

1 Supplemental materials for  
2 **Novel TCF21<sup>high</sup> pericyte subpopulation promotes colorectal cancer metastasis**  
3 **by remodelling perivascular matrix**  
4  
5  
6 **Supplemental methods**  
7 **Cell lines and cell culture**  
8 Human CRC cell lines HCT116, DLD-1, RKO, SW480, SW620, Caco-2, HT29 and  
9 human microvascular endothelial cell-1 (HMEC-1) were purchased from the American  
10 Type Culture Collection (Manassas, VA). Mouse colon cancer cell line MC38 (Cat.  
11 BNCC337716) was from BeNa Culture Collection (Beijing, China). HCT116, DLD-1,  
12 RKO, SW480, SW620 and Caco-2 cells were cultured with DMEM. MC38 cells were  
13 maintained in RPMI-1640. DMEM and RPMI-1640 medium were supplemented with  
14 10% FBS (Cat. FSP500, ExCell Bio, Shanghai, China) and 1% penicillin-streptomycin  
15 (PS). HMEC-1 cells were cultured in endothelial cell medium (ECM, Cat. 1001,  
16 Sciencell research laboratories, Corte Del Cedro Carlsbad, CA) supplemented with 5%  
17 FBS, 1% endothelial cell growth supplement (ECGS), and 1% PS. All cell lines were  
18 cultured at 37 °C in incubator with 5% CO<sub>2</sub>. MC38, HCT116 and DLD-1 cells were  
19 infected with lentivirus harboring luciferase (Genechem, Shanghai, China) to generate  
20 the MC38-luc, HCT116-luc and DLD-1-luc cells, which were then selected with  
21 puromycin (2 µg/mL) for 2 days. All cell lines were authenticated to have no cross-  
22 contamination using a STR Multi-amplification Kit (Microreader™21 ID System) and  
23 tested negative for mycoplasma by the TransDetect® PCR Mycoplasma Detection Kit  
24 (Cat. FM311-01 Transgen, Beijing, China).  
25  
26 **Human samples and specimens**

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

31

### 32 **Mice**

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

50

### 51 **Isolation and culture of TPCs**

52 TPCs were isolated from CRC patients through a microdissection combined with  
53 pericyte medium-based approach that developed by our lab. Briefly, fresh surgical  
54 tumor specimens were obtained from CRC patients with or without liver metastasis.  
55 Information of the CRC patients was listed in **Supplemental Table 5 and**  
56 **Supplemental Table 6**. Tumor tissues were kept in serum-free DMEM containing PS  
57 and placed on the ice, and then washed with pre-chilled PBS in a sterile hood to remove  
58 the blood, adipose tissues. Tumor vessels were separated from perivascular adipose  
59 tissues under a stereomicroscope (Olympus, SZX7). The acquired tumor vessels were  
60 cultured in Pericyte Medium (PM, Cat. 1201, Corning Cellgro, Corning, NY, USA,  
61 Sciencell research laboratories) with 5% FBS, 1% PGS and 1% PS at 37 °C with 5%  
62 CO<sub>2</sub>. TPCs were migrated from the tumor vessels within 14 days, which were then  
63 disassociated by trypsin once the confluence reaches 80%. The purity of the isolated  
64 TPCs were authenticated by STR Multi-amplification Kit (**Supplemental Table 8**).

65

#### 66 **Construction of single cell cDNA libraries**

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

74

#### 75 **Bioinformatic analysis of scRNA-seq**

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

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

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

110

#### 111 **Analysis of the public datasets**

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

116

#### 117 **Construction of MC38-luc-LM3, HCT116-luc-LM3 cells and DLD-1-luc-LM3** 118 **cells**

119 To establish highly metastatic MC38-luc-LM3 cells, MC38-luc cells ( $1 \times 10^5$ ) suspended  
120 in 100  $\mu$ L of Matrigel (Cat. 354248, Corning, NY) were injected into the spleen of male  
121 C57BL/6JGpt mice anesthetized with isoflurane inhalation. Liver metastasis was  
122 detected by bioluminescence imaging. MC38-luc-LM1 cells from the metastatic foci  
123 were isolated by mouse tumor dissociation kit (Cat. 130-096-730, Miltenyi Biotec,  
124 Bergisch Gladbach, Germany) and cultured in complete RPMI-1640 medium and

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

137

### 138 **Flow cytometry**

139 Cells were collected, re-suspended in flow cytometry staining buffer, and  
140 distributed into 1.5 mL EP tubes. Following fixation with 4% paraformaldehyde on ice  
141 and permeabilization with 0.1% Triton X-100 in PBS for 5 min, the cells were incubated  
142 with anti-TCF21 antibody (Cat. AB\_182134, Abcam) or anti-MATN2 antibody (Cat.  
143 AF3044, R&D system,) for 1 h on ice. Then, cells were washed with PBS twice and  
144 incubated with Donkey anti-Rabbit IgG (H+L) Highly Cross-Adsorbed Secondary  
145 Antibody, Alexa Fluor 546 (Cat. AB\_2534016, Invitrogen, Carlsbad, CA, USA) or  
146 Donkey anti-Goat IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor 488  
147 (Cat. AB\_2534102, Invitrogen) for 1 h and analyzed by flow cytometry (BD  
148 Biosciences, San Jose, CA). The size of HCT116-luc, DLD-1-luc, MC38-luc, HCT116-  
149 luc-LM3, DLD-1-luc-LM3 and MC38-luc-LM3 cells was directly evaluated by forward

150 scatter using flow cytometry

151

### 152 **Animal studies**

153 MC38-luc-LM3 cells ( $1 \times 10^5$ ) suspended in 100  $\mu$ L of Matrigel were orthotopically  
154 injected into the cecum wall of PC<sup>lin</sup> mice and PC<sup>lin-KO</sup> mice anesthetized with isoflurane  
155 inhalation. At the end of the experiment, tumors were collected and subjected to  
156 immunohistochemistry and immunofluorescence analysis. The livers were harvested,  
157 photographed, and prepared for H&E staining. For co-injection assays, HCT116-luc-  
158 LM3, DLD-1-luc-LM3 cells, TPC<sub>NM</sub> transfected with lentivirus harboring negative  
159 control shNC (TPC<sub>NM</sub><sup>shNC</sup>) or shITGA5 (TPC<sub>NM</sub><sup>shITGA5</sup>), TPC<sub>LM</sub> transfected with Vector  
160 (TPC<sub>LM</sub><sup>Vector</sup>) or lentivirus expressing ITGA5 (TPC<sub>LM</sub><sup>ITGA5</sup>) were collected. HCT116-  
161 luc-LM3 or DLD-1-luc-LM3 cells ( $4 \times 10^5$ ) were premixed with TPCs ( $1.6 \times 10^6$ ) in 100  
162  $\mu$ L of Matrigel, which was then injected into the cecum wall of BALB/C nude mice. At  
163 the end of the experiment, the mice were sacrificed with CO<sub>2</sub> and the metastatic foci in  
164 mouse liver were analyzed by H&E staining. Orthotopic tumor tissues were obtained  
165 for Masson staining, immunohistochemical staining, immunofluorescence, and  
166 transmission electron microscope analysis.

167

### 168 ***In vivo* cell tracking**

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

174

175 **Vessel permeability assay**

176 PC<sup>lin</sup> and PC<sup>lin-KO</sup> mice bearing MC38 orthotopic xenografts were intravenously (i.v.)  
177 injected with 1 mg of FITC-labeled Dextran-40 kDa (Cat. D1845, Thermo Scientific)  
178 for 10 min. Then, the mice were perfused with 4% PFA and tumors were obtained and  
179 then frozen. Tumor tissues were sectioned, and tumor vessels were stained for CD31  
180 (RRID: AB\_2161028, RD, Minneapolis, MN, USA) followed by Donkey anti-Goat  
181 IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor™ 647 (RRID:  
182 AB\_2534102, Invitrogen, Carlsbad, CA, USA), and the double staining of FITC-  
183 labeled Dextran over CD31-positive vessels indicates vessel permeability.

184

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

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

197

198 **Isolation efficiency of CTCs**

199 To determine the isolation efficiency of CTCs<sup>10-12</sup>, 500 MC38-luc-LM3 cells were

200 spiked into 500  $\mu$ L of blood collected from the healthy C57BL/6JGpt mice by cardiac  
201 puncture. The spiked blood was then treated with blood red cell lysing reagent and the  
202 remaining cells were seeded on a 12-well plate and cultured with DMEM containing  
203 20% FBS. The adherent tumor cells were stained with EpCAM and CD45 and the  
204 EpCAM<sup>+</sup>CD45<sup>-</sup> CTCs were identified by confocal microscopy. The number of  
205 EpCAM<sup>+</sup>CD45<sup>-</sup> CTCs was counted within 12 h at a time of no tumor cell growth. The  
206 efficiency of CTC recovery was calculated using the following equation: Cell recovery  
207 (%) = counts of isolated MC38-luc-LM3 cells/500  $\times$  100%.

208

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

210 ChIP assay was performed according to the manufacture manual of SimpleChIP<sup>®</sup>  
211 Enzymatic Chromatin IP Kit (Cat. 9003, Cell Signaling Technology, MA, USA). Briefly,  
212 TCF21-overexpressing TPCs (TPC<sub>NM</sub><sup>TCF21</sup>) were washed twice in cold PBS buffer and  
213 cross-linked with 1% formaldehyde for 10 min at room temperature and then stopped  
214 by addition of glycine (125 mM). Afterwards, samples were lysed, and chromatin were  
215 obtained on ice. Chromatins were then sonicated to get soluble sheared chromatin  
216 (average DNA length of 150-900 bp). Then, 20  $\mu$ L of chromatin was saved as input and  
217 100  $\mu$ L of chromatin was harvested for immunoprecipitation by anti-TCF21 antibody  
218 (RRID: AB\_10601215, Sigma, Shanghai, China), and anti-IgG was served as the  
219 negative control. 10  $\mu$ g of anti-TCF21 was used in the immunoprecipitation reactions  
220 at 4 °C overnight. Then 30  $\mu$ L of protein A beads was added and the samples were  
221 further incubated for 2 h. After reverse cross-linking and DNA purification,  
222 immunoprecipitated DNA was quantified by real-time PCR. Immunoprecipitated DNA  
223 was used to construct sequencing libraries following the protocol provided by the  
224 NEXTflex<sup>®</sup> ChIP-Seq kit (Cat. NOVA-5143-02, BioScientific, TX, USA) and

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

235

#### 236 **RT-qPCR assay**

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

246

#### 247 **Cell infection and transfection**

248 TPCs derived from CRC patients with non-metastasis (TPC<sub>NM</sub>) were infected with  
249 lentivirus harboring *TCF21* overexpression plasmid for 48 h and selected by puromycin

10

250 (2 mg/mL). Detailed information of *TCF21* overexpression plasmid and vector was  
251 listed as follows: TCF21 (NM\_198392) Human Untagged Clone (Cat. RC220002,  
252 Origene. Rockwell, MD, USA), Cloning vector pCMV6-Entry (Cat. PS100001,  
253 Origene). For TCF21- or MATN2-knockdown experiments, TPCs derived from CRC  
254 patients with liver metastasis (TPC<sub>LM</sub>) were transfected with siRNA for 48 h followed  
255 by subsequent analysis. For ITGA2- and ITGB1-knockdown experiments, TPC<sub>NM</sub> were  
256 transfected with siRNA for 48 h followed by subsequent analysis. Transfection was  
257 performed with Lipofectamine™ 3000 (Cat. L3000015, Invitrogen, Carlsbad, CA,  
258 USA) according to the manufacturer instructions, and the siRNA sequences were listed  
259 in **Supplemental Table 14**. For MATN2-overexpression experiments, TPC<sub>NM</sub> were  
260 transfected with pCMV6-MATN2-overexpressing plasmid (Cat. RC203833, Origene)  
261 or pCMV6-Entry as empty vector for 48 h (Cat. PS100001, Origene). For ITGA5  
262 overexpression experiments, TPC<sub>LM</sub> were infected with lentivirus harboring *ITGA5* or  
263 its corresponding Vector (pGC-FU-3FLAG-CBh-gcGFP-IRES-puromycin)  
264 (Genechem, Shanghai, China). For ITGA5-knockdown experiments, TPC<sub>NM</sub> were  
265 infected with lentivirus harboring shITGA5 or pFU-GW-016 as Vector (Genechem,  
266 Shanghai, China).

267

#### 268 **Y15 and SGI1027 treatment**

269 Y15 and SGI1027 were purchased from Selleck (Shanghai, China) and dissolved in  
270 DMSO. Integrin  $\alpha 5$ -overexpressing TPCs ( $2 \times 10^5$ ) were seeded into 6-well plates and  
271 cultured overnight. The next day, cells were treated with Y15 (5  $\mu$ M) or SGI1027 (2.5  
272  $\mu$ M) for 24 h, and then cells were applied for Western blotting assay and bisulfite  
273 sequencing.

274

275 **Western blotting assay**

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

284

285 **Immunofluorescence analysis**

286 Tissue slices were deparaffinized and incubated with 1×Tris-EDTA (pH 9.0) and 0.05%  
287 Tween for 3 min for antigen retrieval. After that, tumor sections were permeabilized in  
288 0.1% Triton™ X-100, blocked with QuickBlock™ immunostaining blocking solution  
289 (Cat. ST797, Beyotime, Shanghai, China) and incubated with the corresponding  
290 primary antibody overnight at 4 °C. Then, the sections were incubated with the  
291 corresponding secondary antibody for 1 h at room temperature. For nucleus staining,  
292 sections were incubated with 1 µg/mL DAPI (Cat. MBD0015, Sigma) for 15 min. The  
293 slides were photographed with a Zeiss LSM 800 confocal microscope and analyzed  
294 with Image J software (RRID: SCR\_003070, Rawak Software Inc., Stuttgart, Germany).  
295 The primary and secondary antibodies were listed in **Supplemental Table 16**. For  
296 phalloidin immunofluorescence assay, HCT116-luc, DLD-1-luc, MC38-luc, HCT116-  
297 luc-LM3, DLD-1-luc-LM3 or MC38-luc-LM3 cells were plated on the glass bottom  
298 cell culture dish and incubated with DMEM for 24 h. The next day, cells were fixed,  
299 permeabilized with 0.1% Triton™ X-100 and then incubated with Alexa Fluor™ 594-

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

303

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

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

319

#### 320 **RNA sequencing analysis**

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

325 total RNA by Dynabeads Oligo (dT) 25-61005 (Thermo Fisher, CA, USA) using two  
326 rounds of purification. Then the poly(A) RNA was cut into pieces using Magnesium  
327 RNA Fragmentation Module (Cat. e6150, NEB, NY, USA) under 94 °C for 5-7 min.  
328 Then the fragmented RNA pieces were reversely transcribed into cDNA by  
329 SuperScript™ II Reverse Transcriptase (Cat.18064022, Invitrogen, USA) and  
330 sequenced with illumina Novaseq™ 6000 (LC-Bio Technology CO., Ltd., Hangzhou,  
331 China). Then, StringTie and edgeR were used to evaluate the expression levels of all  
332 transcripts. The differentially expressed mRNAs and genes were picked with log<sub>2</sub> (fold  
333 change) >1 or log<sub>2</sub> (fold change) <-1 and with statistical significance (*P* value < 0.05)  
334 by R package-edgeR. The volcano plot revealed the distributions of log<sub>2</sub> fold change  
335 and *P* values for the differentially expressed genes. The GO terms  
336 (<http://www.geneontology.org>) of these differentially expressed genes were annotated.

337

### 338 **Migration and invasion assay**

339 Migration assay was performed with 24-well Boyden chambers (Corning, NY, USA)  
340 containing inserts of polycarbonate membranes with 8 μm-pores. Cells suspended with  
341 100 μL of serum-free medium were seeded in the upper compartment (3×10<sup>4</sup> HCT116  
342 cells or 2×10<sup>4</sup> TPCs). The bottom chamber was filled with different chemoattractants.  
343 For invasion assay, the upper chamber was pre-coated with 30 μL of Matrigel (diluted  
344 at 3:1 using PBS) and incubated for 30 min. Then, PKH67-labeled HCT116 cells or  
345 DLD-1 cells (5×10<sup>3</sup>) mixed with TPCs (2.5×10<sup>4</sup>) were seeded into the upper chamber.  
346 The bottom chamber was filled with PM and DMEM (5:1). Following incubation for  
347 48 h, the upper chamber was fixed with 4% paraformaldehyde for 30 min and then the  
348 cells were stained with 0.1% crystal violet. The non-migrated cells on the upper side of  
349 the membrane were removed with a cotton swab. The cells remaining on the lower

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

352

### 353 **Collagen gel contraction assay**

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

363

### 364 **Cell proliferation assay**

365 Cells ( $5 \times 10^3$ ) were cultured overnight in 96-well plates. The next day, cells were treated  
366 with or without the culture supernatant of TPCs (48-h culture medium) and cell  
367 proliferation was determined by BeyoClick™ EdU Cell Proliferation kit (Cat. C0071S,  
368 Beyotime, Shanghai, China) and analyzed with Image Pro Plus 6 software.

369

### 370 **Adhesion assay**

371 TPCs with the overexpression or knockdown of TCF21 were collected, washed, and  
372 stained with PKH67 (Cat. MINI67, Sigma). PKH67-labeled TPCs ( $2 \times 10^4$ ) were seeded  
373 in a 96-well plate for 2 h. Then, the media was removed, and cells were washed with

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

376

#### 377 **Tube formation assay**

378 Tube formation was performed with a 96-well plate. Matrigel was first coated in the  
379 plates at 37 °C for 30 min. Then, HMEC-1 cells ( $2 \times 10^4$ ) supplemented with 100  $\mu$ L  
380 ECM were seeded in the Matrigel-coated plates. After 2-h incubation, ECM was  
381 replaced, and HMEC-1 cells were incubated with the conditioned medium of TPCs.  
382 The capillary tubes were photographed under an inverted light microscopy, and the  
383 number of tubes was analyzed by Image Pro Plus 6 software.

384

#### 385 **Transmission electron microscopy analysis**

386 Tumor tissues were acquired and fixed in 2.5% glutaraldehyde (Cat. PH9003, Maya  
387 Reagent, Zhejiang, China). All samples were post-fixed in 1% osmium tetroxide (Cat.  
388 23311-10, Polysciences, USA), dehydrated in graded concentration of alcohols, and  
389 then embedded in low-viscosity resin. The embedded tissues were sectioned and stained  
390 with saturated uranyl-acetate and Sato's lead-citrate. Sections were imaged using  
391 JEM1200EX transmission electron microscope equipped with BioScan600W digital  
392 camera (JEOL, Tokyo, Japan).

393

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

395 Tumor tissues were acquired and fixed in 4% paraformaldehyde overnight followed by  
396 washing with PBS twice and sectioned (250  $\mu$ m) using Vibration slice (Leica,  
397 VT1000S). Following blocking with 5% BSA solution for 1 h, the sections were  
398 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<sup>13</sup>. *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 ( $1 \times 10^4$ ) 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<sup>14</sup>. 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

424 region of tumor sections and five 15  $\mu\text{m} \times 15 \mu\text{m}$  AFM stiffness map (16  $\times$  16 raster  
425 series) were acquired for each sample. The Young's modulus of the perivascular region  
426 in each section were determined by Hertz model. Tissue samples were assumed to be  
427 incompressible and a Poisson's ratio of 0.5 was used in the calculation of the Young's  
428 modulus.

429

#### 430 **DNA extraction and bisulfite sequencing**

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

437

#### 438 **Organotypic culture system**

439 Organotypic culture system was constructed with 12-well Boyden chambers (Corning,  
440 NY, USA) containing inserts of polycarbonate membranes with 0.4  $\mu\text{m}$ -pores. HMEC-  
441 1 cells ( $2 \times 10^4$ ) supplemented with 100  $\mu\text{L}$  ECM were seeded in the upper chamber and  
442 incubated overnight. The next day, ECM was removed, and TPCs ( $5 \times 10^5$ ) were  
443 embedded in 1.0 mL collagen I and plated in the chamber. The lower chamber was  
444 filled with complete PM and ECM (PM: ECM, 5:1). Following incubation at 37  $^\circ\text{C}$ , 5%  
445  $\text{CO}_2$  for 5 days, the mechanical properties of the matrix, complex modulus ( $G^*$ ), were  
446 determined with a rheometer (Malvern Kinexus pro<sup>+</sup>, USA) according to previous  
447 report<sup>15</sup>. The elastic modulus (E) was determined from  $G^*$  by assuming a poisson's ratio  
448 ( $\nu$ ) of 0.5 with the expression  $E=2G^*(1+\nu)$  to allow comparison to other published

18

449 work. For invasion assay, TCF21-overexpressing TPCs ( $5 \times 10^5$ ) were embedded in a  
450 matrix mixture of 0.25 mL collagen I and 0.75 mL Matrigel and plated in the chamber.  
451 Following incubation for 5 days, PKH67-labeled HCT116 cells ( $3 \times 10^4$ ) supplemented  
452 with 100  $\mu$ L complete DMEM were plated on the top of matrix and further cultured for  
453 1 day. At the end of experiment, the whole matrix was fixed in 4% overnight and  
454 subjected to immunofluorescence analysis. The migrated HCT116 cells were detected  
455 by staining of EpCAM. The invaded cells were observed under a Zeiss LSM 800  
456 confocal microscope and analyzed with Image Pro Plus 6 software.

457

#### 458 **Statistical analysis**

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

470

#### 471 **References**

- 472 1. Percie du Sert N, Hurst V, Ahluwalia A, *et al.* The ARRIVE guidelines 2.0: Updated  
473 guidelines for reporting animal research. *PLoS Biol* 2020; 18: e3000410.
- 474 2. Aibar S, González-Blas CB, Moerman T, *et al.* SCENIC: single-cell regulatory

- 475 network inference and clustering. *Nat Methods* 2017; 14: 1083-6.
- 476 3. Huynh-Thu VA, Irrthum A, Wehenkel L, *et al.* Inferring regulatory networks from  
477 expression data using tree-based methods. *PLoS One* 2010; 5: e12776.
- 478 4. Shannon P, Markiel A, Ozier O, *et al.* Cytoscape: a software environment for  
479 integrated models of biomolecular interaction networks. *Genome Res* 2003; 13: 2498-  
480 504.
- 481 [dataset] [5] Kinchen J, Chen HH, Parikh K, *et al.* Data from: Composition of the  
482 Colonic Mesenchyme and the Nature of its Plasticity in Inflammatory Bowel Disease.  
483 *Gene Expression Omnibus*, Sep 27, 2018. [https://www.ncbi.nlm.nih.gov/geo/query/  
484 acc.cgi?acc= GSE114374](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE114374).
- 485 [dataset] [6] Rao M, Oh K, Moffitt R, *et al.* Single-Cell RNA-Seq Analysis of  
486 Metastasis-Primary Tumor Dissimilarity in a Patient with Gastrointestinal  
487 Neuroendocrine Cancer. *Gene Expression Omnibus*, Feb 01, 2020.  
488 <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi>.
- 489 7. Harney AS, Arwert EN, Entenberg D, *et al.* Real-Time Imaging Reveals Local,  
490 Transient Vascular Permeability, and Tumor Cell Intravasation Stimulated by TIE2<sup>hi</sup>  
491 Macrophage-Derived VEGFA. *Cancer Discov* 2015; 5: 932-43.
- 492 8. Szczerba BM, Castro-Giner F, Vetter M, *et al.* Neutrophils escort circulating tumour  
493 cells to enable cell cycle progression. *Nature* 2019; 566: 553-7.
- 494 9. Gkoutela S, Castro-Giner F, Szczerba BM, *et al.* Circulating Tumor Cell Clustering  
495 Shapes DNA Methylation to Enable Metastasis Seeding. *Cell* 2019; 176: 98-112.e14.
- 496 10. Nagrath S, Sequist LV, Maheswaran S, *et al.* Isolation of rare circulating tumour  
497 cells in cancer patients by microchip technology. *Nature* 2007; 450: 1235-9.
- 498 11. Ozkumur E, Shah AM, Ciciliano JC, *et al.* Inertial focusing for tumor antigen-  
499 dependent and -independent sorting of rare circulating tumor cells. *Sci Transl Med* 2013;  
500 5: 179ra47.
- 501 12. Ramani VC, Lemaire CA, Triboulet M, *et al.* Investigating circulating tumor cells  
502 and distant metastases in patient-derived orthotopic xenograft models of triple-negative  
503 breast cancer. *Breast Cancer Res* 2019; 21: 98.
- 504 13. Kirmse R, Otto H, Ludwig T. Interdependency of cell adhesion, force generation  
505 and extracellular proteolysis in matrix remodeling. *J Cell Sci* 2011; 124: 1857-66.
- 506 14. Laklai H, Miroshnikova YA, Pickup MW, *et al.* Genotype tunes pancreatic ductal  
507 adenocarcinoma tissue tension to induce matricellular fibrosis and tumor progression.

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

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

511

512

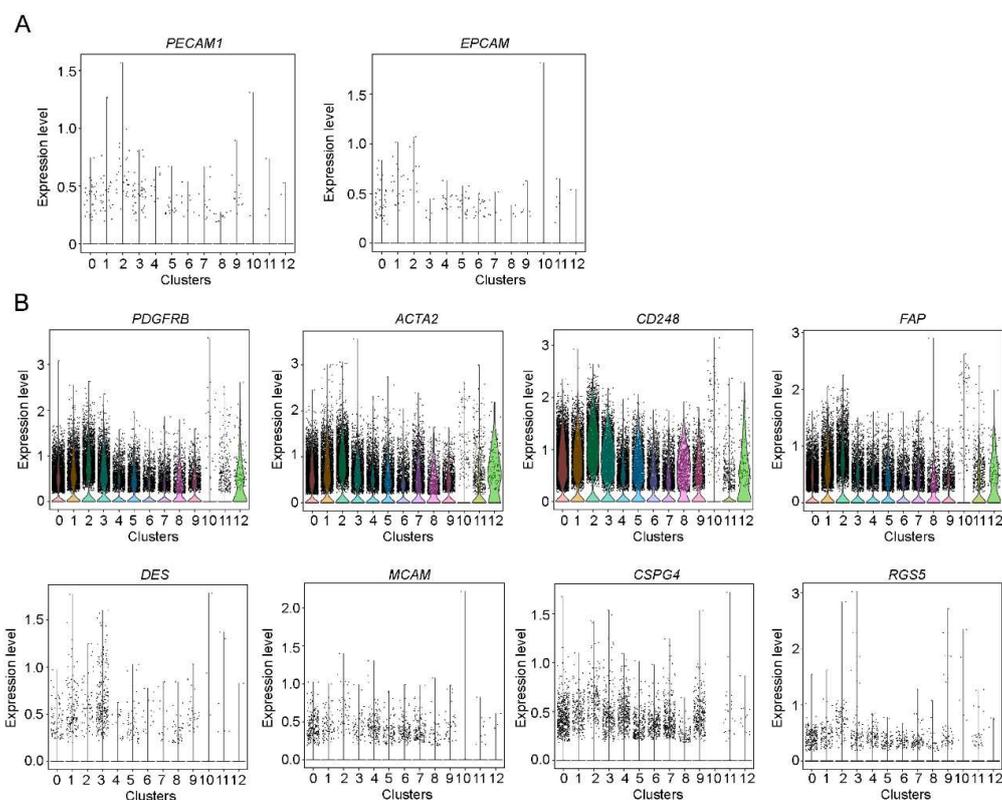
513

514

515

516

517 **Supplemental figures and figure legends**



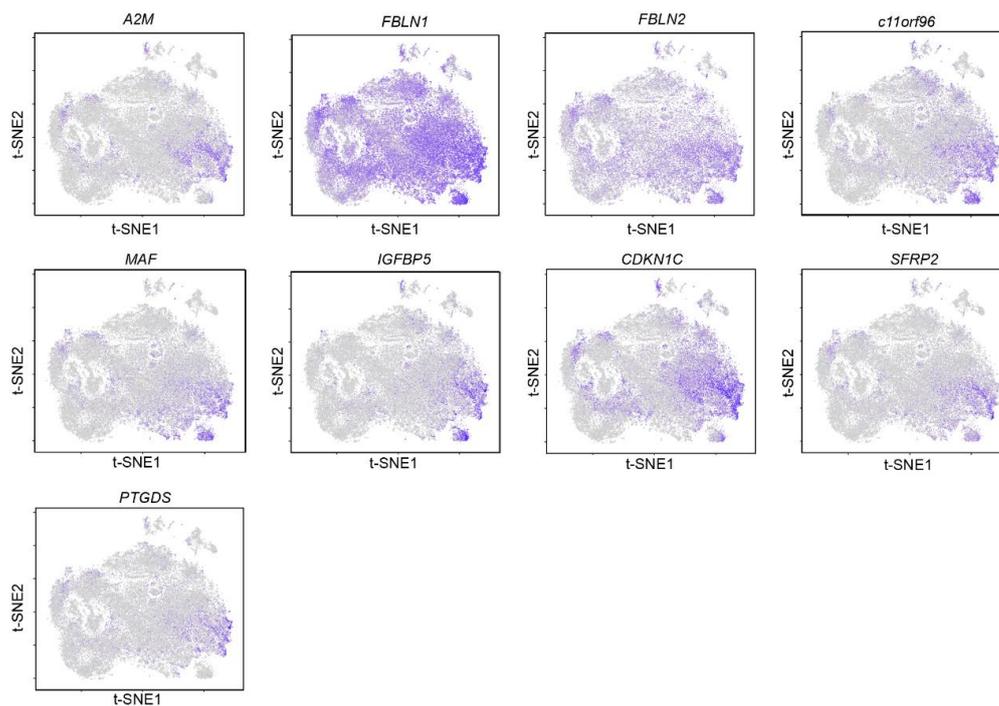
518

519 **Supplemental Figure 1. Transcriptomic characterization of 13 subsets of TPCs. (A)**

520 Gene expression profiles of *PECAM1* and *EPCAM* in distinct subsets of TPCs. **(B)**

521 Gene expression profiles of *PDGFRB*, *ACTA2*, *CD248*, *FAP*, *DES*, *MCAM*, *CSPG4*,

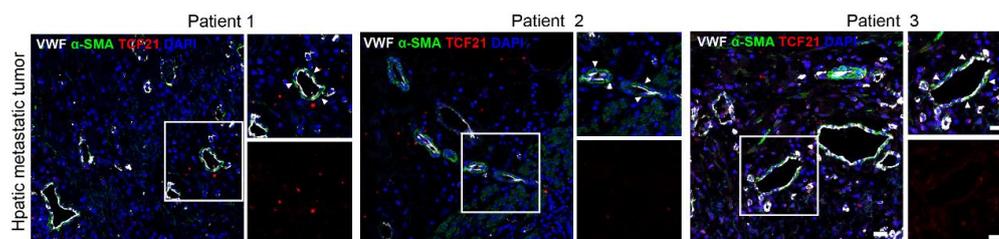
522 and *RGS5* in distinct subsets of TPCs.



523

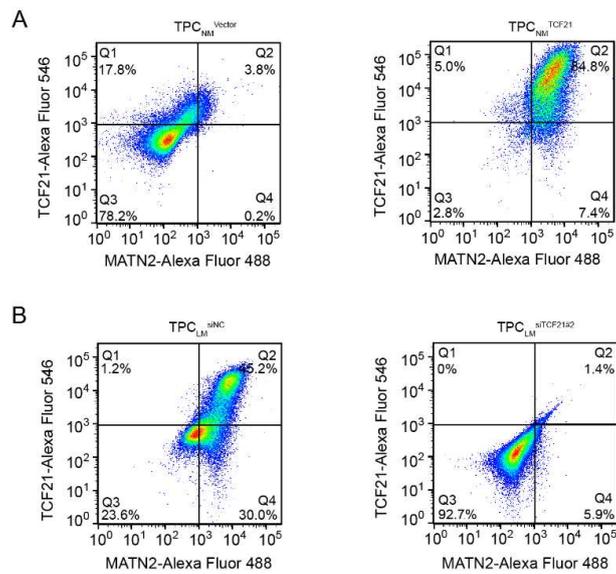
524 **Supplemental Figure 2. Distribution of the matrix-pericyte-related genes in all**525 **subsets of TPCs.** t-SNE visualization of gene distribution of *A2M*, *FBLN1*, *FBLN2*,526 *c11orf96*, *MAF*, *IGFBP5*, *CDKN1C*, *SFRP2* and *PTGDS* in all subsets of TPCs.

527



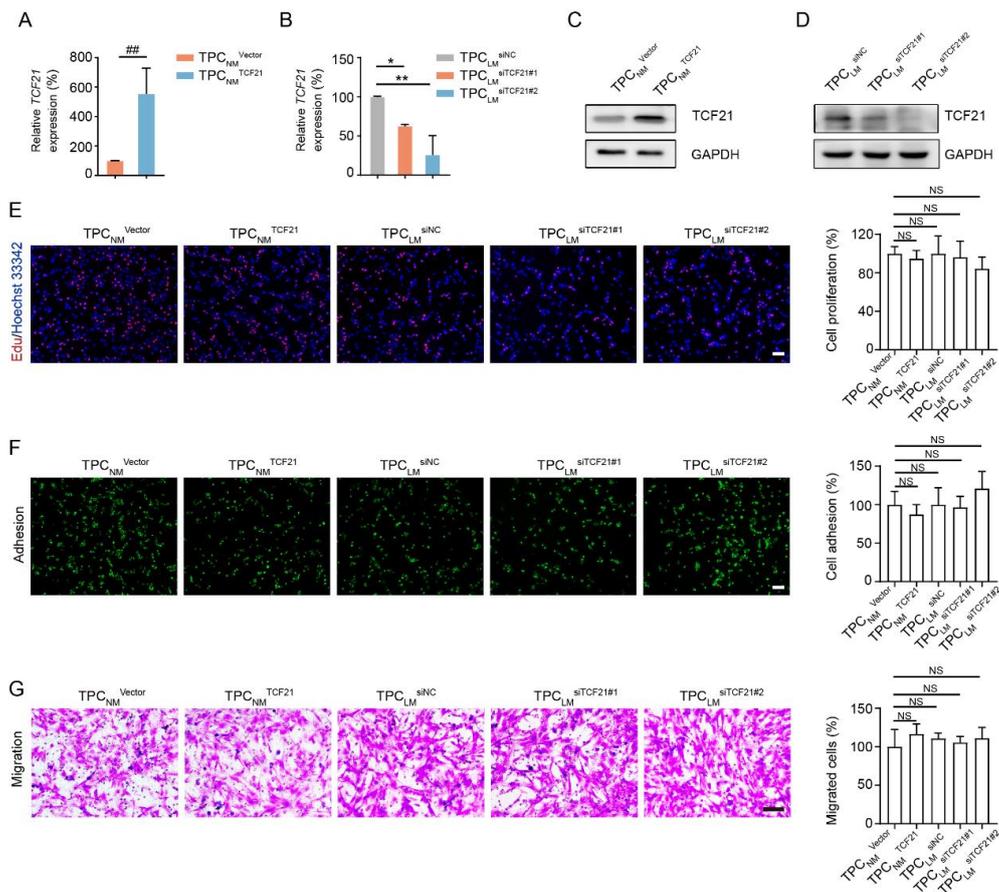
528

529 **Supplemental Figure 3. Determination of TCF21 in TPCs in the hepatic metastatic**530 **tumors from CRC patients.** Representative images of TCF21 staining (red) in TPCs531 ( $\alpha$ SMA<sup>+</sup>, green) in the hepatic metastatic tumors from CRC patients (n = 20). Scale bar,532 20  $\mu$ m.



533

534 **Supplemental Figure 4. The expression of TCF21 is positively correlated with**  
 535 **MATN2 in TPCs. (A)** FCM analysis of the TCF21<sup>+</sup>MATN2<sup>+</sup> TPCs in TPC<sub>NM</sub> infected  
 536 with *TCF21* lentivirus (TPC<sub>NM</sub><sup>TCF21</sup>) or Vector (TPC<sub>NM</sub><sup>Vector</sup>) (n = 3). **(B)** FCM analysis  
 537 of the TCF21<sup>+</sup>MATN2<sup>+</sup> TPCs in TPC<sub>LM</sub> transfected with siRNA targeting TCF21  
 538 (TPC<sub>LM</sub><sup>siTCF21</sup>) or negative control (TPC<sub>LM</sub><sup>siNC</sup>) (n = 3).

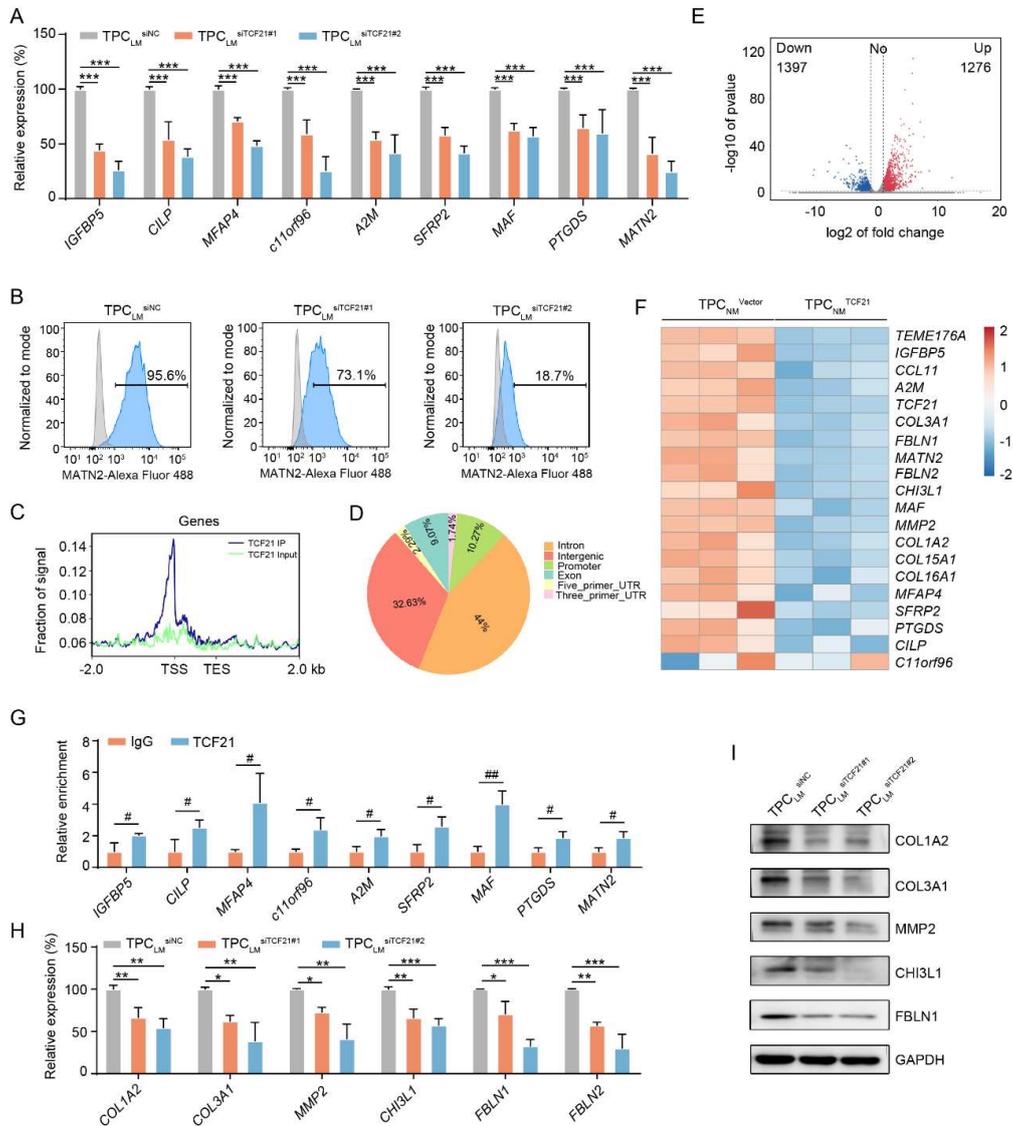


539

540 **Supplemental Figure 5. Effects of TCF21 on the proliferation, adhesion, and**541 **migration of TPCs. (A, B) RT-qPCR analysis of *TCF21* mRNA levels in TCF21-**542 **overexpressing (A) or -knockdown (B) TPCs (n = 3). (C, D) Western blotting analysis**543 **of TCF21 in TCF21-overexpressing (C) or -knockdown (D) TPCs (n = 3). (E)**544 **Representative images and quantification of cell proliferation in TCF21-overexpressing**545 **and -knockdown TPCs (n = 3). Scale bar, 100  $\mu$ m. (F) Representative images and**546 **quantification of TPC adhesion (n = 3). Scale bar, 100  $\mu$ m. (G) Transwell assay for the**547 **migration of TPCs. Quantification of the migrated TPCs is shown (n = 3). Scale bar,**548 **200  $\mu$ m. Data are presented as mean  $\pm$  SEM. NS, not significant. ###*P* < 0.01 by two-**549 **tailed unpaired *t*-test; NS, \**P* < 0.05, \*\**P* < 0.01 by one-way ANOVA followed by**550 **Tukey's post hoc test.**

24

551

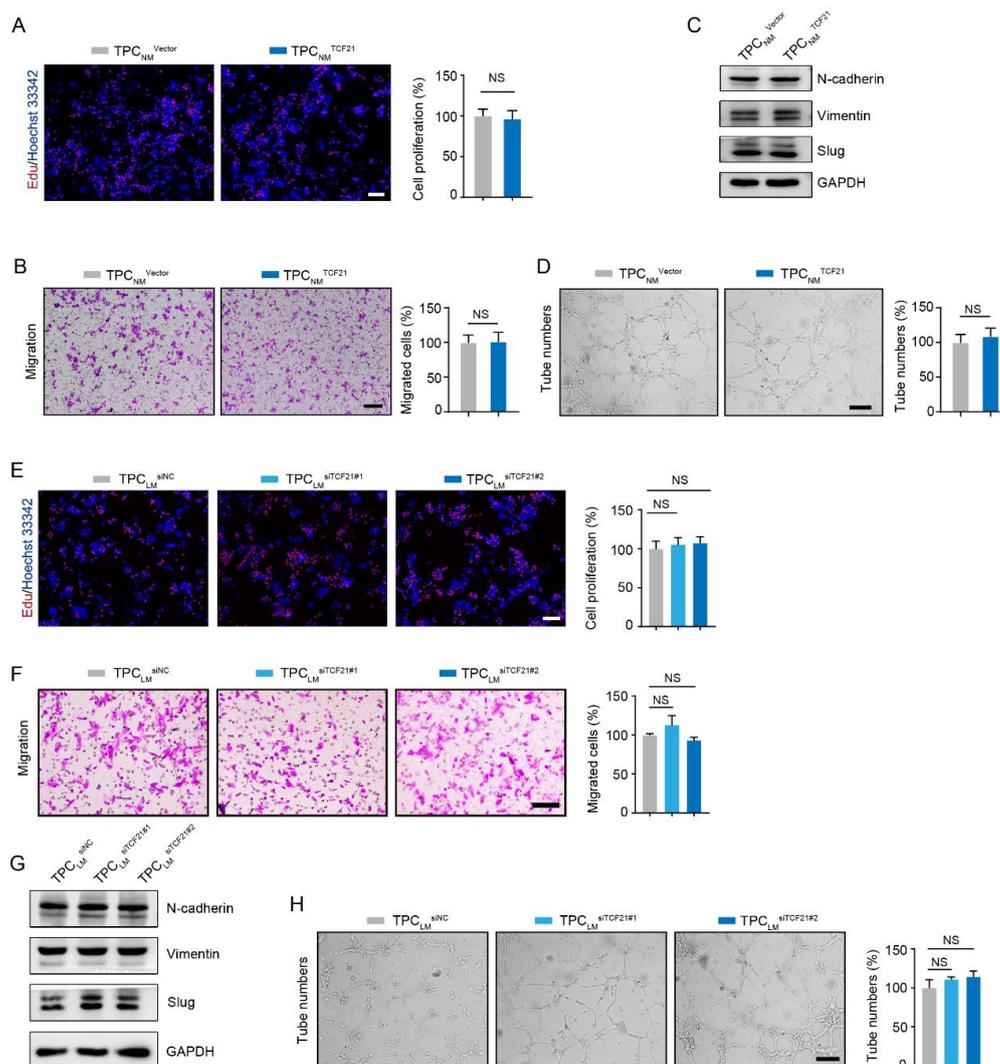


552

553 **Supplemental Figure 6. TCF21 stimulates the generation of matrix-pericytes and**  
 554 **induces ECM remodeling. (A)** RT-qPCR analysis of matrix-pericyte-specific genes in  
 555 TPC<sub>LM</sub> transfected with siRNA targeting TCF21 (TPC<sub>LM</sub><sup>siTCF21</sup>) or negative control  
 556 (TPC<sub>LM</sub><sup>siNC</sup>) (n = 3). **(B)** FCM analysis of MATN2 expression in TCF21-knockdown  
 557 TPCs (n = 3). **(C)** ChIP-seq summary plot of TCF21 enrichment across the indicated  
 558 genomic distance in TCF21-overexpressing TPCs (n = 3). **(D)** The distribution of  
 559 TCF21 peaks on gene elements (n = 3). **(E)** Volcano Plot of TCF21 regulated genes (n

25

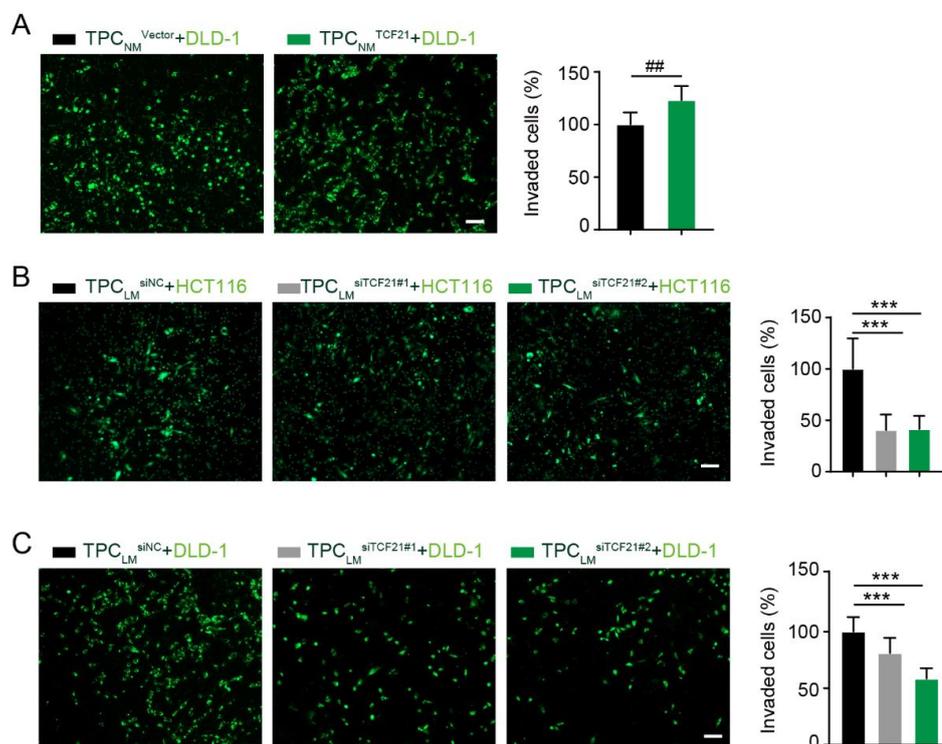
560 = 3); Red dots represent the up-regulated genes and blue dots represent the down-  
 561 regulated genes. **(F)** Heat maps of the differentially expressed genes between  
 562  $TPC_{NM}^{Vector}$  and  $TPC_{NM}^{TCF21}$  ( $n = 3$ ). **(G)** ChIP-qPCR analysis of TCF21 binding at the  
 563 promoter of indicated genes in  $TPC_{NM}^{TCF21}$  ( $n = 3$ ). **(H)** RT-qPCR analysis of  
 564 differentially expressed genes in  $TPC_{LM}^{siNC}$  and  $TPC_{LM}^{siTCF21}$ . **(I)** Western blotting  
 565 analysis of COL1A2, COL3A1, MMP2, CHI3L1, and FBLN1 in  $TPC_{LM}^{siNC}$  and  
 566  $TPC_{LM}^{siTCF21}$  ( $n = 3$ ). Data are presented as mean  $\pm$  SEM, \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P <$   
 567  $0.001$  by one-way ANOVA followed by Tukey's post hoc test, # $P < 0.05$ , ## $P < 0.01$  by  
 568 two-tailed unpaired  $t$ -test.



569

26

570 **Supplemental Figure 7. TCF21 in TPCs has negligible effects on cell migration and**  
571 **angiogenesis. (A)** EdU assay for the proliferation of HCT116 cells primed with  
572 conditioned medium from  $\text{TPC}_{\text{NM}}^{\text{Vector}}$  and  $\text{TPC}_{\text{NM}}^{\text{TCF21}}$  for 48 h (n = 3). Scale bar, 100  
573  $\mu\text{m}$ . **(B)** Transwell assay for cell migration of HCT116 cells. HCT116 cells were seeded  
574 on the upper chamber of the transwell and the bottom compartment was filled with  
575 conditioned medium of  $\text{TPC}_{\text{NM}}^{\text{Vector}}$  and  $\text{TPC}_{\text{NM}}^{\text{TCF21}}$ . After 48 h, the migrated cells were  
576 imaged and counted (n = 3). Scale bar, 100  $\mu\text{m}$ . **(C)** Western blotting analysis of EMT  
577 markers in HCT116 cells primed with conditioned medium of  $\text{TPC}_{\text{NM}}^{\text{Vector}}$  and  
578  $\text{TPC}_{\text{NM}}^{\text{TCF21}}$  (n = 3). **(D)** Representative images and quantification of tube numbers  
579 formed by HMEC-1 cells. HMEC-1 cells suspended with ECM were seeded on the  
580 Matrigel coated 96-well plated. After 2 h, ECM were replaced with the conditioned  
581 medium of  $\text{TPC}_{\text{NM}}^{\text{Vector}}$  and  $\text{TPC}_{\text{NM}}^{\text{TCF21}}$ , and the number of formed tubes was  
582 calculated 2 hours later (n = 3). Scale bar, 100  $\mu\text{m}$ . **(E)** EdU assay for the proliferation  
583 of HCT116 cells primed with conditioned medium of  $\text{TPC}_{\text{LM}}^{\text{siNC}}$  and  $\text{TPC}_{\text{LM}}^{\text{siTCF21}}$  (n =  
584 3). Scale bar, 100  $\mu\text{m}$ . **(F)** Representative images and quantification of migrated  
585 HCT116 cells. HCT116 cells were seeded on the upper chamber of the transwell and  
586 the bottom chamber was filled with the conditioned medium of  $\text{TPC}_{\text{LM}}^{\text{siNC}}$  and  
587  $\text{TPC}_{\text{LM}}^{\text{siTCF21}}$ . After 48 h, the migrated cells were imaged and counted (n = 3). Scale bar,  
588 100  $\mu\text{m}$ . **(G)** Western blotting analysis of EMT markers in HCT116 cells primed with  
589 conditioned medium of  $\text{TPC}_{\text{LM}}^{\text{siNC}}$  and  $\text{TPC}_{\text{LM}}^{\text{siTCF21}}$  (n = 3). **(H)** Tube formation assay  
590 for HMEC-1 cells treated with conditioned medium of  $\text{TPC}_{\text{LM}}^{\text{siNC}}$  and  $\text{TPC}_{\text{LM}}^{\text{siTCF21}}$  as  
591 indicated in **(D)** (n = 3). Scale bar, 100  $\mu\text{m}$ . Data are presented as mean  $\pm$  SEM, NS, not  
592 significant. Two-tailed unpaired *t*-test (**A, B, D**), one-way ANOVA followed by Tukey's  
593 post hoc test (**E, F, H**).



594

595 **Supplemental Figure 8. TCF21 in TPCs promotes invasion of CRC cells. (A)**

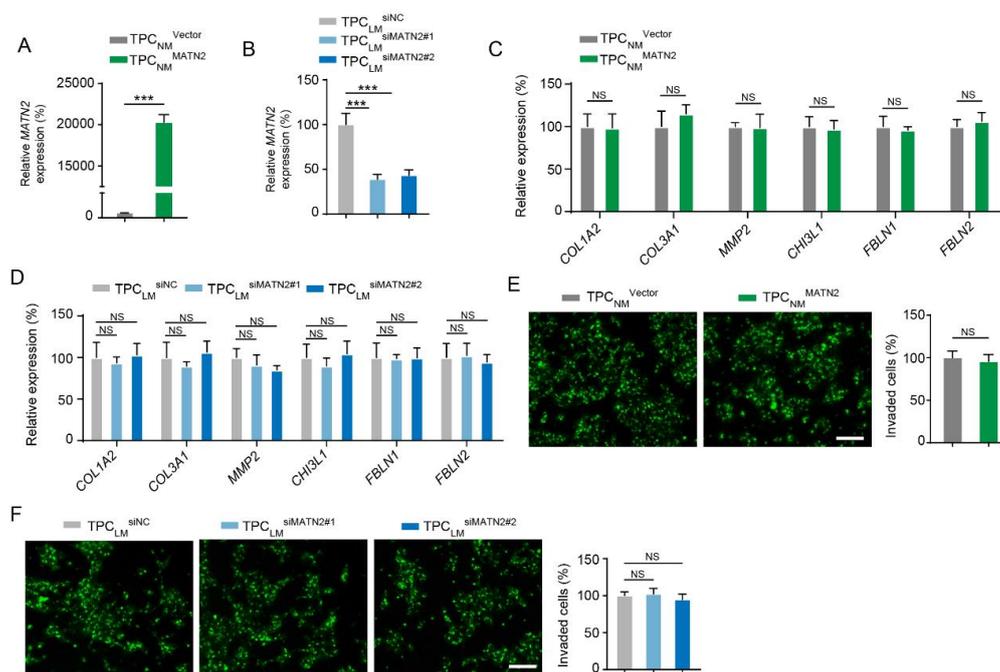
596 Representative images and quantification of invaded DLD-1 cells (green). DLD-1 cells

597 mixed with TPC<sub>NM</sub><sup>Vector</sup> or TPC<sub>NM</sub><sup>TCF21</sup> were seeded into the Matrigel-coated transwell.

598 The invaded DLD-1 cells were photographed and counted after 48 h (n = 3). Scale bar,

599 100 μm. **(B, C)** Transwell assay for invasion of HCT116 cells and DLD-1 cells.600 HCT116 cells **(B)** or DLD-1 cells **(C)** were pre-mixed with TPC<sub>LM</sub><sup>siNC</sup> or TPC<sub>LM</sub><sup>siTCF21</sup>601 and subjected to invasion assay as indicated in **(A)** (n = 3). Scale bar, 100 μm. Data are602 presented as mean ± SEM. ##*P* < 0.01 by two-tailed unpaired *t*-test, \*\*\**P* < 0.001 one-

603 way ANOVA followed by Tukey's post hoc test

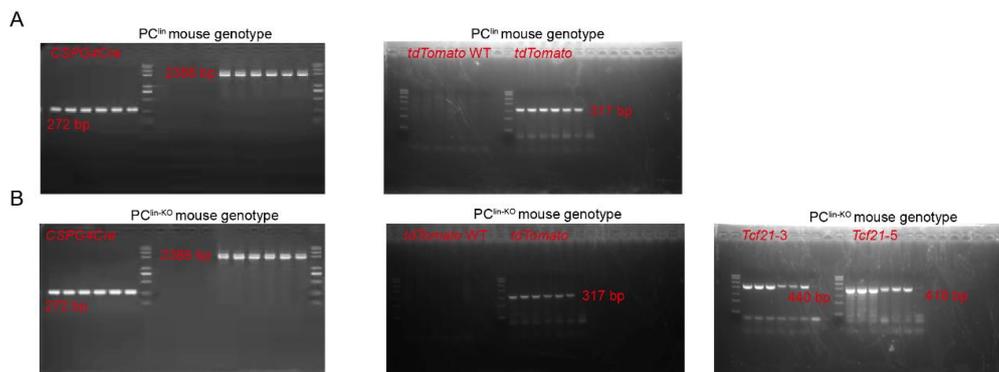


604

605 **Supplemental Figure 9. MATN2 has negligible effects on ECM remodeling and**606 **CRC metastasis. (A, B) RT-qPCR analysis of *MATN2* in MATN2-overexpressing (A)**607 **or -knockdown (B) TPCs (n = 3). (C) RT-qPCR analysis of the ECM-related genes in**608 **TPC<sub>NM</sub><sup>Vector</sup> and TPC<sub>NM</sub><sup>MATN2</sup> (n = 3). (D) RT-qPCR analysis of the indicated genes in**609 **TPC<sub>LM</sub><sup>siNC</sup> and TPC<sub>LM</sub><sup>siMATN2</sup> (n = 3). (E, F) Representative images and quantification**610 **of the invaded DLD-1 cells (green). DLD-1 cells pre-mixed with MATN2-**611 **overexpressing (E) or -knockdown (F) TPCs were seeded into the Matrigel-coated**612 **transwell. The invaded DLD-1 cells were photographed and counted after 48 h (n = 3).**613 **Scale bar, 100  $\mu$ m. Data are presented as mean  $\pm$  SEM, NS, not significant. \*\*\* $P$  < 0.001.**614 **Two-tailed unpaired *t*-test (A, C, E), one-way ANOVA followed by Tukey's post hoc**615 **test (B, D, F).**

616

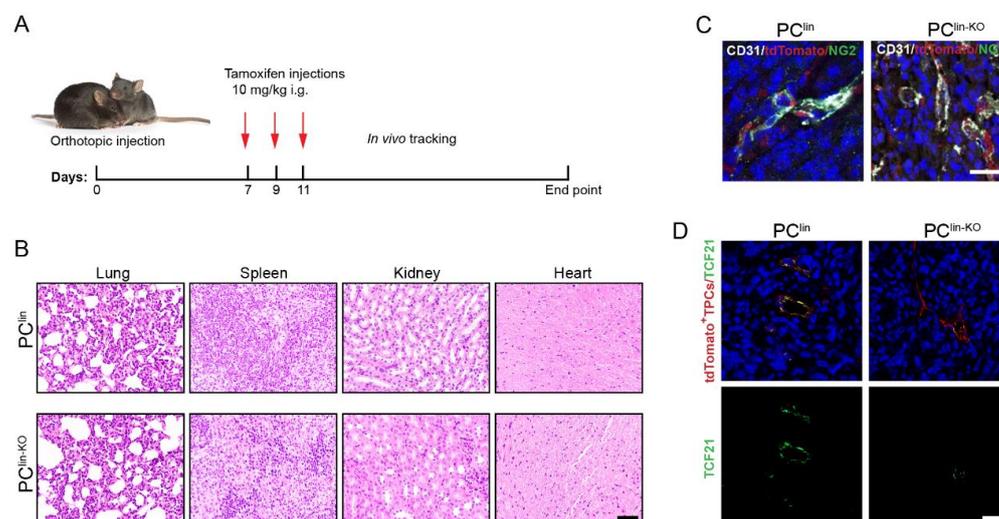
617



618

619 **Supplemental Figure 10. Mouse genotyping. (A)** PCR analysis of the genotype of620  $PC^{lin}$  mouse (n = 6). **(B)** PCR analysis of the genotype of  $PC^{lin-KO}$  mice (n = 6). All mice621 were analyzed by PCR genotyping. PCR analysis of *Cspg4Cre* showed two bands622 identifying homozygous *Cspg4Cre* at 2386 bp (knock in) and 272 bp (wild type); PCR623 analysis of *tdTomato* showed homozygous *tdTomato* with a band at 317 bp (knock in)624 and no signal at 479 bp (wild type); PCR analysis of *Tcf21* indicated homozygous625 *Tcf21<sup>lox/lox</sup>* at 418 bp and 440bp.

626

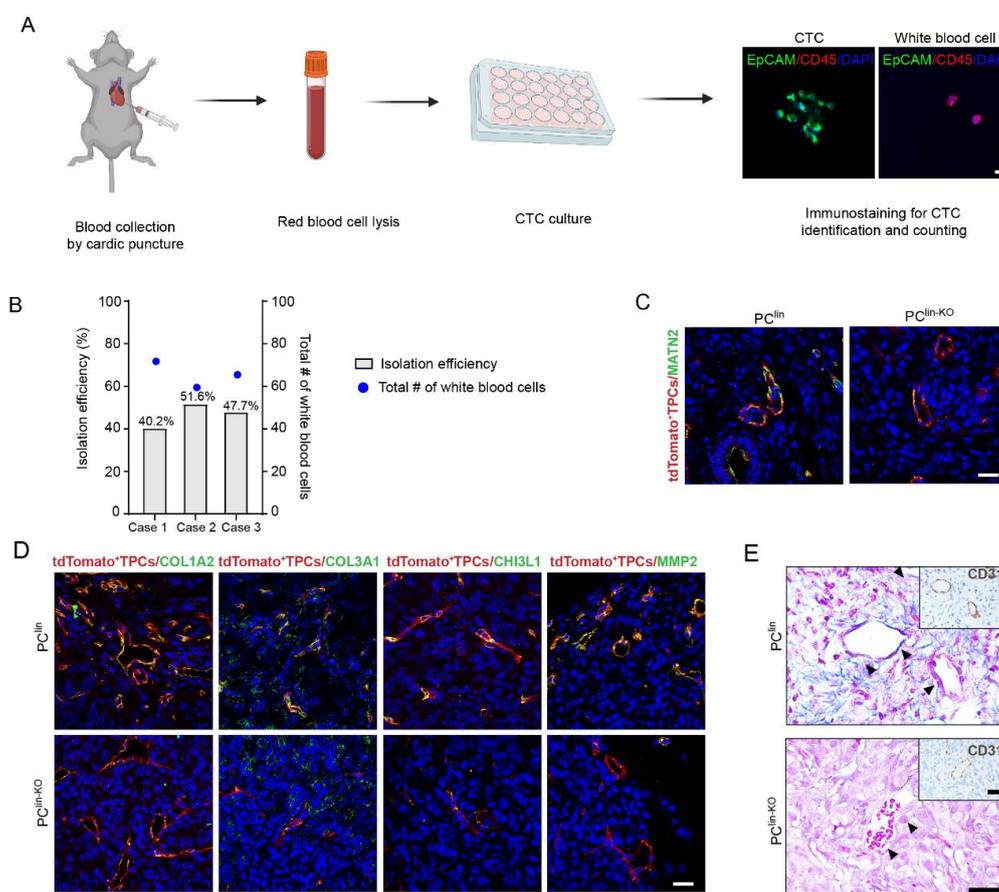


627

628 **Supplemental Figure 11. Characterization of  $PC^{lin}$  and  $PC^{lin-KO}$  mice. (A)**629 Schematic diagram describing the experimental design of *in vivo* experiments.  $PC^{lin}$ 630 mice and  $PC^{lin-KO}$  mice were orthotopically injected with MC38-luc-LM3 cells. After 7

30

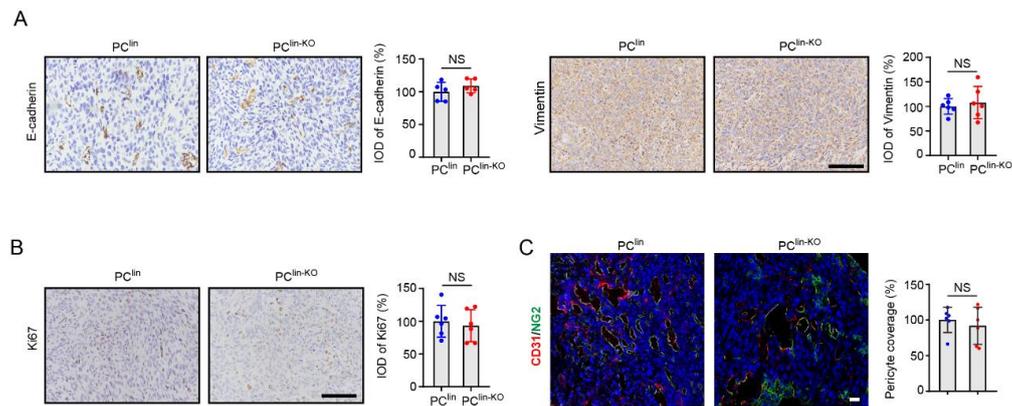
631 days, all mice were treated with tamoxifen (10 mg/kg) through intragastric  
 632 administration every other day for three times. *In vivo* tracking was performed to detect  
 633 tumor liver metastasis. **(B)** Representative images of H&E staining of lung, spleen,  
 634 kidney, heart, and liver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



639

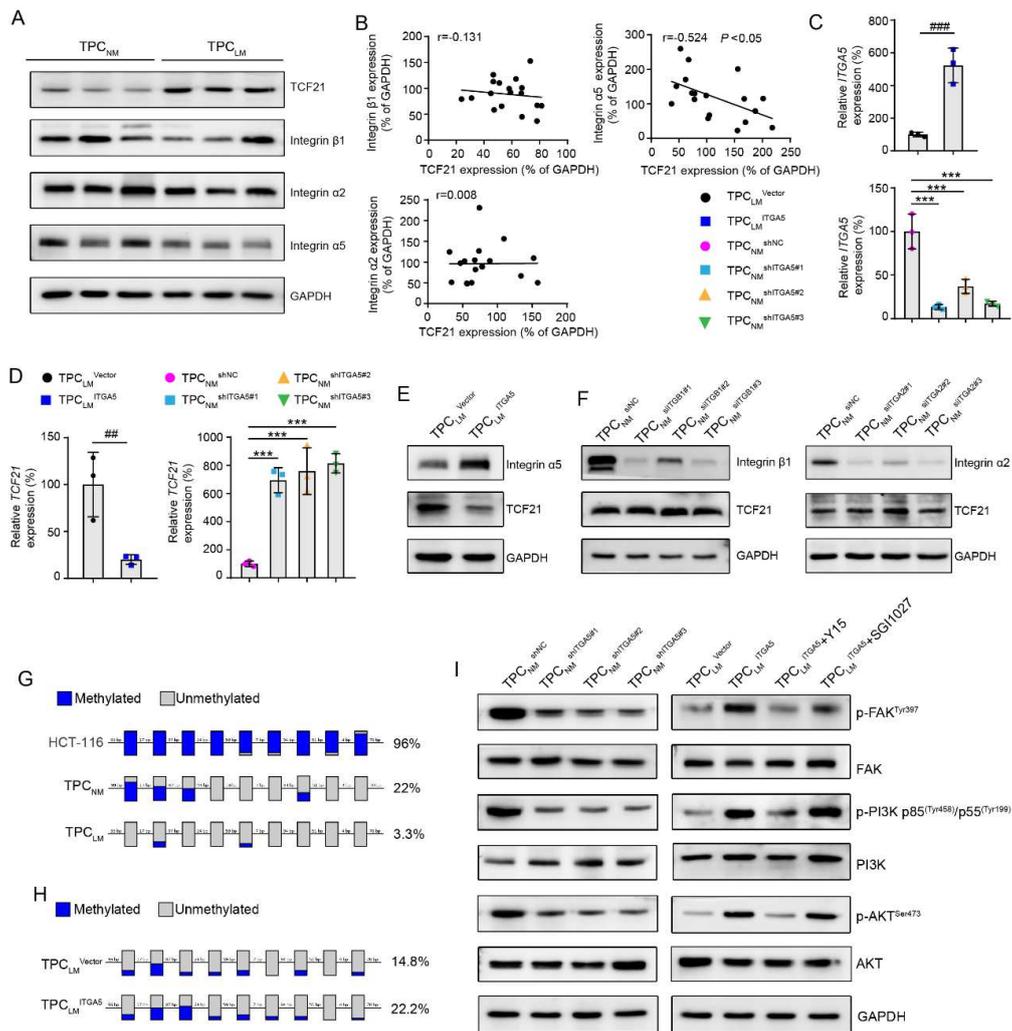
640 **Supplemental Figure 12. Pericyte-specific knockout of *Tcf21* inhibits perivascular**  
 641 **ECM remodeling and CRCLM.** **(A)** Schematic diagram of the isolation and  
 642 identification of CTCs. Scale bar, 20  $\mu$ m. **(B)** Isolation efficiency of CTCs in the blood

643 spiked with MC38-luc-LM3 cells ( $n = 3$ ). **(C)** Immunofluorescence staining for  
 644 MATN2 (green) in TPCs (tdTomato) in primary tumor sections from MC38 allografts  
 645 ( $n = 6$ ). Scale bar, 20  $\mu\text{m}$ . **(D)** Immunofluorescence staining for COL3A1, MMP2,  
 646 COL1A1 and CHI3L1 (green) in TPCs (tdTomato) from primary tumor sections of  
 647 MC38 allografts ( $n = 6$ ). Scale bar, 20  $\mu\text{m}$ . **(E)** Masson staining for perivascular  
 648 collagen in primary tumor sections from MC38 allografts. Tumor vessels were labeled  
 649 with CD31 ( $n = 6$ ). Black arrows indicate the perivascular collagen fibers. Scale bar, 50  
 650  $\mu\text{m}$ .  
 651



652  
 653 **Supplemental Figure 13. Pericyte-specific knockout of *Tcf21* has negligible effects**  
 654 **on EMT and proliferation of CRC cells.** **(A)** Immunohistochemical staining and  
 655 quantification of E-cadherin and Vimentin in orthotopic MC38 tumor tissues ( $n = 6$ ).  
 656 Scale bar, 50  $\mu\text{m}$ . **(B)** Representative images and quantification of Ki67 staining in  
 657 orthotopic MC38 tumor sections ( $n = 6$ ). Scale bar, 50  $\mu\text{m}$ . **(C)** Immunofluorescence  
 658 staining and quantification of pericyte coverage as indicated by CD31 (red) and NG2  
 659 (green) in tumor sections ( $n = 6$ ). Scale bar, 20  $\mu\text{m}$ . Data are presented as mean  $\pm$  SEM.  
 660 NS, not significant. NS by two-tailed unpaired  $t$ -test.

661

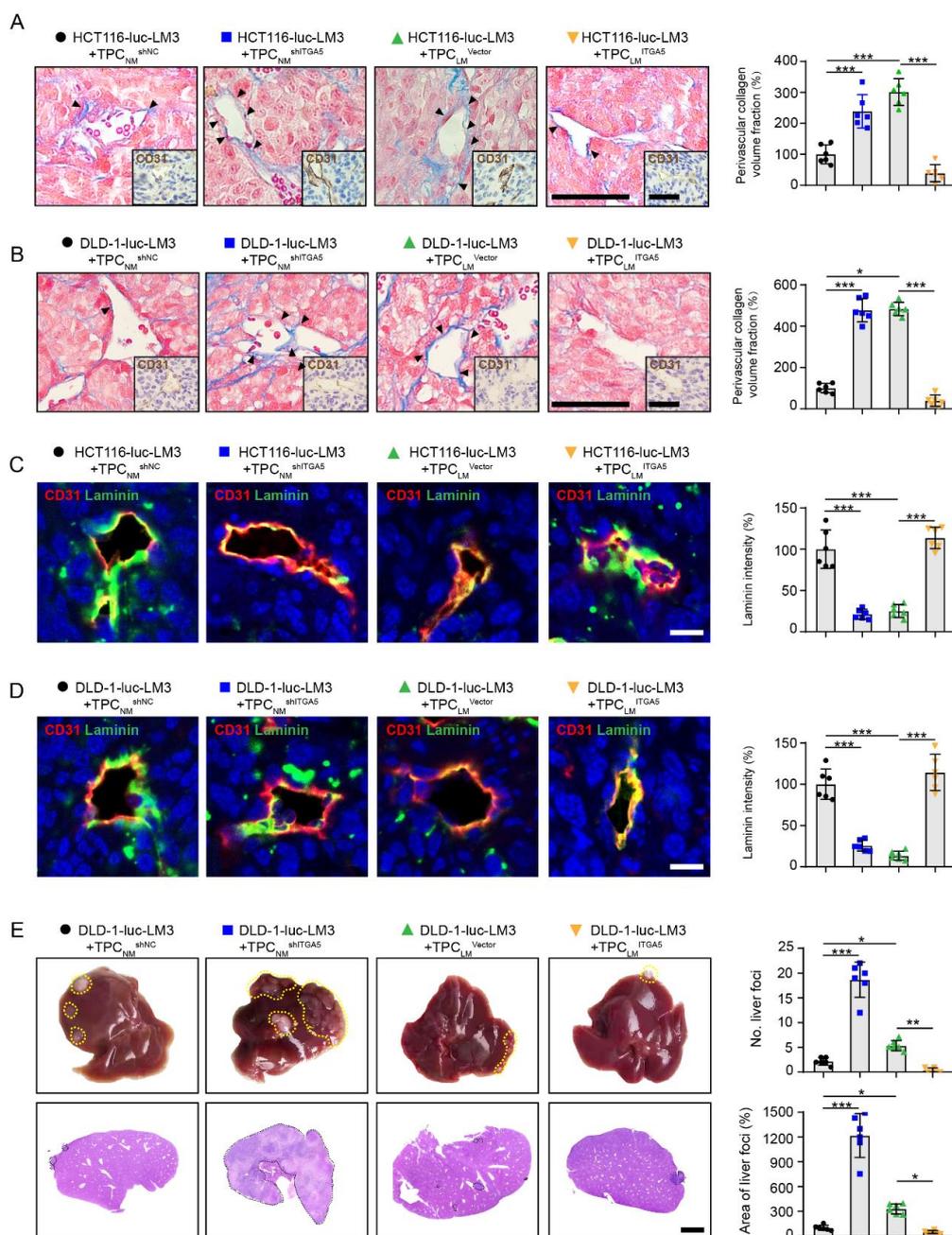


662

663 **Supplemental Figure 14. Effects of integrin  $\alpha$ 2 and integrin  $\beta$ 1 on TCF21 in TPCs.**664 **(A)** Western blotting analysis of TCF21, integrin  $\beta$ 1, integrin  $\alpha$ 2, and integrin  $\alpha$ 5 in665 TPC<sub>NM</sub> and TPC<sub>LM</sub> (n = 3). **(B)** Pearson's correlation analysis of the expression of666 integrin  $\beta$ 1, integrin  $\alpha$ 2 and integrin  $\alpha$ 5 with TCF21 in TPCs (n = 3). **(C)** RT-qPCR667 analysis of the levels of *ITGA5* in the integrin  $\alpha$ 5-knockdown or -overexpressing TPCs668 (n = 3). **(D)** RT-qPCR analysis of the levels of *TCF21* in integrin  $\alpha$ 5-knockdown or -669 overexpressing TPCs (n = 3). **(E)** Western blotting analysis of integrin  $\alpha$ 5 and TCF21670 in integrin  $\alpha$ 5-overexpressing TPC<sub>LM</sub> (n = 3). **(F)** Western blotting analysis of integrin671  $\alpha$ 2, integrin  $\beta$ 1 and TCF21 in integrin  $\beta$ 1-knockdown TPC<sub>NM</sub> (n = 3). **(G)**

33

672 Bisulfite DNA sequencing analysis of *TCF21* promoter region in TPCs or HCT116 cells.  
673 Blue and gray circles represent methylated and unmethylated CpGs, respectively. The  
674 percentage of total methylated CpGs is given on right of each dataset (n = 3). **(H)**  
675 Bisulfite DNA sequencing analysis of *TCF21* promoter region in integrin  $\alpha$ 5-  
676 overexpressing TPC<sub>LM</sub> (n = 3). **(I)** Western blotting analysis of FAK/PI3K/AKT  
677 signaling axis in integrin  $\alpha$ 5-knockdown or -overexpressing TPCs with or without FAK  
678 inhibitor (Y15) or DNMT1 inhibitor (SGI1027) treatment (n = 3). Data are presented  
679 as mean  $\pm$  SEM,  $^{##}P < 0.01$ ,  $^{###}P < 0.001$  by two-tailed unpaired *t*-test,  $^{***}P < 0.001$  by  
680 one-way ANOVA followed by Tukey's post hoc test.



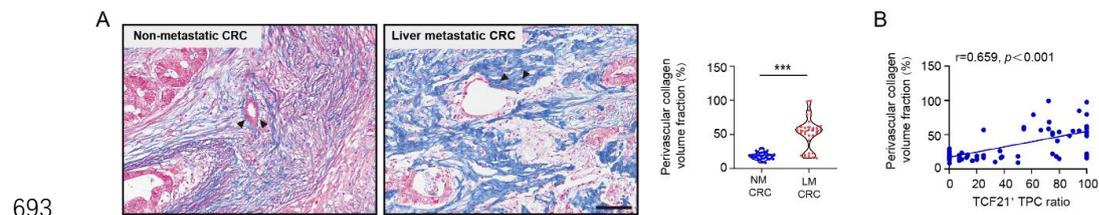
681

682 **Supplemental Figure 15. Loss of integrin  $\alpha 5$  in TPCs promotes perivascular ECM**683 **remodeling and CRCLM. (A, B) Representative images of Masson and CD31 staining**684 **in primary tumor sections (n = 6). Black arrows indicate the perivascular collagen fibers.**685 **Scale bar, 50  $\mu$ m. The quantification of perivascular collagen volume fraction was**686 **shown in the right. (C, D) Immunofluorescence staining and quantification of laminin**

35

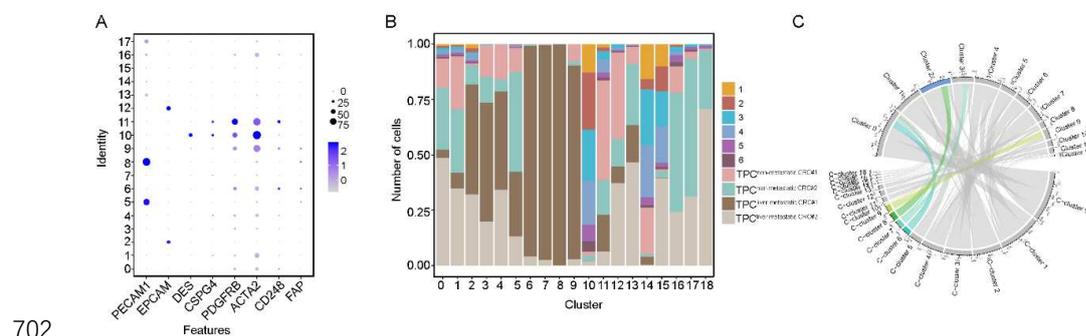
687 (green) around the CD31<sup>+</sup> tumor vessels (red) in HCT116-luc-LM3 xenografts (C) and  
 688 DLD-1-luc-LM3 xenografts (D) (n = 6). Scale bar, 20  $\mu$ m. (E) Representative images  
 689 and H&E staining of liver metastatic foci (n = 6). Yellow and black dotted lines indicate  
 690 the metastatic loci. Scale bar, 2 mm. Data are presented as mean  $\pm$  SEM. \* $P$  < 0.05, \*\* $P$   
 691 < 0.01, \*\*\* $P$  < 0.001 by one-way ANOVA followed by Tukey's post hoc test.

692



694 **Supplemental Figure 16. TCF21 in TPCs is associated with perivascular ECM**  
 695 **deposition.** (A) Masson staining and quantification of perivascular collagen in  
 696 tumors derived from CRC patients with non-metastasis or liver metastasis (n = 75).  
 697 Scale bar, 50  $\mu$ m. \*\*\* $P$  < 0.001 by two-tailed Mann-Whitney test. (B) Pearson's  
 698 correlation analysis of perivascular collagen volume fraction and TCF21<sup>+</sup> TPC ratio  
 699 (n = 75). NM CRC, non-metastatic colorectal cancer, LM CRC, liver metastatic  
 700 colorectal cancer.

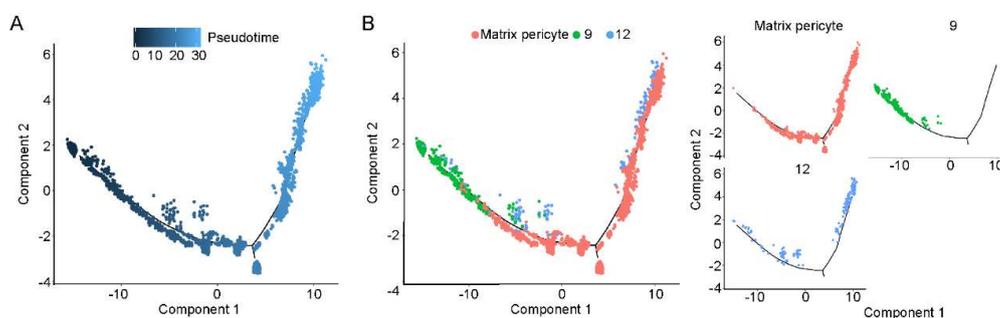
701



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

705 acquired from previous studies. Raw data of pericyte scRNA-seq were collected from  
 706 six samples derived from previous studies (GEO accession GSM3140596,  
 707 GSM3140595, GSM3140594, and GSM3140593) and (GEO accession GSM4159165  
 708 and GSM4159164). Total cells derived from previous studies were classified into 18  
 709 clusters named L-Cluster 0 to L-Cluster 17. Among them, L-Cluster 3, L-Cluster 9, L-  
 710 Cluster 10, and L-Cluster 11 were subjected for further analysis as these four  
 711 populations were positive for *DES*, *CSPG4*, *PDGFRB*, *ACTA2*, *CD248* and *FAP*  
 712 (pericyte markers), but negative for *PECAMI* (endothelial cell marker) and *EPCAM*  
 713 (epithelial cell marker). **(B)** Analysis of the data of the four populations (L-Cluster 3,  
 714 L-Cluster 9, L-Cluster 10, and L-Cluster 11) derived from six samples (Sample 1-6) in  
 715 **(A)** and our scRNA-seq data (Cluster 0-12). The combined pericytes were categorized  
 716 into 19 subpopulations, termed C-Cluster 0 to C-Cluster 18. Among them, C-cluster 0-  
 717 5, C-cluster 10-18 were presented both in our data and the extended data (Sample1-6);  
 718 however, C-cluster 6-9 were specifically revealed in our data, indicating that the  
 719 existing pericyte clusters originated from the previous research<sup>5,6</sup> were included in our  
 720 scRNA-seq data and we discovered four new subsets. **(C)** Comparative analysis of  
 721 pericyte cluster derived from **(B)** (C-cluster 0-18) with our data (Cluster 0-12). Among  
 722 them, C-Cluster 8 was included in Cluster 2 (matrix pericytes).

723

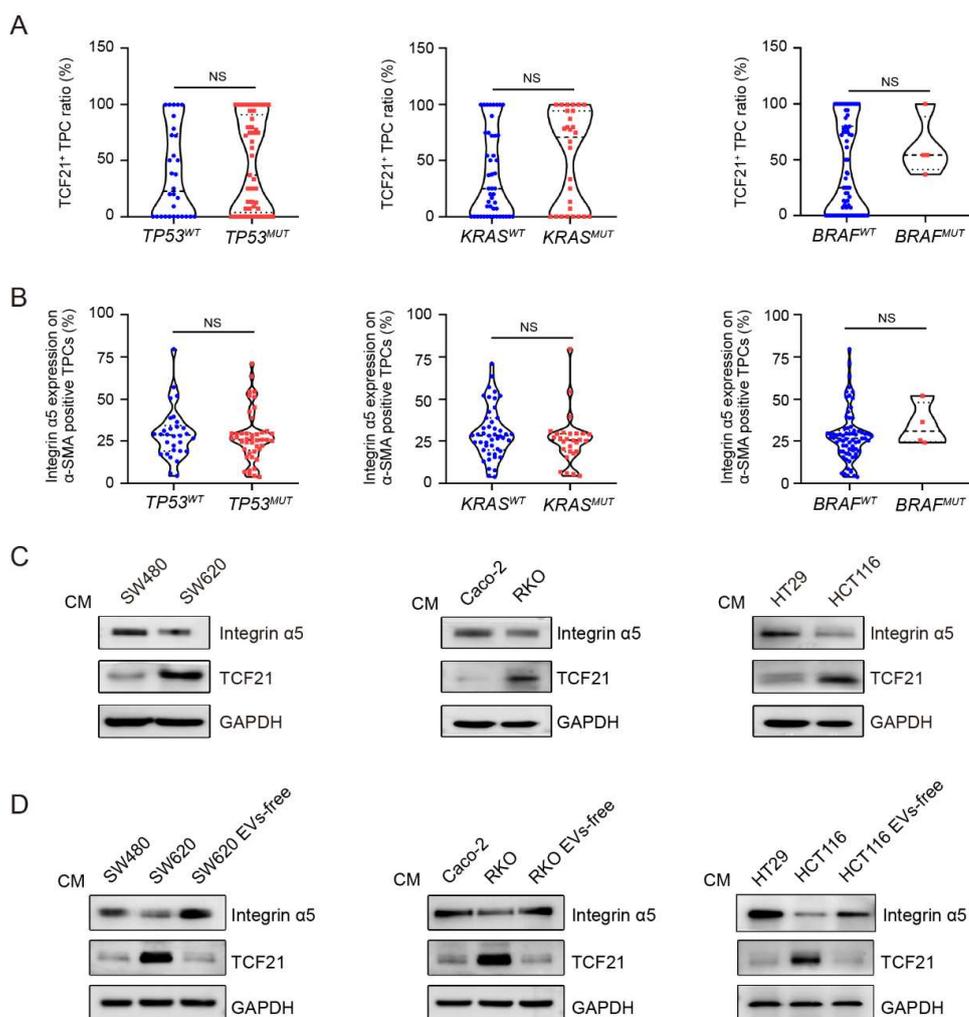


724

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

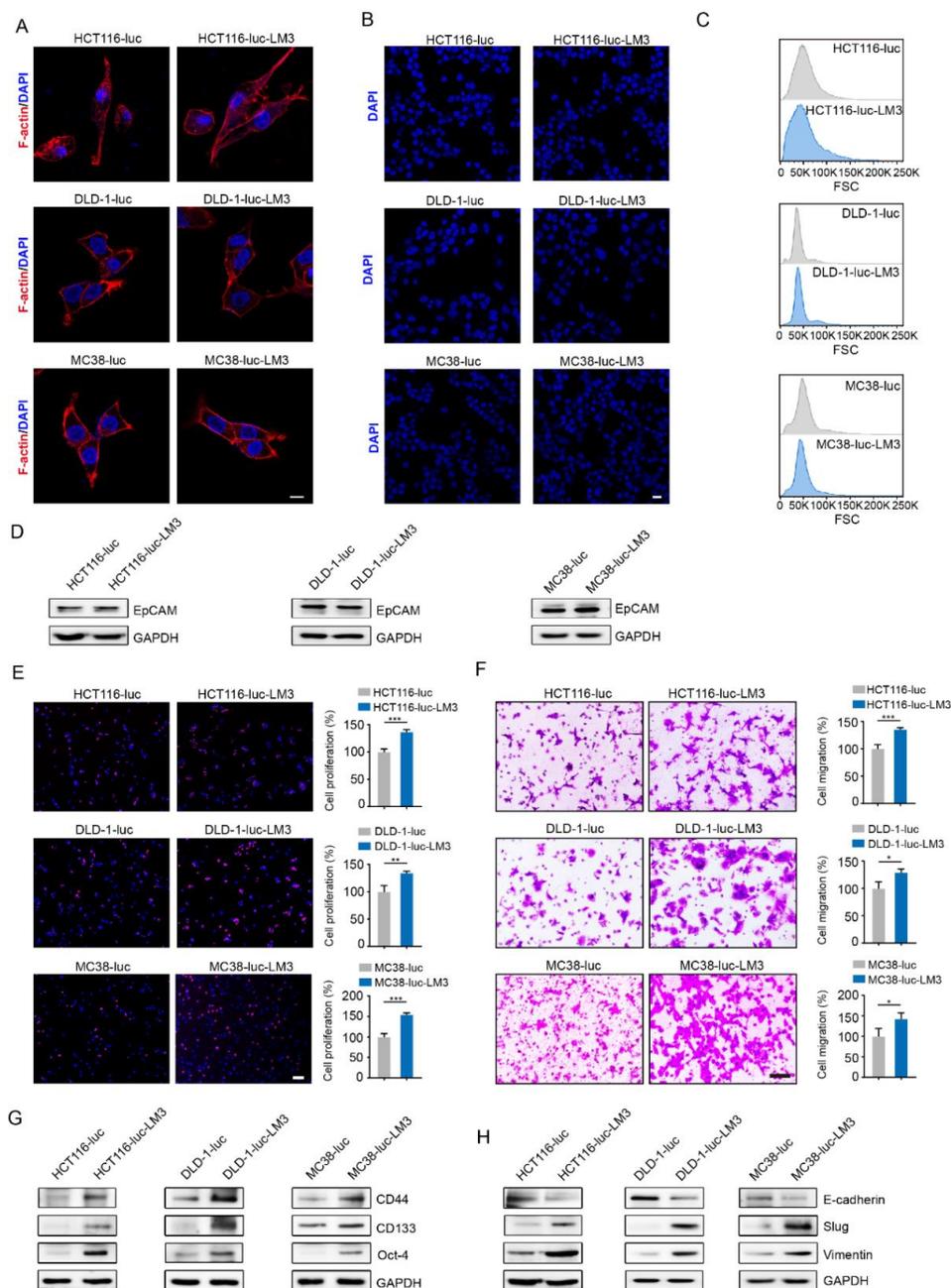
37

726 **pericytes. (A)** Trajectory analysis plot for matrix-pericytes. Cells are ordered in  
 727 pseudo-time colored in a gradient from dark blue to light blue **(B)** The trajectory of the  
 728 differentiation state of matrix-pericytes. Matrix-pericytes were originated from Cluster  
 729 9 and evolved into Cluster 12.



730  
 731 **Supplemental Figure 19. Effects of metastatic CRC cells on the expressions of**  
 732 **integrin α5 and TCF21 in TPCs. (A)** Quantification of TCF21<sup>+</sup> TPC ratio in CRC  
 733 patients with or without *TP53*, *BRAF* and *KRAS* mutation (n = 75). **(B)** Quantification  
 734 of integrin α5 expression in TPCs derived from CRC patients with or without *TP53*,  
 735 *BRAF* and *KRAS* mutation (n = 75). **(C)** Western blotting analysis of integrin α5 and

736 TCF21 in TPCs primed with CM from weakly-metastatic (SW480, Caco-2, HT29) and  
737 highly-metastatic (SW620, HCT116, RKO) CRC cells (n = 3). **(D)** Western blotting  
738 analysis of integrin  $\alpha 5$  and TCF21 in TPCs primed with or without the EVs-free CM of  
739 highly-metastatic CRC cells (n = 3). The EVs-free CM of highly-metastatic CRC cells  
740 were generated by centrifugation at  $1 \times 10^5$  g to remove the EVs. EVs, extracellular  
741 vesicles; CM, conditioned medium. Each sample on the violin plots represents  
742 individual patient data. NS, not significant. NS by two-tailed Mann-Whitney test. *WT*,  
743 *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  $\mu$ m. **(B)** Immunofluorescence analysis of

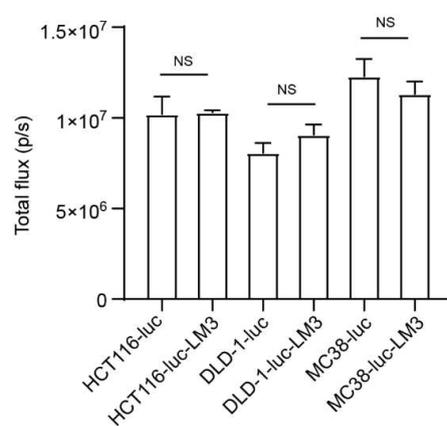
748 nucleus size (n = 3). Cell nucleus was measured after DAPI staining of fixed cells. Scale

749 bar, 20  $\mu$ m. **(C)** FCM analysis of the cell size of parental cells and LM3 cells by Forward

40

750 scatter (n = 3). **(D)** Western blotting analysis of EpCAM in parental cells and LM3 cells  
 751 (n = 3). **(E)** EdU assay for the proliferation of parental cells and LM3 cells (n = 3).  
 752 Scale bar, 100  $\mu$ m. **(F)** Transwell assay for the migration of parental cells and LM3 cells  
 753 (n = 3). Scale bar, 100  $\mu$ m. **(G)** Western blotting analysis of CD44, CD133 and Oct-4  
 754 in parental cells and LM3 cells (n = 3). **(H)** Western blotting analysis of E-cadherin,  
 755 slug and Vimentin in parental cells and LM3 cells (n = 3). Data are presented as mean  
 756  $\pm$  SEM. \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$  by two-tailed unpaired  $t$ -test.

757



758

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

762

763

764

765

766

767

768

769 **Supplemental Table 1.** Clinical characteristics of CRC specimens.

770

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

771

**Supplemental Table 2.** Correlation analysis between the MATN2<sup>+</sup> TPC ratio and the clinicopathologic data.

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

Yes	28 (62.2%)	17 (37.8%)	
<b>KRAS</b> mutation			
No	35 (74.5%)	12 (25.5%)	0.120
Yes	16 (57.1%)	12(42.9%)	
<b>BRAF</b> mutation			
No	49 (69%)	22 (31%)	0.808
Yes	2 (50%)	2 (50%)	

**Supplemental Table 3.** Correlation analysis between the TCF21<sup>+</sup> TPC ratio and the clinicopathologic data.

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

I-II	40 (97.6%)	1 (2.4%)	< 0.001
IV	1 (2.9%)	33 (97.1%)	
<b>Liver metastasis</b>			
No	40 (97.6%)	1 (2.4%)	< 0.001
Yes	1 (2.9%)	33 (97.1%)	
<b>TP53 mutation</b>			
No	18 (60.0%)	12 (40.0%)	0.449
Yes	23 (51.1%)	22 (48.9%)	
<b>KRAS mutation</b>			
No	29 (61.7%)	18 (38.3%)	0.113
Yes	12 (42.9%)	16 (57.1%)	
<b>BRAF mutation</b>			
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.

	$\beta$	S.E.	Wald	<i>P</i>	OR	95% CI
TCF21 <sup>+</sup> 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

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

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

<b>TNM stage</b>	T3N2M1a	T4bN1M1a	T3N2bM1a	T3N2M1a	T4bN1M1a	T4bN1aM1a
<b>Clinical stage</b>	IVa	IVa	IVa	IVa	IVa	IVa
<b>Clinical metastasis</b>	Liver metastasis	Liver metastasis	Liver metastasis	Liver metastasis	Liver metastasis	Liver metastasis
<b>Treatment status</b>	Chemotherapy, radiotherapy	No treatment before surgery				

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

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

---

	JS04431-Tcf21-5wt-tR1	GTTTGCTAACTTGCTGCCACACAC		
4	XM003792- Cspg4-TF1	AAATCTAAGCGCGGGTCTGGC	WT:0bp KI:272bp	T006187 <i>Cspg4</i> - CreERT2
	XM003792- Cspg4-TR1	TGCGAACCTCATCACTCGTTGC		
5	XM003792- Cspg4-TF2	AAATCTAAGCGCGGGTCTGGC	WT:352bp (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

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

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

Locus	HCT116-luc		HCT116-luc-LM3	
	X	Y	X	Y
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	
	X	Y	X	Y
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.

<b>Genes</b>	<b>Forward (5'-3')</b>	<b>Reverse (5'-3')</b>
<i>ACTB</i>	GTTGCTATCCAGGCTGTGCTATCC	GGTGGCAGTGATGGCATGGAC
<i>MFAP4</i>	TGAAGGCACAAGGAGTTCTCT	GGGTAGATGAGGTACACGCC
<i>MMP2</i>	CGACCACAGCCAACTACGATGATG	GTGCCAAGGTCAATGTCAGGAGAG
<i>FBLN1</i>	TGCTCCATCAACGAGACCTG	AGCACTCCCGATTCTCATGG
<i>COL1A2</i>	CCGTGGCAGTGATGGAAGTGTG	CCTTGTTACCGCTCTCTCCTTTGG
<i>CHI3L1</i>	GGCTTCTTCTGAGACTGGTGTGG	CGCTTTCCTGGTCGTCGTATCC
<i>COL3A1</i>	TGTACCAGCCAGACCAGGAAGAC	TGTACCAGCCAGACCAGGAAGAC
<i>FBLN2</i>	GACCGAGGACAGTGAGGAGGAAG	CAGGCAGTGATGTGGACAGGATG
<i>IGFBP5</i>	GTACCTGCCCAATTGTGACC	AAGTCCCCGTCAACGTACTC
<i>MATN2</i>	AGAGGTGTGTGGCTGTGGACTAC	GAGCACTGGCAGACGAAGGAATC
<i>ITGA5</i>	GTCGGGGGCTTCAACTTAGAC	ACAGAGGTAGACAGCACCAC
<i>CILP</i>	CTTTGAGAACCTCCGGGCAT	TCGATCCCCCTCAATCTGGT

<i>C11orf96</i>	TCCAGTTACCAGGCGGTGAT	TGCGTCTTGAAGCGAGACTG
<i>A2M</i>	GAGGCAGAAGGACAATGGCT	ATAGGCGGAGAGGGTCACTT
<i>SFRP2</i>	GCCCGACTTCTCCTACAAGC	CTCCTTCATGGTCTCGTGGC
<i>MAF</i>	CGTCCTCTCCCGAGTTTTTCA	GGCTTCCAAAATGTGGCGTA
<i>PTGDS</i>	CCATGTGCAAGTCTGTGGTG	CATGGTTCGGGTCTCACACT
<i>TCF21</i>	TCCTGGCTAACGACAAATACGA	TTCCCGGCCACCATAAAGG

Supplemental Table 14. Sequences of siRNAs.

SiRNA	Sense (5'-3')	Antisense (5'-3')
siTCF21-1	GGAUUCGAACAAGGAAUUUTT	AAAUCCUUGUUCGAAUCCTT
siTCF21-2	GCUAACGACAAAUACGAGATT	UCUCGUAUUUGUCGUUAGCTT
siITGB1-1	GAACAGAUCUGAUGAAUGATT	UCAUUCAUCAGAUCUGUUCTT
siITGB1-2	GUGGUUUCGAUGCCAUCAUTT	AUGAUGGCAUCGAAACCACTT

siITGB1-3	GAUCAUUGAUGCAUACAAUTT	AUUGUAUGCAUCAAAUGAUCTT
siITGA2-1	CCCGAGCACAUCAUUUUAUATT	UAUAAAUGAUGUGCUCGGGTT
siITGA2-2	GCUGGUGACAUCAGUUGUATT	UACAACUGAUGUCACCAGCTT
siITGA2-3	GUGGUUGUGUGUGAUGAAUTT	AUUCAUCACACACAACCACTT
siMATN2-1	GCAUCCUAAUCUUUGCCAUTT	AUGGCAAAGAUUAGGAUGCTT
siMATN2-2	GCAGUUUGUCACUGGAAUUTT	AAUCCAGUGACAAACUGCTT

---

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

<b>Antibody</b>	<b>RRID</b>	<b>Company</b>
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 $\alpha$ 5	AB_2233962	Cell Signaling Technology
GAPDH	AB_10622025	Cell Signaling Technology
p-FAK <sup>Tyr397</sup>	AB_10891442	Cell Signaling Technology
FAK	AB_2799801	Cell Signaling Technology
p-PI3K p85 <sup>(Tyr458)</sup> /p55 <sup>(Tyr199)</sup>	AB_2895293	Cell Signaling Technology
PI3K	AB_2165248	Cell Signaling Technology
p-AKT <sup>Ser473</sup>	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 $\beta$ 1/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
$\alpha$ -SMA	AB_2799045	Cell Signaling Technology
Integrin $\alpha$ 5	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