.8%) | |
| ***KRAS*** mutation | | | |
| No | 35 (74.5%) | 12 (25.5%) | 0.120 |
| Yes | 16 (57.1%) | 12(42.9%) | |
| ***BRAF*** mutation | | | |
| No | 49 (69%) | 22 (31%) | 0.808 |
| Yes | 2 (50%) | 2 (50%) | |

**Supplemental Table 3.** Correlation analysis between the $TCF21^+$ TPC ratio and the clinicopathologic data.

| Character | $TCF21^+$ TPC ratio (≤44%) | $TCF21^+$ TPC ratio (>44%) | *P* value |
|---|---|---|---|
| **Gender** | | | |
| Female | 16 (48.5%) | 17 (51.5%) | 0.340 |
| Male | 25 (59.5%) | 17 (40.5%) | |
| **Age** | | | |
| < 60 | 11 (55.0%) | 9 (45.0%) | 0.972 |
| ≥60 | 30 (54.5%) | 25 (45.5%) | |
| **Location** | | | |
| Right hemicolon | 6 (35.3%) | 11 (64.7%) | 0.068 |
| Left hemicolon | 35 (60.3%) | 23 (39.7%) | |
| **Differentiation** | | | |
| High/Moderate | 29 (61.7%) | 18 (38.3%) | 0.113 |
| Low | 12 (42.9%) | 16 (57.1%) | |
| **Size** | | | |
| < 5 cm | 27 (61.4%) | 17 (38.6%) | 0.165 |
| ≥5 cm | 14 (45.2%) | 17 (54.8%) | |
| **TNM stage** | | | |

| | | | |
|---|---|---|---|
| I-II | 40 (97.6%) | 1 (2.4%) | < 0.001 |
| IV | 1 (2.9%) | 33 (97.1%) | |
| **Liver metastasis** | | | |
| No | 40 (97.6%) | 1 (2.4%) | < 0.001 |
| Yes | 1 (2.9%) | 33 (97.1%) | |
| ***TP53*** mutation | | | |
| No | 18 (60.0%) | 12 (40.0%) | 0.449 |
| Yes | 23 (51.1%) | 22 (48.9%) | |
| ***KRAS*** mutation | | | |
| No | 29 (61.7%) | 18 (38.3%) | 0.113 |
| Yes | 12 (42.9%) | 16 (57.1%) | |
| ***BRAF*** mutation | | | |
| No | 40 (56.3%) | 31 (43.7%) | 0.478 |
| Yes | 1 (25.0%) | 3 (75.0%) | |

**Supplemental Table 4.** Multivariable logistic regression for clinical and demographic factors between CRC patients with or without liver metastasis.

| | β | S.E. | Wald | *P* | OR | 95% CI |
|---|---|---|---|---|---|---|
| $TCF21^+$ TPC ratio（%） | 7.112 | 1.435 | 24.558 | <0.001 | 1226.464 | 73.636-20427.781 |
| *TP53* mutation | 0.074 | 1.536 | 0.002 | 0.962 | 1.077 | 0.053-21.851 |
| *KRAS* mutation | 0.448 | 1.561 | 0.082 | 0.774 | 1.565 | 0.073-33.387 |
| *BRAF* mutation | 0.875 | 3.392 | 0.067 | 0.796 | 2.399 | 0.003-1850.29 |

Abbreviations: S.E., standard error; OR, odds ratio; CI, confidence interval.

**Supplemental Table 5.** Clinical characteristics of non-metastatic CRC specimens

| Characteristics | Case 1 | Case 2 | Case 3 | Case 4 | Case 5 | Case 6 |
|---|---|---|---|---|---|---|
| **Gender/Age (yr)** | Male (58) | Male (78) | Male (59) | Female (53) | Male (73) | Female (75) |
| **Date of diagnosis** | 20190429 | 20191031 | 20200331 | 20210323 | 20210402 | 20210420 |
| **Tumor type** | Colorectal Adenocarcinoma | Colorectal Adenocarcinoma | Colon Adenocarcinoma | Colon Adenocarcinoma | Colon Adenocarcinoma | Colon Adenocarcinoma |
| **Location** | Rectum | Sigmoid colon | Sigmoid colon | Sigmoid colon | Sigmoid colon | transverse colon |
| **Tumor size (maximum diameter)** | 5 cm | 4 cm | 2 cm | 4 cm | 5 cm | 5 cm |
| **Differentiation** | Moderate | Moderate | Moderate | Moderate | Moderate | Moderate |
| **TNM stage** | T2N0M0 | T4aN0M0 | T3N1bM0 | T4aN1bM0 | T3N0M0 | T3N1aM0 |
| **Clinical stage** | I | IIB | IIIB | IIIB | IIA | IIIB |
| **Clinical metastasis** | No metastasis | No metastasis | No metastasis | No metastasis | No metastasis | No metastasis |
| **Treatment status** | Chemotherapy, | No treatment | No treatment | No treatment | No treatment | No treatment |

| | | | | | | |
|---|---|---|---|---|---|---|
| | radiotherapy | before surgery | before surgery | before surgery | before surgery | before surgery |

**Supplemental Table 6.** Clinical characteristics of liver-metastatic CRC specimens

| **Characteristics** | **Case 1** | **Case 2** | **Case 3** | **Case 4** | **Case 5** | **Case 6** |
|---|---|---|---|---|---|---|
| **Gender/Age (yr)** | Female (61) | Male (51) | Male (78) | Male (64) | Female (51) | Female (59) |
| **Date of diagnosis** | 20210324 | 20190516 | 20191010 | 20190103 | 20190626 | 20200320 |
| **Tumor type** | Rectal Adenocarcinoma | Colorectal Adenocarcinoma | Colorectal Adenocarcinoma | Colorectal Adenocarcinoma | Colorectal Adenocarcinoma | Colorectal Adenocarcinoma |
| **Location** | Rectum | Descending colon and ileum | Sigmoid colon | Sigmoid colon | Descending colon and ileum | Descending colon and ileum |
| **Tumor size (maximum diameter)** | 6 cm | 4 cm | 6 cm | 5 cm | 6 cm | 7 cm |
| **Differentiation** | Moderate | Moderate | Low | Low | Moderate | Low |

| | | | | | | |
|---|---|---|---|---|---|---|
| **TNM stage** | T3N2M1a | T4bN1M1a | T3N2bM1a | T3N2M1a | T4bN1M1a | T4bN1aM1a |
| **Clinical stage** | IVa | IVa | IVa | IVa | IVa | IVa |
| **Clinical metastasis** | Liver metastasis | Liver metastasis | Liver metastasis | Liver metastasis | Liver metastasis | Liver metastasis |
| **Treatment status** | Chemotherapy, radiotherapy | No treatment before surgery | No treatment before surgery | No treatment before surgery | No treatment before surgery | No treatment before surgery |

**Supplemental Table 7.** Primers for mouse genotyping

| Mice | Primer name | Primer Sequence | Product size | Gene type |
|---|---|---|---|---|
| 1 | Rosa26-tF1 | CCCAAAGTCGCTCTGAGTTGTTA | Wt=479bp | T002249 *Rosa26-CAG-LSL-Cas9-tdTomato* |
| | Rosa26-tR1 | TCGGGTGAGCATGTCTTTAATCT | | |
| 2 | tdTomato-tF1 | CGGCATGGACGAGCTGTACAAG | KI=317bp | |
| | WPRE-tR2 | TCAGCAAACACAGTGCACACCAC | | |
| 3 | JS04431-Tcf21-5wt-tF1 | GATCCTTCAAATGACTCCAGGCC | WT: 314bp<br>Fl: 418bp | T013083 *Tcf21*-flox |

| | | | | |
|---|---|---|---|---|
| | JS04431-Tcf21-5wt-tR1 | GTTTGCTAACTTGCTGCCACACAC | | |
| 4 | XM003792- Cspg4-TF1 | AAATCTAAGCGCGGGTCTGGC | WT:0bp<br>KI:272bp | T006187 *Cspg4*-CreERT2 |
| | XM003792- Cspg4-TR1 | TGCGAACCTCATCACTCGTTGC | | |
| 5 | XM003792- Cspg4-TF2 | AAATCTAAGCGCGGGTCTGGC | WT:352bp<br>(KI=2386bp) | |
| | XM003792- Cspg4-TR2 | GGACCATGAGTGCAGTCCCCATA | | |

**Supplemental Table 8.** STR profiles of TPCs.

| Marker | Allele 1 | Allele 2 | Allele 3 | Allele 4 |
|---|---|---|---|---|
| D19S433 | 13 | 14 | | |
| D5S818 | 11 | 13 | | |
| D21S11 | 29 | 31.2 | | |
| D18S51 | 13 | 14 | | |
| D6S1043 | 12 | 13 | | |
| AMEL | X | Y | | |
| D3S1358 | 16 | 17 | | |
| D13S317 | 10 | | | |
| D7S820 | 8 | 12 | | |
| D16S539 | 9 | 11 | | |
| CSF1PO | 12 | 14 | | |
| Penta D | 9 | 12 | | |
| D2S441 | 10 | 11 | | |
| vWA | 14 | 15 | | |
| D8S1179 | 10 | 14 | | |
| TPOX | 8 | 10 | | |
| Penta E | 11 | 12 | | |
| TH01 | 6 | 8 | | |
| D12S391 | 19 | 23 | | |
| D2S1338 | 22 | 23 | | |
| FGA | 20 | 23 | | |

Conclusion of cell identification: The results of STR typing showed that there were no multiple alleles at each locus. No cross contamination of human cells was found in the cells.

**Supplemental Table 9.** Clinical characteristics of the resected CRC patients

| Characteristics | Case 1 | Case 2 | Case 3 | Case 4 |
|---|---|---|---|---|
| **Gender/Age (yr)** | Male (58) | Male (78) | Male (51) | Male (78) |
| **Date of diagnosis** | 20190429 | 20191031 | 20190516 | 20191010 |
| **Tumor type** | Colorectal Adenocarcinoma | Colorectal Adenocarcinoma | Colorectal Adenocarcinoma | Colorectal Adenocarcinoma |
| **Location** | Rectum | Sigmoid colon | Descending colon and ileum | Sigmoid colon |
| **Tumor size (maximum diameter)** | 5 cm | 4 cm | 4 cm | 6 cm |
| **Differentiation** | Moderate | Moderate | Moderate | Low |
| **TNM stage** | T2N0M0 | T4aN0M0 | T4bN1M1a | T3N2bM1a |
| **Clinical stage** | I | IIB | Ⅳa | Ⅳa |
| **Clinical metastasis** | No metastasis | No metastasis | Liver metastasis | Liver metastasis |
| **Treatment status** | No treatment before surgery | No treatment before surgery | No treatment before surgery | No treatment before surgery |



Li X, *et al. Gut* 2023; 72:710–721. doi: 10.1136/gutjnl-2022-327913

**Supplemental Table 10.** STR profiles of HCT116-luc cells and HCT116-luc-LM3 cells.

| Locus | HCT116-luc | | HCT116-luc-LM3 | |
|---|---|---|---|---|
| Amelogenin | X | Y | X | Y |
| D5S818 | 10 | 11 | 10 | 11 |
| D13S317 | 10 | 12 | 10 | 12 |
| D7S820 | 11 | 12 | 11 | 12 |
| D16S539 | 11 | | 11 | |
| vWA | 17 | | 17 | |
| TH01 | 8 | 9 | 8 | 9 |
| TPOX | 8 | | 8 | |
| CSF1PO | 7 | 10 | 7 | 10 |
| The number of matched peaks | | | | 15 |
| Percent match between the query and the database profile： | | | | 100% |

Conclusion of cell identification: ①The results of STR typing of the cell DNA of this strain show that there is no multi-allelic phenomenon at each locus. No human cell cross-contamination is found in the cells. ②The submitted HCT116-luc-LM3 cells are 100% match for HCT116-luc cells.

**Supplemental Table 11.** STR profiles of DLD-1-luc cells and DLD-1-luc-LM3 cells.

| Locus | DLD-1-luc | | DLD-1-luc-LM3 | |
|---|---|---|---|---|
| Amelogenin | X | Y | X | Y |
| D5S818 | 13 | | 13 | |
| D13S317 | 8 | 11 | 8 | 11 |
| D7S820 | 10 | 12 | 10 | 12 |

| | | | | | |
|---|---|---|---|---|---|
| D16S539 | 12 | 13 | 12 | 13 | |
| vWA | 18 | 19 | 18 | 19 | |
| TH01 | 7 | 9.3 | 7 | 9.3 | |
| TPOX | 8 | 11 | 8 | 11 | |
| CSF1PO | 12 | | 12 | | |
| The number of matched peaks | | | | | 16 |
| Percent match between DLD-1-luc and DLD-1-luc-LM3： | | | | | 100% |

Conclusion of cell identification: ①The results of STR typing of the cell DNA of this strain show that there is no multi-allelic phenomenon at each locus. No human cell cross-contamination is found in the cells. ②The submitted DLD1-luc-LM3 cells are 100% match for DLD-1-luc cells.

**Supplemental Table 12.** STR profiles of MC38-luc cells and MC38-luc-LM3 cells.

| Locus | MC38-luc | | MC38-luc-LM3 | |
|---|---|---|---|---|
| 18-3 | 16 | | 16 | |
| 4-2 | 19.3 | 20.3 | 19.3 | 20.3 |
| 6-7 | 14 | 15 | 14 | 15 |
| 19-2 | 13 | | 13 | |
| 1-2 | 19 | | 19 | |
| 7-1 | 26.2 | | 26.2 | |
| 8-1 | 16 | | 16 | |
| 1-1 | 16 | | 16 | |
| 3-2 | 13 | 14 | 13 | 14 |
| 2-1 | 16 | | 16 | |
| 15-3 | 22.3 | | 22.3 | |
| 6-4 | 18 | | 18 | |
| 13-1 | 17.1 | | 17.1 | |
| 11-2 | 16 | | 16 | |
| 17-2 | 15 | | 15 | |
| 12-1 | 17 | | 17 | |

| | | | |
|---|---|---|---|
| 5-5 | 17 | 17 | |
| X-1 | 27 | 27 | |
| | The number of matched peaks | | 21 |
| | Percent match between the query and the database profile | | 100% |

Conclusion of cell identification: ①The results of STR typing of the cell DNA of this strain show that there is no multi-allelic phenomenon at each locus. No human cell cross-contamination is found in the cells. ②The cells of MC38-luc and MC-38-luc-LM3 are mouse cells from a single source, and there is no human-derived cell contamination. ③The submitted MC38-luc-LM3 cells are 100% match for MC38-luc cells.

**Supplemental Table 13.** Primers for qPCR.

| Genes | Forward （5’-3’） | Reverse （5’-3’） |
|---|---|---|
| *ACTB* | GTTGCTATCCAGGCTGTGCTATCC | GGTGGCAGTGATGGCATGGAC |
| *MFAP4* | TGAAGGCACAAGGAGTTCTCT | GGGTAGATGAGGTACACGCC |
| *MMP2* | CGACCACAGCCAACTACGATGATG | GTGCCAAGGTCAATGTCAGGAGAG |
| *FBLN1* | TGCTCCATCAACGAGACCTG | AGCACTCCCGATTCTCATGG |
| *COL1A2* | CCGTGGCAGTGATGGAAGTGTG | CCTTGTTACCGCTCTCTCCTTTGG |
| *CHI3L1* | GGCTTCTTCTGAGACTGGTGTTGG | CGCTTTCCTGGTCGTCGTATCC |
| *COL3A1* | TGTACCAGCCAGACCAGGAAGAC | TGTACCAGCCAGACCAGGAAGAC |
| *FBLN2* | GACCGAGGACAGTGAGGAGGAAG | CAGGCAGTGATGTGGACAGGATG |
| *IGFBP5* | GTACCTGCCCAATTGTGACC | AAGTCCCCGTCAACGTACTC |
| *MATN2* | AGAGGTGTGTGGCTGTGGACTAC | GAGCACTGGCAGACGAAGGAATC |
| *ITGA5* | GTCGGGGGCTTCAACTTAGAC | ACAGAGGTAGACAGCACCAC |
| *CILP* | CTTTGAGAACCTCCGGGCAT | TCGATCCCCCTCAATCTGGT |

| | | |
|---|---|---|
| *C11orf96* | TCCAGTTACCAGGCGGTGAT | TGCGTCTTGAAGCGAGACTG |
| *A2M* | GAGGCAGAAGGACAATGGCT | ATAGGCGGAGAGGGTCACTT |
| *SFRP2* | GCCCGACTTCTCCTACAAGC | CTCCTTCATGGTCTCGTGGC |
| *MAF* | CGTCCTCTCCCGAGTTTTTCA | GGCTTCCAAAATGTGGCGTA |
| *PTGDS* | CCATGTGCAAGTCTGTGGTG | CATGGTTCGGGTCTCACACT |
| *TCF21* | TCCTGGCTAACGACAAATACGA | TTTCCCGGCCACCATAAAGG |

**Supplemental Table 14.** Sequences of siRNAs**.**

| **SiRNA** | **Sense （5’-3’）** | **Antisense （5’-3’）** |
|---|---|---|
| siTCF21-1 | GGAUUCGAACAAGGAAUUUTT | AAAUUCCUUGUUCGAAUCCTT |
| siTCF21-2 | GCUAACGACAAAUACGAGATT | UCUCGUAUUUGUCGUUAGCTT |
| siITGB1-1 | GAACAGAUCUGAUGAAUGATT | UCAUUCAUCAGAUCUGUUCTT |
| siITGB1-2 | GUGGUUUCGAUGCCAUCAUTT | AUGAUGGCAUCGAAACCACTT |

| | | |
|---|---|---|
| siITGB1-3 | GAUCAUUGAUGCAUACAAUTT | AUUGUAUGCAUCAAUGAUCTT |
| siITGA2-1 | CCCGAGCACAUCAUUUAUATT | UAUAAAUGAUGUGCUCGGGTT |
| siITGA2-2 | GCUGGUGACAUCAGUUGUATT | UACAACUGAUGUCACCAGCTT |
| siITGA2-3 | GUGGUUGUGUGUGAUGAAUTT | AUUCAUCACACACAACCACTT |
| siMATN2-1 | GCAUCCUAAUCUUUGCCAUTT | AUGGCAAAGAUUAGGAUGCTT |
| siMATN2-2 | GCAGUUUGUCACUGGAAUUTT | AAUUCCAGUGACAAACUGCTT |

**Supplemental table 15.** Antibodies for Western blotting

| Antibody | RRID | Company |
|---|---|---|
| TCF21 | AB_10601215 | SIGMA |
| COL1A2 | AB_10679394 | Abcam |
| N-cadherin | AB_1310479 | Abcam |
| Vimentin | AB_10562134 | Abcam |
| Slug | AB_777968 | Abcam |
| FBLN1 | AB_2553938 | Invitrogen |
| Integrin α5 | AB_2233962 | Cell Signaling Technology |
| GAPDH | AB_10622025 | Cell Signaling Technology |
| p-FAK$^{Tyr397}$ | AB_10891442 | Cell Signaling Technology |
| FAK | AB_2799801 | Cell Signaling Technology |
| p-PI3K p85$^{(Tyr458)}$/p55$^{(Tyr199)}$ | AB_2895293 | Cell Signaling Technology |
| PI3K | AB_2165248 | Cell Signaling Technology |
| p-AKT$^{Ser473}$ | AB_2315049 | Cell Signaling Technology |
| AKT | AB_2225340 | Cell Signaling Technology |
| EpCAM | Cat. GB12274 | Servicebio |
| E-cadherin | AB_2728770 | Cell Signaling Technology |
| CHI3L1 | Cat. AF8379 | Beyotime |
| COL3A1 | Cat. AF6531 | Beyotime |
| Integrin alpha 2 | Cat. AF2332 | Beyotime |
| Integrin β1/CD29 | Cat. AF0207 | Beyotime |
| MMP2 | Cat. GB11130 | Servicebio |
| CD44 | AB_2750879 | Cell Signaling Technology |
| CD133 | AB_2721172 | Cell Signaling Technology |

| | | |
|---|---|---|
| Oct-4 | AB_823583 | Cell Signaling Technology |
| HRP conjugated Goat Anti-Rabbit IgG (H+L) | Cat. GB23303 | Servicebio |
| HRP conjugated Goat Anti-Mouse IgG (H+L) | Cat. GB23301 | Servicebio |
| HRP conjugated Donkey Anti-Goat IgG (H+L) | Cat. GB23404 | Servicebio |

**Supplemental table 16.** Antibody list for immunofluorescence.

| **Antibody** | **RRID** | **Company** |
|---|---|---|
| MATN2 | AB_2811126 | Abcam |
| VWF | AB_298501 | Abcam |
| COL1A2 | AB_10679394 | Abcam |
| Laminin | AB_298179 | Abcam |
| α-SMA | AB_2799045 | Cell Signaling Technology |
| Integrin α5 | AB_2233962 | Cell Signaling Technology |
| TCF21 | AB_10601215 | SIGMA |
| CD31 | AB_2161028 | R&D Systems |
| CD45 | AB_306361 | Abcam |
| NG2 | AB_11213678 | MERK |
| MMP2 | Cat. GB11130 | Servicebio |
| CHI3L1 | Cat. AF8379 | Beyotime |
| COL3A1 | Cat. AF6531 | Beyotime |
| EpCAM | Cat. GB12274 | Servicebio |

| | | |
|---|---|---|
| Alexa Fluor 647-Donkey anti-Sheep IgG (H+L) | AB_2535865 | Invitrogen |
| Alexa Fluor 546-Donkey anti-Rabbit IgG (H+L) | AB_2534016 | Invitrogen |
| Alexa Fluor 488-Donkey anti-Rabbit IgG (H+L) | AB_2535792 | Invitrogen |
| Alexa Fluor 488-Donkey anti-Goat IgG (H+L) | AB_2534102 | Invitrogen |

**Supplemental table 17.** Antibody list for immunohistochemistry.

| **Antibody** | **RRID** | **Company** |
|---|---|---|
| CD31 | AB_2161028 | R&D Systems |
| E-Cadherin | AB_2728770 | CST |
| Vimentin | AB_10562134 | Abcam |
| Ki67 | Cat. GB11030 | Servicebio |
| HRP- Goat Anti-Rabbit IgG (H+L) | Cat. GB23303 | Servicebio |
| HRP -Goat Anti-Mouse IgG (H+L) | Cat. GB23301 | Servicebio |
| HRP-Donkey Anti-Goat IgG (H+L) | Cat. GB23404 | Servicebio |