Original research

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

Xiaobo Li,[1] Jinghua Pan,[2] Tongzheng Liu,[1] Wenqian Yin,[1] Qun Miao,[1] Zhan Zhao,[2] Yufeng Gao,[3] Wei Zheng,[3] Hang Li,[4] Rong Deng,[5] Dandan Huang,[6] Shenghui Qiu,[2] Yiran Zhang,[2] Qi Qi,[7] Lijuan Deng,[8] Maohua Huang,[1] Patrick Ming-Kuen Tang,[9] Yihai Cao ,[10] Minfeng Chen,[1] Wencai Ye,[1] Dongmei Zhang [1]

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/gutjnl-2022-327913).

For numbered affiliations see end of article.

Correspondence to
Professor Dongmei Zhang, Professor Wencai Ye and Dr Minfeng Chen, College of Pharmacy, Jinan University, Guangzhou, Guangdong, China; dmzhang701@jnu.edu.cn, chywc@aliyun.com, minfengchen@jnu.edu.cn and Professor Yihai Cao, Department of Microbiology, Tumor and Cell Biology, Karolinska Institute, Stockholm, Stockholm, Sweden; yihai.cao@ki.se

XL, JP and TL contributed equally.

Received 22 May 2022
Accepted 20 August 2022
Published Online First
7 September 2022

To cite: Li X, Pan J, Liu T, *et al*. *Gut* 2023;**72**:710–721.

ABSTRACT

Objective Haematogenous dissemination is a prevalent route of colorectal cancer (CRC) metastasis. However, as the gatekeeper of vessels, the role of tumour pericytes (TPCs) in haematogenous metastasis remains largely unknown. Here, we aimed to investigate the heterogeneity of TPCs and their effects on CRC metastasis.

Design TPCs were isolated from patients with CRC with or without liver metastases and analysed by single-cell RNA sequencing (scRNA-seq). Clinical CRC specimens were collected to analyse the association between the molecular profiling of TPCs and CRC metastasis. RNA-sequencing, chromatin immunoprecipitation-sequencing and bisulfite-sequencing were performed to investigate the TCF21-regulated genes and mechanisms underlying integrin α5 on *TCF21* DNA hypermethylation. Pericyte-conditional *Tcf21*-knockout mice were constructed to investigate the effects of TCF21 in TPCs on CRC metastasis. Masson staining, atomic force microscopy, second-harmonic generation and two-photon fluorescence microscopy were employed to observe perivascular extracellular matrix (ECM) remodelling.

Results Thirteen TPC subpopulations were identified by scRNA-seq. A novel subset of TCF21^{high} TPCs, termed 'matrix–pericytes', was associated with liver metastasis in patients with CRC. TCF21 in TPCs increased perivascular ECM stiffness, collagen rearrangement and basement membrane degradation, establishing a perivascular metastatic microenvironment to instigate colorectal cancer liver metastasis (CRCLM). *Tcf21* depletion in TPCs mitigated perivascular ECM remodelling and CRCLM, whereas the coinjection of TCF21^{high} TPCs and CRC cells markedly promoted CRCLM. Mechanistically, loss of integrin α5 inhibited the FAK/PI3K/AKT/DNMT1 axis to impair *TCF21* DNA hypermethylation in TCF21^{high} TPCs.

Conclusion This study uncovers a previously unidentified role of TPCs in haematogenous metastasis and provides a potential diagnostic marker and therapeutic target for CRC metastasis.

SIGNIFICANCE OF THIS STUDY

WHAT IS ALREADY KNOWN ON THIS SUBJECT?

- ⇒ Colorectal cancer (CRC) metastasises mainly to the liver, which results from haematogenous dissemination.
- ⇒ Tumour pericytes (TPCs) are the major components of tumour vessels; their heterogeneity and functions in CRC metastasis remain largely unknown.

WHAT ARE THE NEW FINDINGS?

- ⇒ The heterogeneity of TPCs derived from patients with CRC was dissected by single-cell RNA sequencing.
- ⇒ This study elucidated an undefined role of TPCs in promoting CRC metastasis.
- ⇒ We identified a novel TPC-related prometastatic signalling pathway for CRC metastasis.

HOW MIGHT IT IMPACT ON CLINICAL PRACTICE IN THE FORESEEABLE FUTURE?

- ⇒ This study provides new insights into the heterogeneity of TPCs and their prometastatic effects on CRC.
- ⇒ Therapeutic targeting of the prometastatic TCF21^{high} TPCs may represent a new paradigm for cancer therapy.

INTRODUCTION

Colorectal cancer (CRC) is the third most common malignant tumour worldwide, and colorectal cancer liver metastasis (CRCLM) counts as the leading cause of disease mortality.[1] CRCLM results from haematogenous dissemination,[2] which refers to tumour cell intravasation into circulation and extravasation from blood vessels, followed by the formation of metastatic foci in liver.[3] Tumour intravasation is a critical and rate-limiting step of haematogenous metastasis, during which tumour cells invade the perivascular extracellular matrix (ECM), breach the endothelial barrier and enter blood circulation.[4 5] During intravasation, the invading tumour cells contact endothelial and immune cells to form a tumour microenvironment of metastasis (TMEM).[6] Pericytes are contractile cells embedded within the capillary walls and wrapped around endothelial cells, where they act as the supervisors of endothelial cells to regulate vessel stabilisation and vascular permeability.[7] However, the role of tumour pericytes (TPCs) in haematogenous metastasis remains controversial.

TPCs attached to endothelial cells act as physiological barriers to inhibit tumour cell intravasation. Genetic depletion of NG2$^+$ or PDGFRβ$^+$ pericytes or pharmaceutical inhibition of pericyte recruitment has been shown to enhance vascular permeability and intratumoral hypoxia, which promotes tumour cell epithelial-to-mesenchymal transition (EMT) and tumour metastasis.[8–10] Tumor-derived PDGF-BB induces pericyte–fibroblast transition (PFT), leading to cell detachment from blood vessels or secretion of interleukin-33 to recruit tumor-associated macrophages, enhancing haematogenous metastasis.[11 12] Tumor-derived exosomes activate KLF4 in pericytes, resulting in cell detachment and abundant deposition of fibronectin in secondary organs, thereby establishing a premetastatic niche to facilitate haematogenous metastasis.[13] However, TPCs also play a prometastatic role. CD45$^-$VLA-1^{bri} or endosialin-expressing pericytes in primary tumours facilitate haematogenous metastasis by promoting tumour cell intravasation in a cell contact-dependent manner without altering the tumour vasculature structure or permeability.[14 15] These contradictory effects of TPCs on haematogenous metastasis may be associated with their heterogeneity,[16] which has not been determined.

TCF21, a member of the basic helix–loop–helix family of transcription factors, is critical for embryogenesis. Deficiency of TCF21 leads to abnormalities in multiple organs and neonatal lethality.[17] High expression of TCF21 in coronary artery smooth muscle cells inhibits their differentiation and promotes their migration and proliferation, contributing to stable atherosclerotic plaques and reducing the risk of coronary artery disease.[18 19] TCF21 also drives the expression of inflammatory genes and deposition of collagen IV in visceral adipose stem cells.[20] TCF21 acts as a tumour suppressor in various types of tumours, which is low expressed in tumour cells owing to its promoter hypermethylation.[17] However, the expression profile and function of TCF21 in TPCs remain unknown.

Here, we employed single-cell RNA sequencing (scRNA-seq) to dissect the heterogeneity of TPCs derived from patients with CRC with or without liver metastases, and 13 distinct TPC subpopulations were identified. Among them, a novel subset of TCF21^{high} TPCs, termed matrix–pericytes, was correlated with CRCLM. TCF21 in TPCs increased perivascular collagen deposition and rearrangement and basement membrane degradation, facilitating the establishment of the perivascular metastatic microenvironment (PMM) to enhance tumour cell intravasation. Pericyte-specific knockout of *Tcf21* inhibited CRCLM by reducing perivascular ECM remodelling, maintaining the integrity of the basement membrane, and decreasing circulating tumour cells (CTCs). Moreover, the expression of TCF21 in matrix–pericytes was regulated by the loss of integrin α5, which conferred to the DNA hypermethylation of *TCF21* via the FAK/PI3K/AKT-DNMT1 axis.

MATERIALS AND METHODS

The detailed materials and methods can be found in the online supplemental materials.

RESULTS

Identification of a distinct subpopulation of TPCs associated with CRC metastasis

To dissect the heterogeneity of TPCs and assess their contribution to CRC metastasis, TPCs were isolated from primary tumour tissues of patients with CRC with or without liver metastases and were analysed by scRNA-seq using the 10× chromium platform. Single-cell transcriptomes were generated from 50 000 cells, which were juxtaposed in t-stochastic neighbour embedding (t-SNE) space to identify distinct clusters (figure 1A). The gene expression profiles of TPCs were classified into 13 distinct clusters (figure 1B), and their origins were determined using known pericyte markers (online supplemental figures 1A, B). The top 10 cluster-specific genes are shown in figure 1C . The proportion of cells in cluster 2 showed the greatest difference among all subsets between patients with CRC with (18.5%) and without (3.4%) liver metastases (figure 1D), indicating that this subset was associated with CRCLM. Gene Ontology (GO) enrichment analysis indicated that the upregulated genes in samples from patients with CRC with liver metastases were associated with several GO terms related to the ECM (figure 1E), including *MATN2*, *CHI3L1*, *COL3A1*, *COL1A2*, *CILP*, *MMP2*, *MFAP4*, *FBLN1* and *FBLN2* (figure 1F and online supplemental figure 2). Therefore, cells in cluster 2 were defined as matrix–pericytes. One of the genes exclusively enriched in cluster 2, *MATN2*, which encodes the protein matrilin-2 (MATN2), was selected as a biomarker for matrix–pericytes. The ratio of MATN2$^+$ TPCs was higher in tumour sections derived from patients with CRC with liver metastases than in those without liver metastases (figure 1G and online supplemental table 1). In addition, receiver operating characteristic (ROC) curve analysis showed that the optimal cut-off percentage of MATN2$^+$ TPCs to predict CRCLM with high sensitivity and specificity was 30% (figure 1H). Kaplan-Meier survival analysis showed that overall survival (OS) (figure 1I) and disease-free survival (DFS) (figure 1J) were shorter in patients with a high (>30%) ratio of MATN2$^+$ TPCs. Furthermore, the correlation analysis between the MATN2$^+$ TPC ratio and clinicopathological parameters showed that the MATN2$^+$ TPC ratio had a positive correlation with the TNM stage and liver metastasis of patients with CRC (online supplemental table 2). Taken together, these data indicate that matrix–pericytes are notably associated with CRC metastasis.

TCF21 in TPCs is associated with CRC metastasis

Key regulators of matrix–pericytes were identified using the single-cell regulatory network inference and clustering (SCENIC) pipeline, which connects cis-regulatory sequence information with scRNA-seq data.[21] SCENIC analysis showed that the regulon activity of TCF21 was the highest in matrix–pericytes (figure 2A), which was confirmed in the t-SNE space of all cells (figure 2B). In addition, the number of TCF21^{high} TPCs was markedly increased in samples from patients with CRC with liver metastases compared with those without liver metastases (figure 2C). In addition, the ratio of TCF21$^+$ TPCs was increased in tumour sections from patients with CRC with liver metastases (figure 2D and online supplemental table 1). However, TCF21 was undetectable in TPCs in hepatic metastatic tumours derived from patients with CRC (online supplemental figure 3). Pearson's correlation coefficient analysis showed a positive correlation between the ratio of TCF21$^+$ TPCs and MATN2$^+$ matrix–pericytes ($r=0.805$, $P<0.001$; figure 2E). Additionally, flow cytometry analysis showed that the proportion of TCF21$^+$MATN2$^+$ TPCs was significantly increased in TPCs derived from patients with CRC without metastases (TPC$_{NM}$) infected with lentivirus harbouring TCF21-overexpressing plasmid (TPC$_{NM}$^{TCF21}) compared with those harbouring vector (TPC$_{NM}$^{Vector}) (online supplemental figure 4A). In contrast, knockdown of TCF21 in TPCs derived from patients with CRC with liver metastases (TPC$_{LM}$) showed opposite effects (online supplemental figure 4B). These data indicate that TCF21$^+$ TPCs are closely associated with matrix–pericytes. ROC curve analysis revealed that the optimal percentage of TCF21$^+$ TPCs for predicting liver metastases in patients with CRC was 44% (figure 2F). Kaplan-Meier survival analysis suggested that OS (figure 2G) and DFS (figure 2H) were significantly shorter in

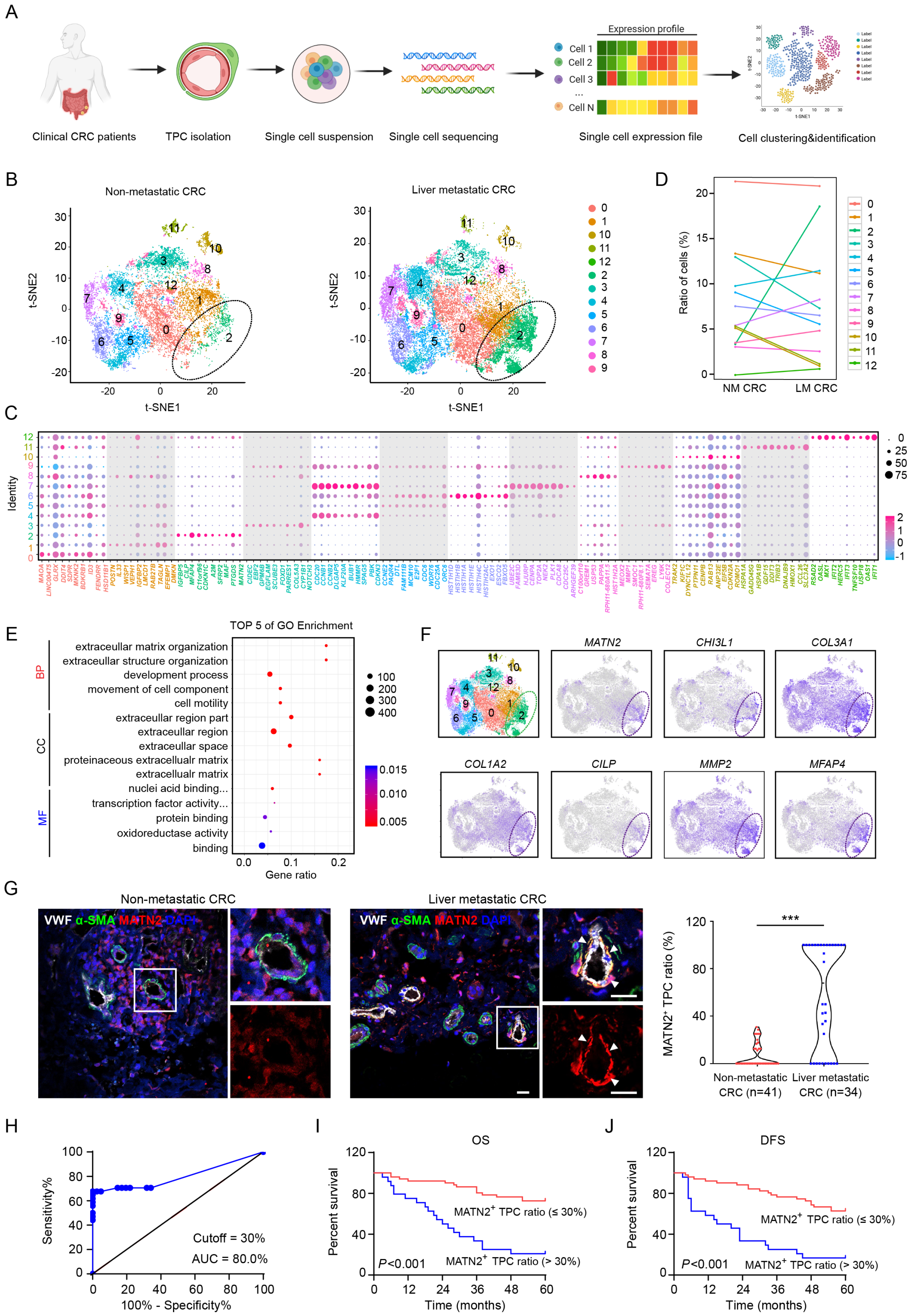


Figure 1 Transcriptomic characterisation of matrix-pericytes in TPCs derived from CRC patients. (A) Schematic diagram of scRNA-seq for TPCs. (B) t-SNE visualisation of TPC subsets derived from primary tumour tissues of CRC patients with (n=2) or without (n=2) liver metastasis. (C) Dot plots of the top ten marker genes in each TPC subset. (D) Percentage of each subset of TPCs. (E) GO analysis of the upregulated genes in cluster 2. (F) t-SNE visualisation of TPCs at all stages is combined, overlaid with the expression of indicated genes. (G) Immunofluorescence analysis and quantification of MATN2$^+$ matrix-pericyte ratio in primary tumour tissues from CRC patients with (n=34) or without (n=41) liver metastasis. white arrows indicate the staining of MATN2 in TPCs. scale bar, 20 µm. each sample on the violin plots represents individual patient data. ***$P < 0.001$ by two-tailed Mann-Whitney test. (H) ROC curve analysis of MATN2$^+$ TPC ratio in CRC patients (n=75). (I, J) Kaplan-Meier analysis of the OS (I) or DFS (J) in CRC patients with high or low MATN2$^+$ matrix-pericyte ratio (based on 30% cut-off, n=75). $P<0.001$ by Log-rank (Mantel-Cox) test. BP, biological process; CC, cellular component; CRC, colorectal cancer; DAPI, 4',6-diamidino-2-phenylindole; DFS, disease-free survival; GO, Gene Ontology; LM CRC, liver metastatic colorectal cancer; MF, molecular function; NM CRC, non-metastatic colorectal cancer; OS, overall survival; ROC, receiver operating characteristic; TPC, tumour pericyte; t-SNE, t-stochastic neighbour embedding; VWF, von willebrand factor.

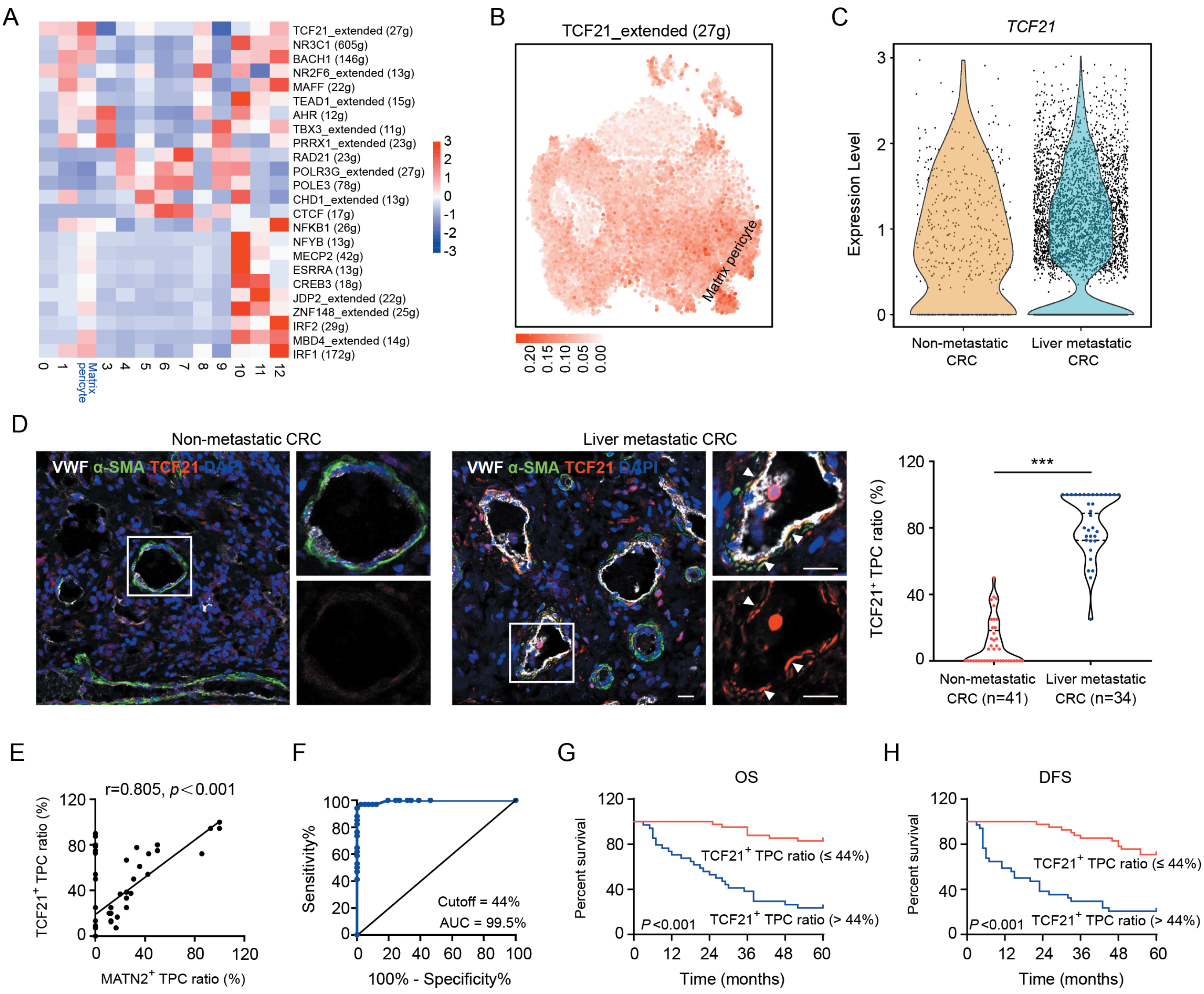


Figure 2 TCF21 in matrix–pericytes is positively associated with liver metastasis of CRC. (A) Heat map analysis of the regulon activities in distinct subsets of TPCs. (B) t-SNE visualisation of TCF21 regulon modulus in all TPC subsets. (C) Expression of *TCF21* in matrix–pericytes derived from CRC patients with or without liver metastasis. (D) Immunofluorescence staining and quantification of TCF21$^+$ TPC ratio in tumour sections from CRC patients with (n=34) or without (n=41) liver metastases.White arrows indicate TCF21 staining in TPCs. Scale bar, 20 µm. Each sample on the violin plots represents individual patient data. ***P<0.001 by two-tailed Mann-Whitney test. (E) Pearson's correlation analysis of TCF21$^+$ TPC ratio and MATN2$^+$ matrix–pericyte ratio (n=75). (F) ROC curve analysis of TCF21$^+$ TPC ratio in CRC patients (n=75). (G, H) Kaplan-Meier analysis of OS (G) or DFS (H) in CRC patients with high or low TCF21$^+$ TPC ratio (based on the 44% cut-off, n=75). P<0.001 by log-rank (Mantel-Cox) test. CRC, colorectal cancer; DAPI, 4',6-diamidino-2-phenylindole; DFS, disease-free survival; OS, overall survival; ROC, receiver operating characteristic; TPC, tumour pericyte; t-SNE, t-stochastic neighbour embedding; VWF, von willebrand factor.

patients with higher (>44%) ratio of TCF21$^+$ TPCs. Furthermore, the TCF21$^+$ TPC ratio was notably correlated with TNM stage and liver metastasis in patients with CRC (online supplemental table 3). These findings indicate that TCF21 in TPCs can serve as a predictive biomarker for CRC metastasis.

TCF21 in TPCs contributes to the phenotype transition of matrix–pericytes and the remodelling of perivascular ECM

The positive correlation between TFC21 upregulation in TPCs and CRCLM suggested that TCF21 may be involved in the activation of TPCs and may promote the phenotype transition of matrix–pericytes. Depletion or overexpression of TCF21 in TPCs had negligible effects on cell proliferation, adhesion and migration (online supplemental figure 5). The levels of nine genes in matrix–pericytes, including *IGFBP5*, *CILP*, *MFAP4*, *c11orf96*, *A2M*, *SFRP2*, *MAF*, *PTGDS* and *MATN2*, were significantly higher in TPC_{NM}^{TCF21} than in TPC_{NM}^{Vector} (figure 3A). Conversely, the levels of these genes were significantly lower in TPC_{LM} transfected with siTCF21 ($TPC_{LM}^{siTCF21}$) than in those transfected with negative control siRNA (TPC_{LM}^{siNC}) (online supplemental figure 6A). In addition, the percentage of MATN2$^+$ matrix–pericytes was increased by overexpression of TCF21 in TPC_{NM} and decreased by the knockdown of TCF21 in TPC_{LM} (figure 3B and online supplemental figure 6B).

Genes acting downstream of TCF21 were investigated in TPCs using whole-transcriptome RNA-seq and ChIP-seq assays (online supplemental figure 6C, D). RNA-seq assays showed that 1276

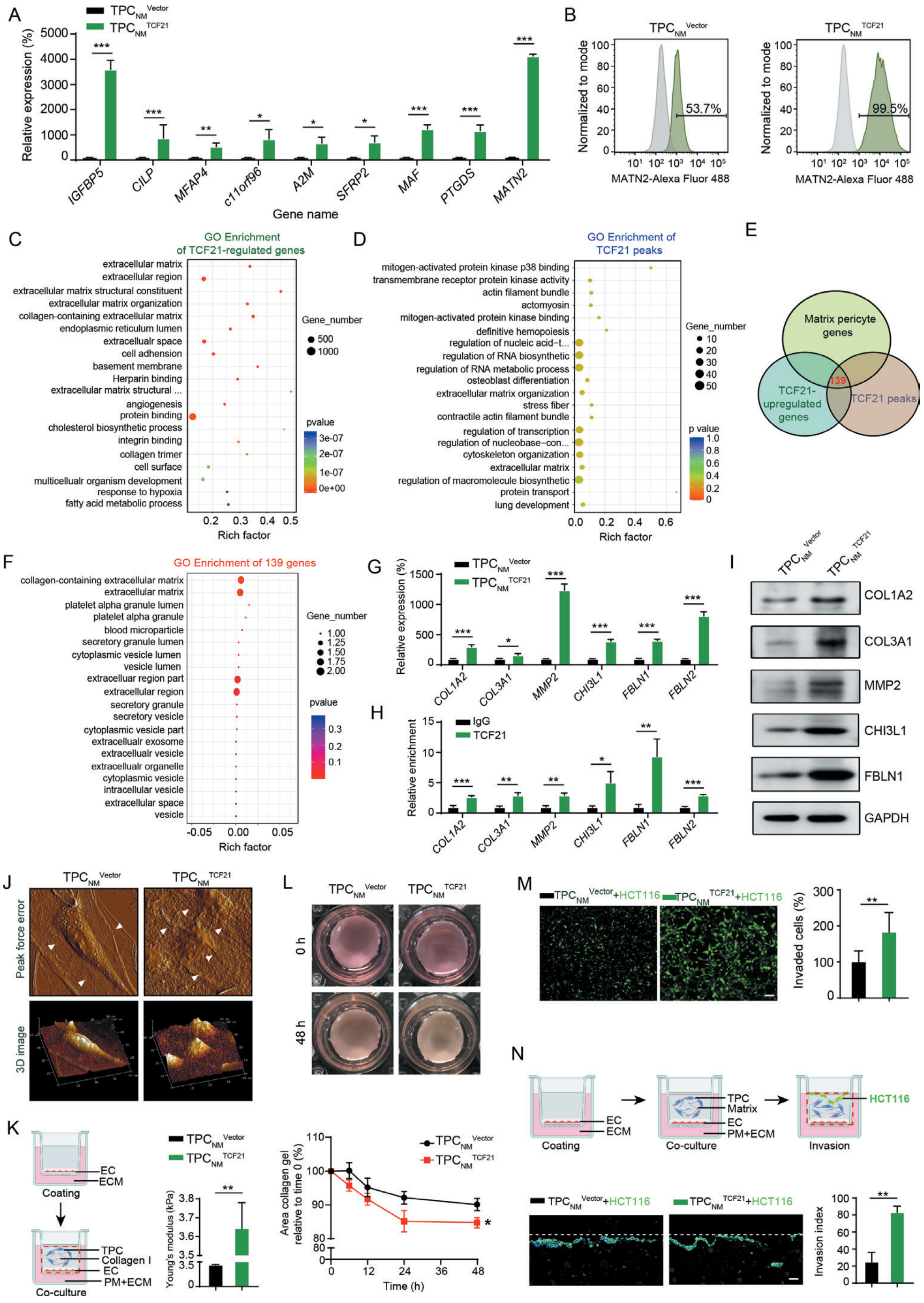


Figure 3 TCF21 is essential for the formation of matrix–pericytes and facilitates CRC cell invasion by ECM remodelling. (A) RT-qPCR analysis for matrix–pericyte-related genes in TPC$_{NM}^{Vector}$ or TPC$_{NM}^{TCF21}$ (n=3). (B) FCM analysis for MATN2 expression in TPC$_{NM}^{Vector}$ and TPC$_{NM}^{TCF21}$ (n=3). (C) GO terms of differentially expressed genes derived from RNA-seq in TPC$_{NM}^{TCF21}$ compared with those in TPC$_{NM}^{Vector}$. (D) GO terms of TCF21 peaks derived from ChIP-Seq in TPC$_{NM}^{TCF21}$. (E) Venn diagram showing the number of overlapped genes derived from TCF21-upregulated genes, TCF21 peaks and matrix–pericyte genes. (F) GO analysis for the intersected genes derived from E. (G) RT-qPCR analysis for the indicated genes in TPC$_{NM}^{Vector}$ and TPC$_{NM}^{TCF21}$ (n=3). (H) ChIP-qPCR analysis for the binding of TCF21 with the promoter of indicated genes in TPC$_{NM}^{TCF21}$ (n=3). (I) Western blotting analysis for the expression of indicated proteins in TPC$_{NM}^{Vector}$ and TPC$_{NM}^{TCF21}$ (n=3). (J) AFM imaging of topography and roughness of the collagen fibres incubated with TPC$_{NM}^{Vector}$ or TPC$_{NM}^{TCF21}$ (n=3). White arrows indicate the collagen fibres. (K) Young's modulus detection of the stiffness of collagen gels (n=3). Schematics of the 3D coculture model (left) and Young's modulus detection (right) are shown. Red dotted line indicates the removed part. (L) Representative images and quantification of collagen contraction caused by TPC$_{NM}^{Vector}$ or TPC$_{NM}^{TCF21}$ (n=3). (M) Representative images and quantification of invaded HCT116 cells (green) (n=3). Scale bar, 50 µm. (N) Schematics of organotypical culture system for studying the invasion of tumour cells (upper). Representative images and quantification of HCT116 cells that invaded into the TPC-containing matrix (bottom) (n=3). Scale bar, 20 µm. Red dotted line indicates the removed part; white lines indicate the initiation. Data are presented as mean±SEM. *$P<0.05$, **$P<0.01$, ***$P<0.001$ by two-tailed unpaired t-test. AFM, atomic force microscopy; ChIP-qPCR, chromatin immunoprecipitation-quantitative PCR; CRC, colorectal cancer; EC, endothelial cell; ECM, endothelial culture medium; GO, Gene Ontology; RT-qPCR, real-time quantitative PCR; TPC, tumour pericyte.

genes were upregulated, and 1397 genes were downregulated in TPC_{NM}^{TCF21} compared with TPC_{NM}^{Vector} (online supplemental figure 6E). A heatmap of the RNA-seq results showed that the expression of matrix–pericyte-specific genes, including *MFAP4*, *CILP*, *MATN2*, *MMP2*, *COL3A1*, *COL1A2*, *CHI3L1* and *FBLN1*, was significantly higher in TPC_{NM}^{TCF21} (online supplemental figure 6F). GO enrichment analysis of the RNA-seq and ChIP-seq results showed that genes regulated by TCF21 were associated with the categories 'ECM organisation' and 'ECM' (figure 3C,D), which were similar to the expression profiles in matrix–pericytes. Moreover, comparison of the TCF21 peaks detected by ChIP-seq analysis, TCF21-induced genes detected by RNA-seq and matrix–pericyte genes detected by scRNA-seq identified 139 overlapping genes (figure 3E). GO analysis of these overlapping genes showed their associations with the categories 'collagen-containing ECM' and ECM (figure 3F). These results were confirmed by RT-qPCR and ChIP-qPCR, and the binding of TCF21 to the promoter of matrix–pericyte-specific genes was significantly upregulated in TCF21-overexpressing TPC_{NM} (figure 3G,H, and online supplemental figure 6G and H), which were confirmed by western blotting (figure 3I and online supplemental figure 6I). These results indicate that TCF21 is critical for the generation of matrix–pericytes.

ECM remodelling is crucial for tumour metastasis,[22] which is represented by aberrant collagen production and cross-linking leading to tissue stiffness, thus promoting tumour metastasis.[23 24] Moreover, degradation of the vascular basement membrane by proteases such as matrix metalloproteinases facilitates the escape of tumour cells from primary tumour sites.[25] We proposed that TCF21 may play a role in perivascular ECM remodelling and CRC metastasis. Neither overexpression nor knockdown of TCF21 in TPCs altered tumour cell proliferation, migration, EMT or endothelial cell tube formation in vitro (online supplemental figure 7A–H). Atomic force microscopy (AFM) revealed that TPC_{NM}^{TCF21} possessed greater ability than TPC_{NM}^{Vector} in inducing local alignment of collagen, reorganisation of curly and sheet-like collagen fibres into radial and bundling structures, and increased roughness (figure 3J). Moreover, TPC_{NM}^{TCF21} was superior to TPC_{NM}^{Vector} in stiffening the mechanical properties of collagen (figure 3K) and increasing collagen contraction (figure 3L), thereby enhancing ECM stiffness. To further assess the effects of TCF21-mediated perivascular ECM remodelling on the transmigration of CRC cells, PKH-67-labelled HCT116 or DLD-1 cells mixed with TPC_{NM}^{Vector}, TPC_{NM}^{TCF21}, TPC_{LM}^{siNC} or $TPC_{LM}^{siTCF21}$ were seeded into Matrigel. TPC_{NM}^{TCF21} was better than TPC_{NM}^{Vector} in facilitating CRC cell invasion through the Matrigel-coated transwell membranes (figure 3M and online supplemental figure 8A), whereas $TPC_{LM}^{siTCF21}$ reduced the invasion of CRC cells compared with TPC_{LM}^{siNC} (online supplemental figure 8B, C). The experiments performed on an organotypical culture system showed that the number of invaded HCT116 cells was significantly increased when matrix, consisting of collagen I and Matrigel, was premixed with TPC_{NM}^{TCF21} compared with TPC_{NM}^{Vector} (figure 3N). Collectively, these results indicate that TCF21 in matrix–pericytes induces perivascular ECM remodelling and facilitates CRC cell invasion through the perivascular matrix.

We further investigated whether MATN2 exerted prometastatic effects similar to those of TCF21. Given that $TCF21^{high}$ TPCs facilitated tumour metastasis through ECM remodelling, the levels of genes encoding ECM proteins, such as *COL1A2*, *COL3A1*, *MMP2*, *CHI3L1* and *FBLN1*, were examined by RT-qPCR in TPCs with overexpression or knockdown of MATN2 (online supplemental figure 9A, B). Our results showed that MATN2 in TPCs had negligible effects on the levels of these ECM-related genes (online supplemental figure 9C, D). Additionally, MATN2 in TPCs had negligible effects on tumour cell invasion (online supplemental figure 9E, F). These data indicate that $MATN2^+$ TPCs could not give metastatic phenotypes to tumour cells similar to $TCF21^{high}$ TPCs, which may merely serve as a characteristic marker for matrix–pericytes.

Knockout of *Tcf21* in TPCs inhibits CRC metastasis

To determine whether TCF21 in TPCs is essential for CRC metastasis in vivo, pericyte lineage-tracing mice (PC^{lin}) and tamoxifen-inducible *Cspg4*-driven pericyte-specific *Tcf21* knockout mice ($PC^{lin\text{-}KO}$) were generated (figure 4A and online supplemental figure 10A, B). Tamoxifen administration to both PC^{lin} and $PC^{lin\text{-}KO}$ mice resulted in the permanent labelling of pericytes with tdT fluorescence, and the knockout of *Tcf21* in $PC^{lin\text{-}KO}$ mice. To examine the effects of TCF21 in TPCs on CRC metastasis, PC^{lin} and $PC^{lin\text{-}KO}$ mice were orthotopically injected with luciferase-labelled tumour cells (MC38-luc-LM3), followed by treatment with tamoxifen for 1 week (online supplemental figure 11A). Tamoxifen treatment had negligible effects on the non-tumour tissues (online supplemental figure 11B) but induced Cre activity in TPCs, as indicated by tdT fluorescence in both PC^{lin} and $PC^{lin\text{-}KO}$ mice and loss of TCF21 in $PC^{lin\text{-}KO}$ mice (online supplemental figure 11C, D). In vivo, the TPC-specific deletion of *Tcf21* significantly inhibited the formation of liver metastases (figure 4B). The area and number of liver metastatic foci were significantly decreased in $PC^{lin\text{-}KO}$ mice compared with those in PC^{lin} mice (figure 4C). Furthermore, the number of $EpCAM^+CD45^-$ CTCs was significantly lower in $PC^{lin\text{-}KO}$ mice than that in PC^{lin} mice (figure 4D and online supplemental figure 12A, B). Knockout of *Tcf21* in TPCs had negligible effects on primary tumour growth and EMT, as indicated by equal levels of E-cadherin, vimentin, and Ki67 in tumour sections from PC^{lin} and $PC^{lin\text{-}KO}$ mice (online supplemental figure 13A, B). These data indicate that TCF21 in TPCs promotes CRC metastasis, independent on tumour growth and EMT.

The effects of TCF21 in TPCs on vascular structure and function were further evaluated by assessing pericyte coverage and the percentage of $MATN2^+$ matrix–pericytes in tumour vessels. Immunofluorescence staining showed that deletion of *Tcf21* in TPCs had negligible effects on total pericyte coverage (online supplemental figure 13C), whereas the number of $MATN2^+$ matrix–pericytes in primary tumour tissues of MC38-luc-LM3 CRCLM xenografts was lower in $PC^{lin\text{-}KO}$ mice than in PC^{lin} mice (figure 4E and online supplemental figure 12C). In addition, the expression of ECM remodelling-related proteins, including MMP2, COL1A2, COL3A1 and CHI3L1, in TPCs from PC^{lin} mice was significantly higher than those from $PC^{lin\text{-}KO}$ mice (figure 4F and online supplemental figure 12D). Masson staining showed that the density of fibrillar collagen components was significantly decreased in tumour sections from $PC^{lin\text{-}KO}$ mice compared with that in PC^{lin} mice (figure 4G and online supplemental figure 12E). Transmission electron microscopy analysis indicated that the collagen bundles at the perivascular region were thinner in tumour sections from $PC^{lin\text{-}KO}$ than in those from PC^{lin} mice (figure 4H). Collagen deposition is required for tissue stiffness and the local invasiveness of tumour cells, and changes in collagen orientation also contribute to tumour metastasis.[26] The structure of perivascular collagen was further examined by second-harmonic generation and two-photon excited fluorescence. Perivascular collagen fibres in $PC^{lin\text{-}KO}$ mice were arranged irregularly around the tumour vessels with curly and

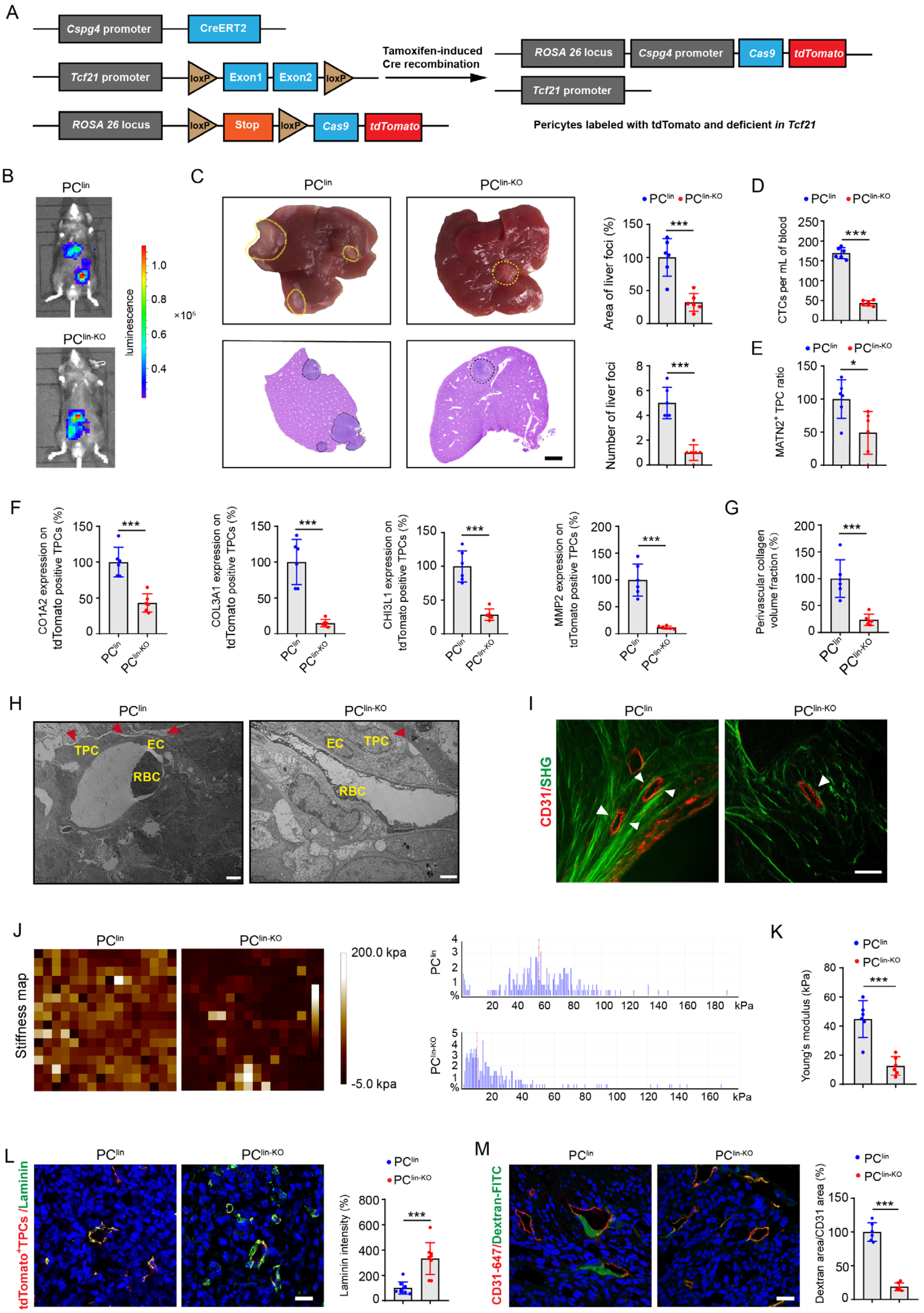


Figure 4 Pericyte-specific knockout of *Tcf21* inhibits the remodelling of perivascular ECM and CRC metastasis. (A) schematic diagram of the construction of *Cspg4* (NG2) lineage-tracing *Tcf21* inducible knockout mice. (B) Representative images and quantification of bioluminescence signals of mice orthotopically injected with MC38-luc-LM3 cells (n=6). (C) Representative images and H&E analysis of livers derived from MC38-luc-LM3 xenograft-bearing mice (n=6). Yellow and black dotted lines indicate the liver metastatic loci. Scale bar, 2 mm. (D) Quantification of CTCs (n=6). (E) Quantification of MATN2$^+$ TPC ratio in primary tumour sections (n=6). (F) Quantification of the expression of COL3A1, MMP2, COL1A1 and CHI3L1 on TPCs (tdTomato) in primary tumour sections (n=6). (G) Quantification of perivascular collagen in primary tumour sections (n=6). (H) Representative TEM images of collagen fibres surrounding TPCs (n=6). Red arrowheads indicate the perivascular collagen fibres. Scale bar, 2 µm. (I) Representative images of perivascular collagen organisation in tumours imaged by SHG (green) and CD31 (red) (n=6). White arrowheads indicate the perivascular collagen fibres. Scale bar, 50 µm. (J,K) Representative AFM images and quantification of the stiffness of the perivascular area in primary tumour sections (n=6). (L) Immunofluorescence staining and quantification of laminin (green) surrounding TPCs (tdTomato) in primary tumour sections (n=6). Scale bar, 20 µm. (M) Immunofluorescence staining and quantification of FITC–dextran 40 kD (green) surrounding vessels labelled with CD31 in primary tumour sections (n=6). Scale bar, 20 µm. Data are presented as mean±SEM. *P< 0.05, ***P <0.001 by two-tailed unpaired t-test. AFM, atomic force microscopy; CRC, colorectal cancer; CTC, circulating tumour cell; EC, endothelial cell; ECM, extracellular matrix; RBC, red blood cell; SHG, second-harmonic generation; TEM, transmission electron microscopy; TPC, tumour pericyte.

unfixed outlines, whereas perivascular collagen in PC^{lin} mice had an almost parallel orientation and were tightly embraced surrounded the tumour vessels (figure 4I). The stiffness of perivascular areas was determined by AFM and analysed with Young's modulus patterns, and we found that the perivascular areas in $PC^{lin\text{-}KO}$ mice were significantly softer than those in PC^{lin} mice (figure 4J,K). As the basement membrane serves as a physical barrier for tumour cell intravasation, degradation of the basement membrane could weaken its barrier function and increase vessel permeability.[27] Our results showed that the expression of laminin, the most abundant component in the vascular basement membrane, was significantly higher in $PC^{lin\text{-}KO}$ mice than that in PC^{lin} mice (figure 4L), and deletion of *Tcf21* in TPCs resulted in decreased vessel permeability in $PC^{lin\text{-}KO}$ mice compared with PC^{lin} mice (figure 4M). Taken together, TCF21 in matrix–pericytes might be a key regulator of perivascular ECM remodelling, which establishes a PMM to facilitate CRC cell intravasation.

Loss of integrin α5 attenuates DNA methylation of *TCF21* promoter and increases TCF21 expression in TPCs

Integrins, the key receptors of ECM components, function as mechanical transducers to facilitate tumour metastasis.[28] Our results showed that *ITGA2*, *ITGA5* and *ITGB1* were significantly decreased in matrix–pericytes among the 13 clusters of TPCs (figure 5A). Integrin α2 and integrin β1 were unrelated to TCF21, whereas integrin α5 expression was inversely correlated with TCF21 expression in TPCs (online supplemental figure 14A, B). Gain-of-function and loss-of-function experiments showed that TCF21 was negatively regulated by integrin α5 at both the mRNA and protein levels (figure 5B and online supplemental figure 14C-E), whereas integrin α2 and integrin β1 had negligible effects on TCF21 expression (online supplemental figure 14F).

Hypermethylation of *TCF21* DNA suppresses TCF21 expression in various types of tumour cells, including non-small cell lung cancer, head and neck squamous cell carcinoma, and CRC.[29–31] Evaluation of the methylation of CpG islands within the promoter of *TCF21* in TPCs showed that TCF21 expression in TPCs was inversely correlated with its methylation status, as indicated by the decreased 5-mC level of *TCF21* promoter of TPC_{LM} compared with TPC_{NM} (online supplemental figure 14G). In addition, integrin α5 positively regulated *TCF21* DNA hypermethylation, and loss of integrin α5 in TPCs attenuated DNA hypermethylation of *TCF21* promoter in TPCs (figure 5C and online supplemental figure 14H).

DNMT1 has been reported to regulate the DNA hypermethylation of *TCF21* in lung cancers.[32] Our results showed that depletion of integrin α5 reduced the expression of DNMT1 in TPC_{NM} by inhibiting the FAK/PI3K/AKT axis, whereas overexpression of integrin α5 in TPC_{LM} exerted opposite effects, which could be reversed by the FAK inhibitor Y15 (figure 5D and online supplemental figure 14I). Furthermore, either Y15 or DNMT inhibitor SGI1027 decreased the expression of DNMT1 and suppressed the DNA hypermethylation of *TCF21* (figure 5E), followed by an increase in the expression of TCF21 in TPC_{LM}^{ITGA5} (figure 5D). These results indicate that the loss of integrin α5 upregulates TCF21 in TPCs by suppressing the DNA hypermethylation of *TCF21*.

The effects of integrin α5 in TPCs on CRC metastasis in vivo were investigated in an orthotopic xenograft model generated by the co-injection of HCT116-luc-LM3 or DLD1-luc-LM3 cells and integrin α5-overexpressing or -knockdown TPCs (figure 5F). Compared with TPC_{NM}^{shNC}, $TPC_{NM}^{shITGA5}$ increased collagen density in perivascular areas of tumour tissues (figure 5G and online supplemental figure 15A, B), facilitated perivascular basement membrane degradation (online supplemental figure 15C, D) and significantly increased the number and area of metastatic foci in the liver (figure 5H and online supplemental figure 15E). Conversely, co-injection of tumour cells and TPC_{LM}^{ITGA5} showed the opposite effects compared with those with TPC_{LM}^{Vector} (figure 5G,H, and online supplemental figure 15A–E). These data indicate that loss of integrin α5 in TPCs promotes CRC metastasis.

$TCF21^{high}$ matrix–pericytes correlate with perivascular ECM remodelling and liver metastasis in patients with CRC

To determine whether the above findings were applicable to patients with CRC, perivascular collagen deposition and alignment were assessed in patients with CRC with and without liver metastases. Compared with patients with CRC without liver metastases, the perivascular collagen abundance, and the reorientation of perivascular collagen fibres into a radial alignment were more prominent in patients with CRC with liver metastases (figure 6A,B, and online supplemental figure 16A, B). In addition, the stiffness of perivascular areas was significantly enhanced in tumour tissues from patients with CRC with liver metastases compared with those without liver metastases (figure 6C,D). Moreover, the expression of COL1A2, COL3A1 and CHI3L1 (figure 6E), as well as MMP2, a key protease involved in the degradation of basement membrane (figure 6F), was higher in TPCs from patients with CRC with liver metastases than those without liver metastases. Correspondently, the expression of integrin α5 was lower in TPCs from patients with CRC with liver metastases than in those without liver metastases (figure 6G). Pearson correlation analysis indicated that the levels of MMP2, COL1A2, COL3A1 and CHI3L1 in TPCs were positively correlated, whereas the level of integrin α5 in TPCs was negatively correlated with the ratio of $TCF21^{high}$ TPCs (figure 6E–G). These clinical data indicate that integrin α5 loss-induced upregulation of TCF21 is an important regulator of perivascular ECM remodelling and the establishment of the PMM, thus facilitating CRC metastasis.

DISCUSSION

Cellular heterogeneity in the tumour microenvironment (TME) may result in distinct pathological phenotypes and different responses to cancer therapy.[33] The heterogeneity of tumour cells,[34] tumor-infiltrating immune cells,[35] cancer-associated fibroblasts (CAFs),[36] and endothelial cells[37] has been extensively investigated by scRNA-seq, which has also been employed to analyse pericytes derived from tumours, enteritis, and normal GI tract.[38–41] However, the specific phenotype and function of pericytes in tumour progression are largely unknown. Therefore, evaluating the heterogeneity of TPCs may provide new insights into the mechanisms underlying haematogenous metastasis. Here, the heterogeneity of TPCs was first revealed by scRNA-seq. By comparing the subpopulations of TPCs in our study with those in previous studies,[39 42] four novel subpopulations of TPCs were specifically identified in our work (online supplemental figure 17A–C).

Loss of NG2 is an important hallmark of PFT, during which pericytes shed from the blood vessels where they were originally attached.[11] We found that the expression of NG2 in clusters 10, 11 and 12 was extremely low, suggesting that PFT may have occurred in these subpopulations. However, the expression of NG2 in cluster 2 was not significantly decreased, indicating that

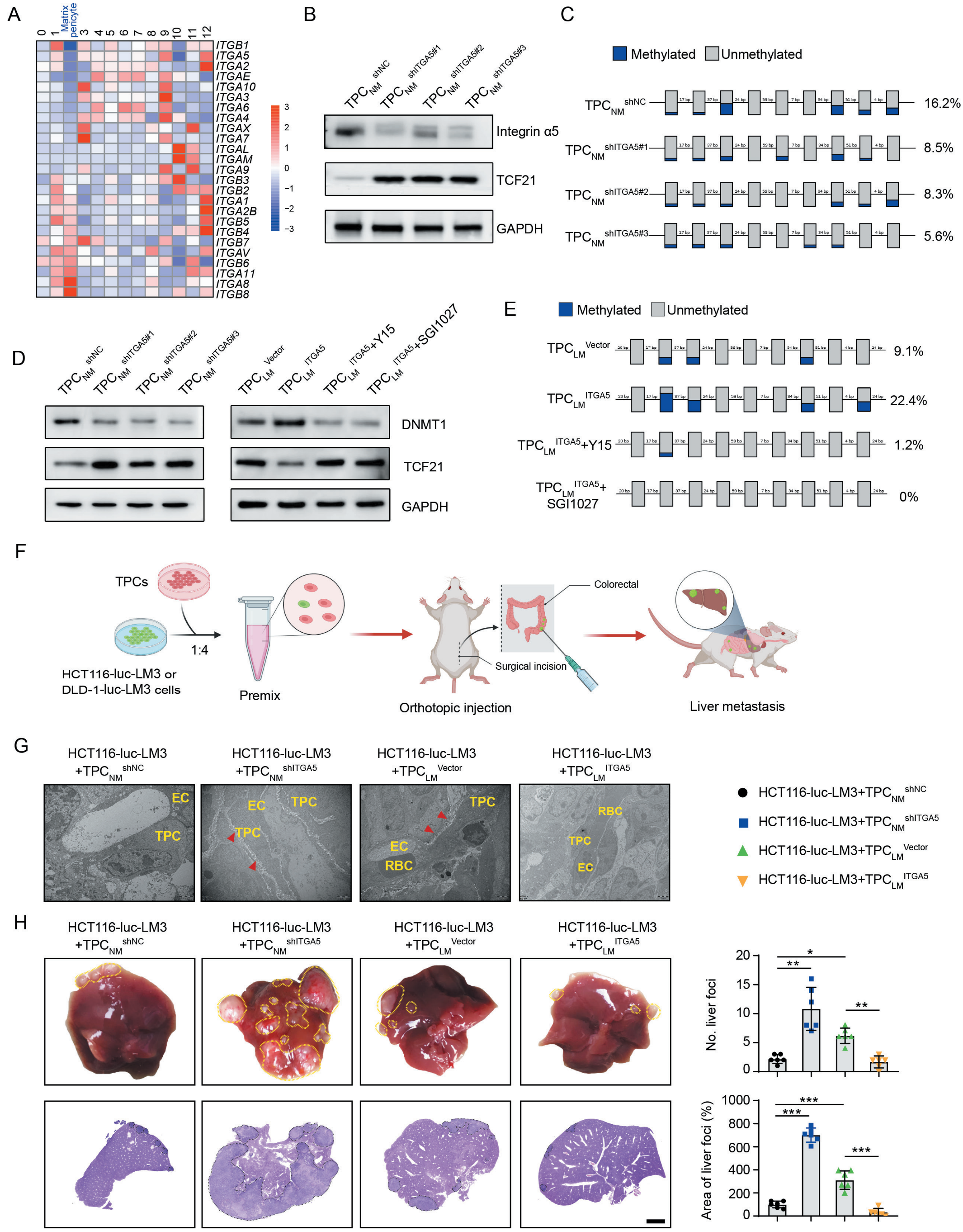


Figure 5 Loss of integrin α5 inhibits hypermethylation of *TCF21* promoter and upregulates the expression of TCF21 in TPCs. (A) Heat map of the expression of integrins in all subsets of TPCs. (B) Western blotting analysis of integrin α5 and TCF21 in integrin α5-knockdown TPC_{NM} (n=3). (C) BSP analysis for *TCF21* promotor region in integrin α5-knockdown TPC_{NM} (n=3). Blue and grey circles represent the methylated and unmethylated CpGs, respectively. The percentage of total methylated CpGs is given on the right. (D) Western blotting analysis of DNMT1 and TCF21 in integrin α5-knockdown (left) or integrin α5-overexpressing (right) TPCs treated with or without FAK inhibitor (Y15) or DNMT1 inhibitor (SGI1027) (n=3). (E) BSP analysis for the methylation and unmethylation levels of *TCF21* in TPC_{LM} treated with or without Y15 and SGI1027 (n=3). (F) Schematic diagram of the in vivo experiments. (G) Representative TEM images of collagen fibres surrounding TPCs in primary HCT116 orthotopic xenografts (n=6). Red arrowheads indicate perivascular collagen fibres. Scale bar, 2 µm. (H) Representative images of the whole liver and H&E analysis 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 analysis of variance followed by Tukey's post hoc test. BSP, bisulfite sequencing PCR; CRC, colorectal cancer; EC, endothelial cell; RBC, red blood cell; TEM, transmission electron microscopy; TPC, tumour pericyte.

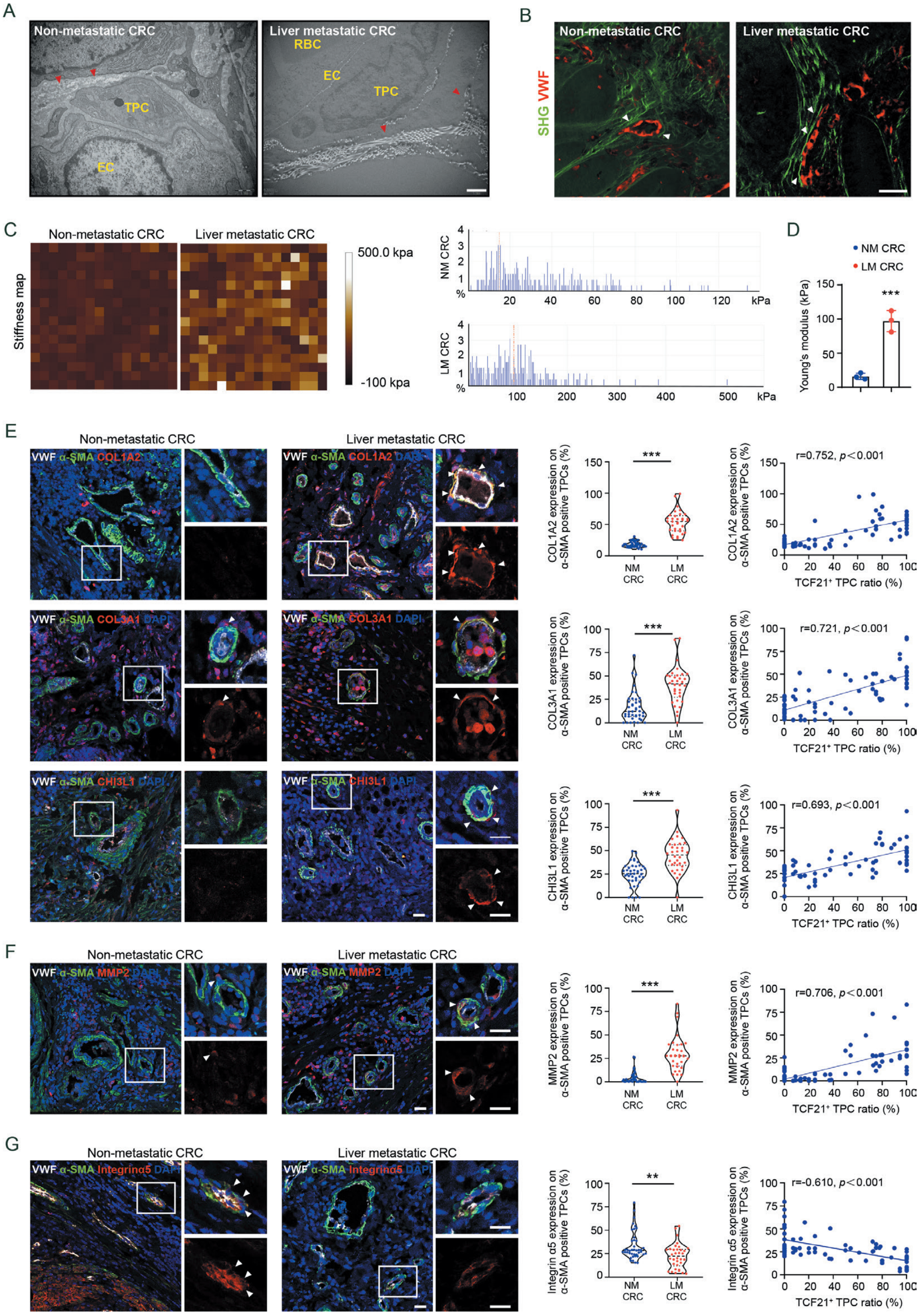


Figure 6 TCF21 in matrix–pericytes is associated with the remodelling of perivascular ECM in primary tumours derived from patients with CRC. (A) Representative TEM images of collagen fibres surrounding TPCs in primary tumour tissues from patients with CRC (n=3). Red arrowheads indicate the perivascular collagen fibres. Scale bar, 1 µm. (B) Representative images of perivascular collagen structure and density in patients with CRC imaged by SHG (green) and VWF (red) (n=3). White arrowheads indicate perivascular collagen fibres. Scale bar, 50 µm. (C,D) Representative AFM images and quantification of stiffness in perivascular area in primary tumour tissues from patients with CRC. Data are presented as mean±SEM (n=3).***P<0.001 by two-tailed unpaired t-test. (E–G) Immunofluorescence staining and quantification of COL1A2, COL3A1, CHI3L1 (E), MMP2 (F) and integrin α5 expression (G) in αSMA^+ TPCs in primary tumour sections from patients with CRC (n=75), and Pearson's correlation analysis for COL1A2, COL3A1, CHI3L1, MMP2 or integrin α5 expression in αSMA^+ TPCs with $TCF21^+$ TPC ratio (%) are shown (n=75). White arrowheads indicate the staining of indicated proteins in TPCs. Scale bar, 20 µm. Each sample on the violin plots represents individual patient data. **P<0.01, ***P<0.001 by two-tailed Mann-Whitney test. AFM, atomic force microscopy; DAPI, 4',6-diamidino-2-phenylindole; EC, endothelial cell; ECM, extracellular matrix; LM CRC, liver metastatic colorectal cancer; NM CRC, non-metastatic colorectal cancer; RBC, red blood cell; SHG, second-harmonic generation; TEM, transmission electron microscopy; TPC, tumour pericyte; VWF, von willebrand factor.

these TPCs retained perivascular characteristics despite being activated. Furthermore, trajectory analysis indicated that cluster 2 originated from cluster 9 and later evolved into cluster 12 (online supplemental figure 18A, B), further demonstrating that cluster 2 was a newly identified subset of TPCs. Therefore, our findings improve understanding of the heterogeneity of TPCs.

To initiate metastasis, tumour cells must break through the collagen-enriched ECM by matrix deposition and degradation, which generate 'tracks' or 'tunnels' that help tumour cells pass.[43] On reaching blood vessels, these tumour cells interact with proangiogenic TIE2^{high}/VEGF^{high} macrophages and luminal endothelial cells to construct a TMEM, which supports the transendothelial migration of tumour cells into blood circulation.[6] During intravasation, tumour cells physically contact with tumour endothelial cells via juxtracrine and paracrine signalling.[5] Although TPCs attach to endothelial cells, their roles in tumour cell intravasation and the underlying mechanisms remain unclear. Our study demonstrates that matrix–pericytes are one of the TPC subpopulations promote CRC metastasis by establishment of PMM, providing an unidentified function of TPCs in haematogenous metastasis.

TCF21 in tumour cells is considered to inhibit tumour growth and metastasis,[17] whereas the expression and function of TPCs in haematogenous metastasis remain largely unknown. In contrast to its suppressive role in tumour cells and CAFs,[44] TCF21 promoted phenotypical transition of TPCs into matrix–pericytes, which facilitated the haematogenous metastasis of CRC. The distinct functions of TCF21 in TPCs in haematogenous metastasis may be caused by the different methylation levels of the *TCF21*. DNA methylation of *TCF21* was lower in matrix–pericytes than in tumour cells, which increased the expression of TCF21 in TPCs from patients with CRC with liver metastases. DNA methylation of TCF21 was regulated by the FAK/AKT/DNMT1 signalling pathway, whereas the expression and activation of FAK in TPCs are negatively correlated with tumour angiogenesis, tumour growth, and metastasis.[45] However, the mechanism by which cancer cell regulates integrin α5 and TCF21 expression in TPCs has not been revealed. Metastatic CRC cells could rely on gene mutations[46] including *TP53*, *BRAF* and *KRAS* or extracellular vehicles (EVs)[47] to promote tumour metastasis. We found that the *TP53*, *BRAF* and *KRAS* mutations were not independent predictors of CRCLM (online supplemental table 4), and the above gene mutations were not associated with the TCF21$^+$ TPC ratio or the expression of integrin α5 in TPCs (online supplemental figure 19A and B). Interestingly, we found that metastatic CRC cells could reduce the expression of integrin α5 while increase the expression of TCF21 in an EV-dependent manner (online supplemental figure 19C, D). Nevertheless, the mechanisms underlying tumor-derived EVs and other factors involved in the regulation of TPCs require further investigation.

In conclusion, this study used scRNA-seq to reveal the heterogeneity of TPCs in patients with CRC and to identify a novel subpopulation of TCF21^{high} TPCs associated with CRCLM. These findings revealed the effects and mechanisms of TCF21^{high} TPCs in the construction of the PMM to facilitate CRC metastasis by remodelling the perivascular ECM and provided a potential diagnostic marker for haematogenous metastasis.

Author affiliations
[1]College of Pharmacy, Jinan University, Guangzhou, Guangdong, China
[2]Department of General Surgery, Jinan University First Affiliated Hospital, Guangzhou, Guangdong, China
[3]Chinese Academy of Sciences Shenzhen Institutes of Advanced Technology, Shenzhen, Guangdong, China
[4]College of Life Science and Technology, Jinan University, Guangzhou, Guangdong, China
[5]Guangdong Key Laboratory of Nasopharyngeal Carcinoma Diagnosis and Therapy, Sun Yat-sen University Cancer Center, Guangzhou, Guangdong, China
[6]Department of Coloproctology & Guangdong Provincial Key Laboratory of Colorectal and Pelvic Floor Diseases, Sun Yat-sen University Sixth Affiliated Hospital, Guangzhou, Guangdong, China
[7]Department of Pharmacology, School of Medicine, Jinan University, Guangzhou, Guangdong, China
[8]School of Traditional Chinese Medicine, Jinan University, Guangzhou, Guangdong, China
[9]Department of Anatomical and Cellular Pathology, Prince of Wales Hospital, The Chinese University of Hong Kong, Hong Kong
[10]Department of Microbiology, Tumor and Cell Biology, Karolinska Institute, Stockholm, Stockholm, Sweden

Correction notice This article has been corrected since it published Online First. Figure one and the supplementary files have been corrected.

Acknowledgements We are grateful to Guangzhou Genedenovo Biotechnology Co. for assistance with sequencing and/or bioinformatics analysis of single-cell RNA sequencing. We thank Guangzhou LC-Bio Technology Co. for their assistance with RNA-seq and CHIP-seq. We thank GL for assistance with atomic force microscopy and data analysis.

Contributors DZ, WY, MC and YC designed and supervised the experiments, revised the manuscript and were responsible for the drafts of the manuscript, as well as all requested revisions. QQ, MH, LD and PM-KT critically revised the manuscript. MC, XL, JP and TL wrote the manuscript and analysed the data. XL, MC, WY, QM, ZZ and SQ performed experiments. JP, DH, RD and YZ collected human colorectal cancer tissues, reviewed the pathological sections and assessed preclinical and clinical samples. YG and WZ directed the second-harmonic generation and two-photon-excited fluorescence experiments. HL directed the Young modulus measurements.

Funding This work was supported by the National Natural Science Foundation of China (81973340, 81803566, 81973341, 81773758 and U1801287), Local Innovative and Research Teams Project of Guangdong Pearl River Talents Program (2017BT01Y036), Natural Science Foundation of Guangdong Province (2019A1515010144, 2019A1515110543, A1515110543, 2021A1515110242, 2019A1515011934 and 2020A1515010071), Ministry of Science and Technology of China (2018ZX09711001-008-008), National High-Level Personnel of Special Support Program (Zhang Dongmei), National Key R&D Program of China (2017YFC1703800), Key-Area Research and Development Program of Guangdong Province (2020B1111110004, 2021B1111110004), Science and Technology Program of Guangzhou (202002030010), Guangdong Province Key Laboratory of Pharmacodynamic Constituents of TCM and New Drugs Research, College of Pharmacy (2020B1212060076), Science and Technology Projects in Guangzhou (202102070001), Young S&T Talent Training ProgramProgramme of Guangdong Provincial Association for S&T, China (SKXRC202216).

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, conduct, reporting or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available in a public, open access repository. scRNA-seq data have been deposited in NCBI's Gene Expression Omnibus and are accessible through the GEO accession number GSE199726. RNA-seq and ChIP-seq data were deposited in NCBI's Gene Expression Omnibus and are accessible through GEO Series accession numbers GSE200064 and GSE200065, respectively.

ORCID iDs

Yihai Cao http://orcid.org/0000-0003-1308-0065
Dongmei Zhang http://orcid.org/0000-0003-2378-2458

REFERENCES

1 Siegel RL, Miller KD, Fuchs HE, *et al*. Cancer statistics, 2021. *CA Cancer J Clin* 2021;71:7–33.
2 Enquist IB, Good Z, Jubb AM, *et al*. Lymph node-independent liver metastasis in a model of metastatic colorectal cancer. *Nat Commun* 2014;5:3530.
3 Welch DR, Hurst DR. Defining the hallmarks of metastasis. *Cancer Res* 2019;79:3011–27.
4 Lambert AW, Pattabiraman DR, Weinberg RA. Emerging biological principles of metastasis. *Cell* 2017;168:670–91.
5 Reymond N, d'Água BB, Ridley AJ. Crossing the endothelial barrier during metastasis. *Nat Rev Cancer* 2013;13:858–70.
6 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.
7 Armulik A, Genové G, Betsholtz C. Pericytes: developmental, physiological, and pathological perspectives, problems, and promises. *Dev Cell* 2011;21:193–215.
8 Cooke VG, LeBleu VS, Keskin D, *et al*. Pericyte depletion results in hypoxia-associated epithelial-to-mesenchymal transition and metastasis mediated by Met signaling pathway. *Cancer Cell* 2012;21:66–81.
9 Xian X, Håkansson J, Ståhlberg A, *et al*. Pericytes limit tumor cell metastasis. *J Clin Invest* 2006;116:642–51.
10 Keskin D, Kim J, Cooke VG, *et al*. Targeting vascular pericytes in hypoxic tumors increases lung metastasis via angiopoietin-2. *Cell Rep* 2015;10:1066–81.
11 Hosaka K, Yang Y, Seki T, *et al*. Pericyte-fibroblast transition promotes tumor growth and metastasis. *Proc Natl Acad Sci U S A* 2016;113:E5618–27.
12 Yang Y, Andersson P, Hosaka K, *et al*. The PDGF-BB-SOX7 axis-modulated IL-33 in pericytes and stromal cells promotes metastasis through tumour-associated macrophages. *Nat Commun* 2016;7:11385.
13 Murgai M, Ju W, Eason M, *et al*. Klf4-Dependent perivascular cell plasticity mediates pre-metastatic niche formation and metastasis. *Nat Med* 2017;23:1176–90.
14 Sinha D, Chong L, George J, *et al*. Pericytes promote malignant ovarian cancer progression in mice and predict poor prognosis in serous ovarian cancer patients. *Clin Cancer Res* 2016;22:1813–24.
15 Viski C, König C, Kijewska M, *et al*. Endosialin-Expressing pericytes promote metastatic dissemination. *Cancer Res* 2016;76:5313–25.
16 Lyle LT, Lockman PR, Adkins CE, *et al*. Alterations in pericyte subpopulations are associated with elevated blood-tumor barrier permeability in experimental brain metastasis of breast cancer. *Clin Cancer Res* 2016;22:5287–99.
17 Ao X, Ding W, Zhang Y, *et al*. Tcf21: a critical transcription factor in health and cancer. *J Mol Med* 2020;98:1055–68.
18 Nagao M, Lyu Q, Zhao Q, *et al*. Coronary Disease-Associated Gene *TCF21* Inhibits Smooth Muscle Cell Differentiation by Blocking the Myocardin-Serum Response Factor Pathway. *Circ Res* 2020;126:517–29.
19 Wirka RC, Wagh D, Paik DT, *et al*. Atheroprotective roles of smooth muscle cell phenotypic modulation and the TCF21 disease gene as revealed by single-cell analysis. *Nat Med* 2019;25:1280–9.
20 Akama T, Chun T-H. Transcription factor 21 (TCF21) promotes proinflammatory interleukin 6 expression and extracellular matrix remodeling in visceral adipose stem cells. *J Biol Chem* 2018;293:6603–10.
21 Suo S, Zhu Q, Saadatpour A, *et al*. Revealing the critical regulators of cell identity in the mouse cell atlas. *Cell Rep* 2018;25:1436–45.
22 Winkler J, Abisoye-Ogunniyan A, Metcalf KJ, *et al*. Concepts of extracellular matrix remodelling in tumour progression and metastasis. *Nat Commun* 2020;11:5120.
23 Shen Y, Wang X, Lu J, *et al*. Reduction of liver metastasis stiffness improves response to bevacizumab in metastatic colorectal cancer. *Cancer Cell* 2020;37:800–17.
24 Reuten R, Zendehroud S, Nicolau M, *et al*. Basement membrane stiffness determines metastases formation. *Nat Mater* 2021;20:892–903.
25 Koikawa K, Ohuchida K, Ando Y, *et al*. Basement membrane destruction by pancreatic stellate cells leads to local invasion in pancreatic ductal adenocarcinoma. *Cancer Lett* 2018;425:65–77.
26 Provenzano PP, Eliceiri KW, Campbell JM, *et al*. Collagen reorganization at the tumor-stromal interface facilitates local invasion. *BMC Med* 2006;4:38.
27 Kalluri R. Assembly and role in tumour angiogenesis. *Nat Rev Cancer* 2003;3:422–33.
28 Paszek MJ, Zahir N, Johnson KR, *et al*. Tensional homeostasis and the malignant phenotype. *Cancer Cell* 2005;8:241–54.
29 Smith LT, Lin M, Brena RM, *et al*. Epigenetic regulation of the tumor suppressor gene TCF21 on 6q23-q24 in lung and head and neck cancer. *Proc Natl Acad Sci U S A* 2006;103:982–7.
30 Chen B, Zeng C, Ye Y, *et al*. Promoter methylation of TCF21 may repress autophagy in the progression of lung cancer. *J Cell Commun Signal* 2018;12:423–32.
31 Dai Y, Duan H, Duan C, *et al*. Tcf21 functions as a tumor suppressor in colorectal cancer through inactivation of PI3K/Akt signaling. *Onco Targets Ther* 2017;10:1603–11.
32 Wu H, Zhou J, Zeng C, *et al*. Curcumin increases exosomal TCF21 thus suppressing exosome-induced lung cancer. *Oncotarget* 2016;7:87081–90.
33 Marusyk A, Janiszewska M, Polyak K. Intratumor heterogeneity: the Rosetta stone of therapy resistance. *Cancer Cell* 2020;37:471–84.
34 Kim C, Gao R, Sei E, *et al*. Chemoresistance evolution in triple-negative breast cancer delineated by single-cell sequencing. *Cell* 2018;173:879–93.
35 Zheng C, Zheng L, Yoo J-K, *et al*. Landscape of infiltrating T cells in liver cancer revealed by single-cell sequencing. *Cell* 2017;169:1342–56.
36 Muhl L, Genové G, Leptidis S, *et al*. Single-Cell analysis uncovers fibroblast heterogeneity and criteria for fibroblast and mural cell identification and discrimination. *Nat Commun* 2020;11:3953.
37 Goveia J, Rohlenova K, Taverna F, *et al*. An integrated gene expression landscape profiling approach to identify lung tumor endothelial cell heterogeneity and angiogenic candidates. *Cancer Cell* 2020;37:21–36.
38 Zhao Q, Eichten A, Parveen A, *et al*. Single-Cell transcriptome analyses reveal endothelial cell heterogeneity in tumors and changes following antiangiogenic treatment. *Cancer Res* 2018;78:2370–82.
39 Kinchen J, Chen HH, Parikh K. Data from: composition of the colonic mesenchyme and the nature of its plasticity in inflammatory bowel disease. *Gene Expression Omnibus* 2018 https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc= GSE114374
40 Kim J-E, Fei L, Yin W-C, *et al*. Single cell and genetic analyses reveal conserved populations and signaling mechanisms of gastrointestinal stromal niches. *Nat Commun* 2020;11:334.
41 Teuwen L-A, De Rooij LPMH, Cuypers A, *et al*. Tumor vessel co-option probed by single-cell analysis. *Cell Rep* 2021;35:109253.
42 Rao M, Oh K, Moffitt R. Single-Cell RNA-seq analysis of Metastasis-Primary tumor dissimilarity in a patient with gastrointestinal neuroendocrine cancer. *Gene Expression Omnibus* 2020 https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi
43 Gaggioli C, Hooper S, Hidalgo-Carcedo C, *et al*. Fibroblast-led collective invasion of carcinoma cells with differing roles for RhoGTPases in leading and following cells. *Nat Cell Biol* 2007;9:1392–400.
44 Hussain A, Voisin V, Poon S, *et al*. Distinct fibroblast functional states drive clinical outcomes in ovarian cancer and are regulated by TCF21. *J Exp Med* 2020;217:e20191094.
45 Lechertier T, Reynolds LE, Kim H, *et al*. Pericyte FAK negatively regulates Gas6/Axl signalling to suppress tumour angiogenesis and tumour growth. *Nat Commun* 2020;11:2810.
46 Bartolini A, Cardaci S, Lamba S, *et al*. BCAM and Lama5 mediate the recognition between tumor cells and the endothelium in the metastatic spreading of KRAS-mutant colorectal cancer. *Clin Cancer Res* 2016;22:4923–33.
47 Zhao S, Mi Y, Zheng B, *et al*. Highly-metastatic colorectal cancer cell released miR-181a-5p-rich extracellular vesicles promote liver metastasis by activating hepatic stellate cells and remodelling the tumour microenvironment. *J Extracell Vesicles* 2022;11:e12186.

Supplemental materials for

Novel TCF21^{high} pericyte subpopulation promotes colorectal cancer metastasis by remodeling perivascular matrix

Supplemental methods

Cell lines and cell culture

Human CRC cell lines HCT116, DLD-1, RKO, SW480, SW620, Caco-2 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

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*$^{\mathit{tm1(CAG\text{-}LSL\text{-}Cas9\text{-}tdTomato)}}$/ Gpt; Cat. T002249), *Cspg4*-CreERT2 mice (B6/JGpt-*Cspg4*$^{\mathit{em1Cin(CreERT2\text{-}P2A)}/\mathrm{Gpt}}$; T006187), and *Tcf21*-flox mice (B6/JGpt-*Tcf21*$^{\mathit{em1Cflox}}$/Gpt; T013083) were obtained from GemPharmatech Co., Ltd (Nanjing, Jiangsu, China). Pericyte lineage tracing mice (PC$^{\mathrm{lin}}$) were generated by crossing mice carrying a tamoxifen-inducible Cre recombinase driven by the pericyte-specific *Cspg4* promoter (Tg$^{\mathrm{Cspg4\text{-}CreERT2}}$) with mice carrying a Cre-responsive reporter gene (tandem dimer Tomato (*tdT*)) inserted at the *ROSA26* locus (*ROSA*$^{\mathrm{tdT/+}}$). The PC$^{\mathrm{lin}}$ mice were further crossed with mice harboring both *Tcf21* alleles flanked by LoxP sites (*Tcf21*$^{\mathit{flox/flox}}$) to generate tamoxifen-inducible *Cspg4*-driven pericyte-specific *Tcf21* knockout mice (PC$^{\mathrm{lin\text{-}KO}}$). 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 ≤ 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 (1×10^5) 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 (1×10^5) 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 (4×10^5) were premixed with TPCs (1.6×10^6) 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. F6434, 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* for 48 h and selected by puromycin (2 mg/mL). Detailed information of *TCF21* lentivirus and vector was listed as follows: TCF21 (NM_004460) Human Untagged Clone (Cat. SC117372, Origene. Rockwell, MD, USA), Cloning vector PCMV6-XL5 (Cat. PCMV6XL5, 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 (2×10^5) 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^{TM} X-100, blocked with QuickBlock^{TM} 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^{TM} 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.1896649, 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 (3×10^4 HCT116 cells or 2×10^4 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 (5×10^3) mixed with TPCs (2.5×10^4) 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 5×10^6 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 (5×10^3) 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 (2×10^4) were seeded in a 96-well plate for 2 h. Then, the media was removed, and cells were washed with

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 (2×10^4) 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 (5 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 with Donkey anti-Goat IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor™ 488 (RRID: AB_2534102, Invitrogen) and immersed in PBS for SHG/TPEF microscopy (Nikon, Tokyo, Japan). TPEF was utilized for visualization of tumor vessels stained by CD31 (red) and SHG was used to visualize collagen structure (green) at 790 nm excitation light.

Atomic Force Microscopy (AFM) measurement

For collagen organization, coatings of isotropic collagen I and AFM detection were performed according to previous report[13]. *In vitro* fibrillogenesis of collagen-I (Cat. 08-115, Merck; 4.4 mg/mL) was initiated by diluting the collagen I solution tenfold in fibrillogenic buffer (50 mM glycine, 200 mM KCl, pH 9.0). After mixture for 10 min at room temperature, 60 µL of the diluted solution were added to the 22 mm-silicide coverslips and incubated overnight at 37 °C. The next day, coverslips were washed with PBS twice and plated in 12-well plate. Then, TPCs (1×10^4) were seeded on the Collagen I-coated coverslip and cultured for 5 days. AFM Imaging was performed with a NanoWizard II AFM (JPK-Instruments, Berlin, Germany) mounted on an inverted microscope (Axiovert 200, Zeiss, Jena, Germany). Scanning of samples was performed at a scan rate of 0.25 Hz and five fields were recorded for each sample. For perivascular stiffness measurements, AFM was performed according to modified published procedures[14]. Tissues were acquired and embedded within OCT. Then, the frozen tissue blocks were sectioned at a thickness of 20 µm and immersed in proteinase inhibitor-contained PBS at room temperature. The next day, tumor sections were applied for AFM quantification of Young's modulus (Bruker, USA). Briefly, silicon nitride cantilevers with a spring constant of 0.15 N/m were attached by a borosilicate glass 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 (2×10^4) supplemented with 100 μL ECM were seeded in the upper chamber and incubated overnight. The next day, ECM was removed, and TPCs (5×10^5) 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+ ν) to allow comparison to other published work. For invasion assay, TCF21-overexpressing TPCs (5×10^5) 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 (3×10^4) 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.cg*i*.

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.

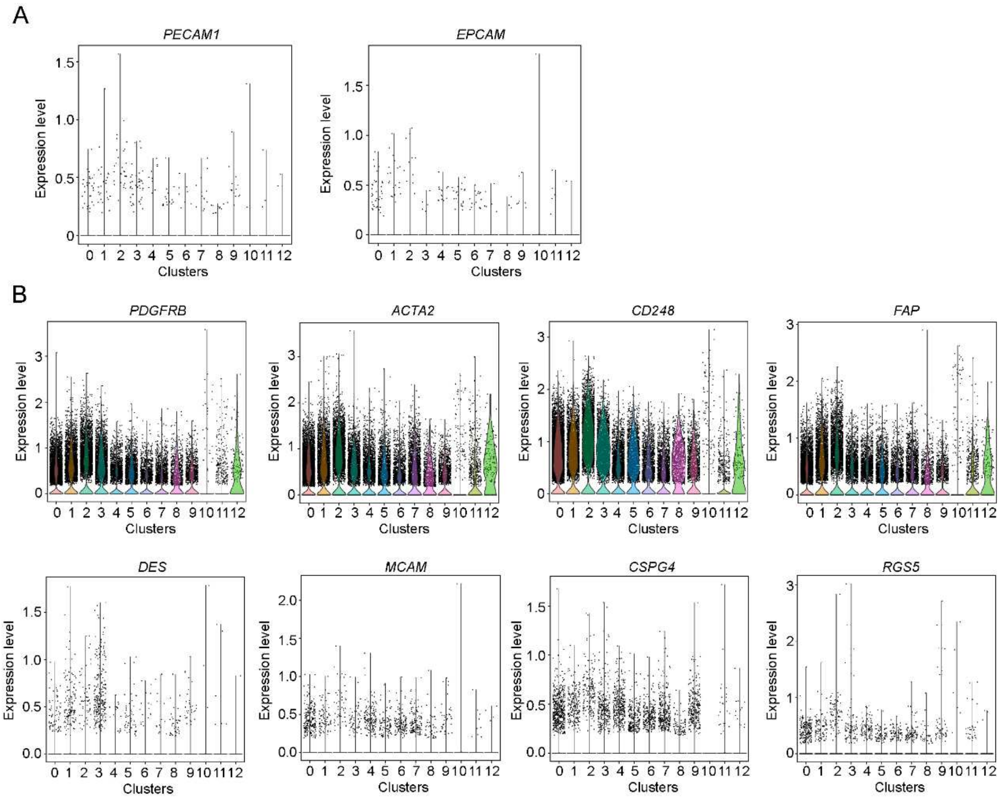
13. Kirmse R, Otto H, Ludwig T. Interdependency of cell adhesion, force generation and extracellular proteolysis in matrix remodeling. *J Cell Sci* 2011; 124: 1857-66.

14. Laklai H, Miroshnikova YA, Pickup MW, *et al*. Genotype tunes pancreatic ductal adenocarcinoma tissue tension to induce matricellular fibrosis and tumor progression.

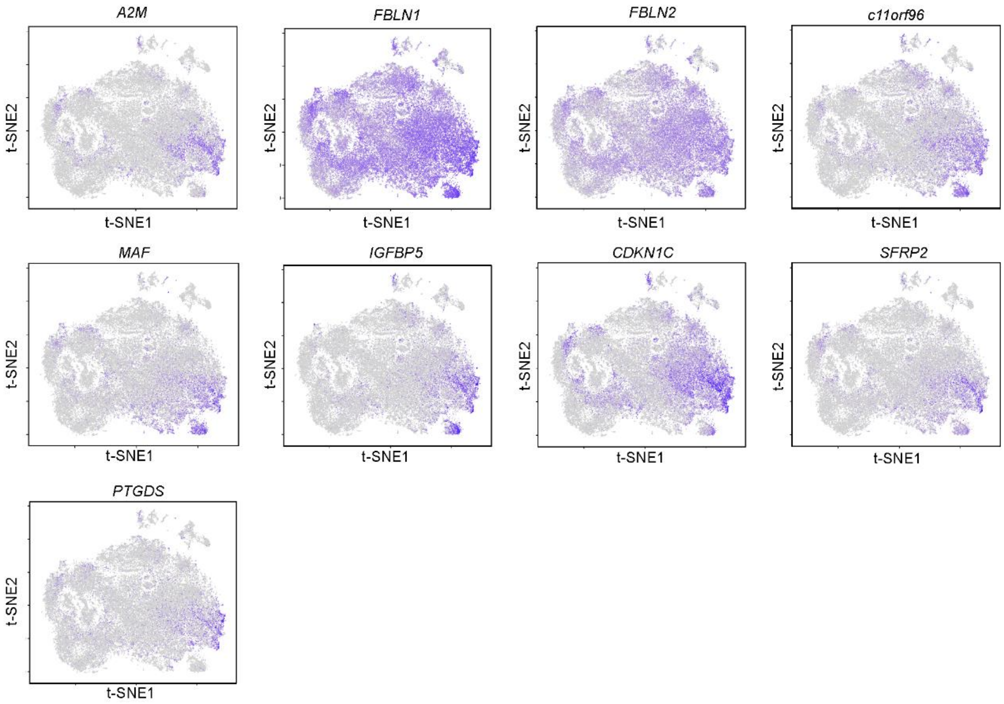
Nat Med 2016; 22: 497-505.

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

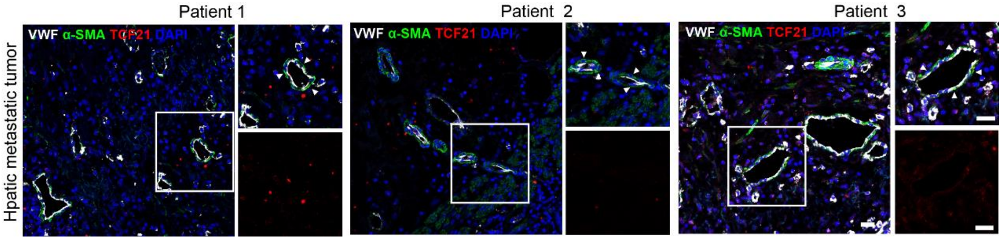
Supplemental figures and figure legends



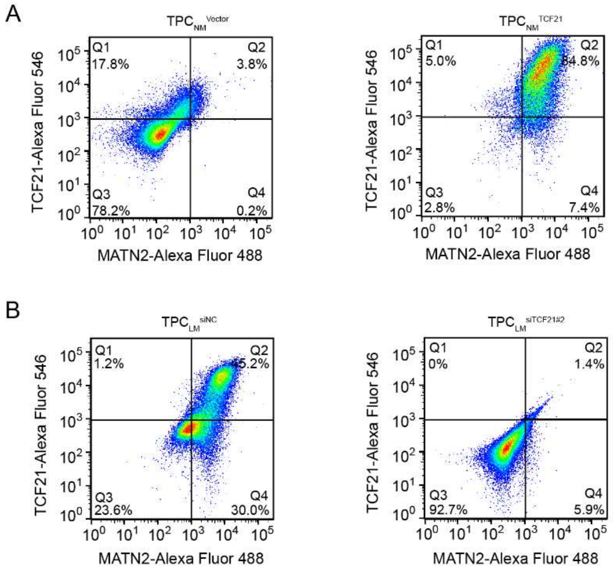
Supplemental Figure 1. Transcriptomic characterization of 13 subsets of TPCs. (A) Gene expression profiles of *PECAM1* and *EPCAM* in distinct subsets of TPCs. **(B)** Gene expression profiles of *PDGFRB*, *ACTA2*, *CD248*, *FAP*, *DES*, *MCAM*, *CSPG4,* and *RGS5* in distinct subsets of TPCs.



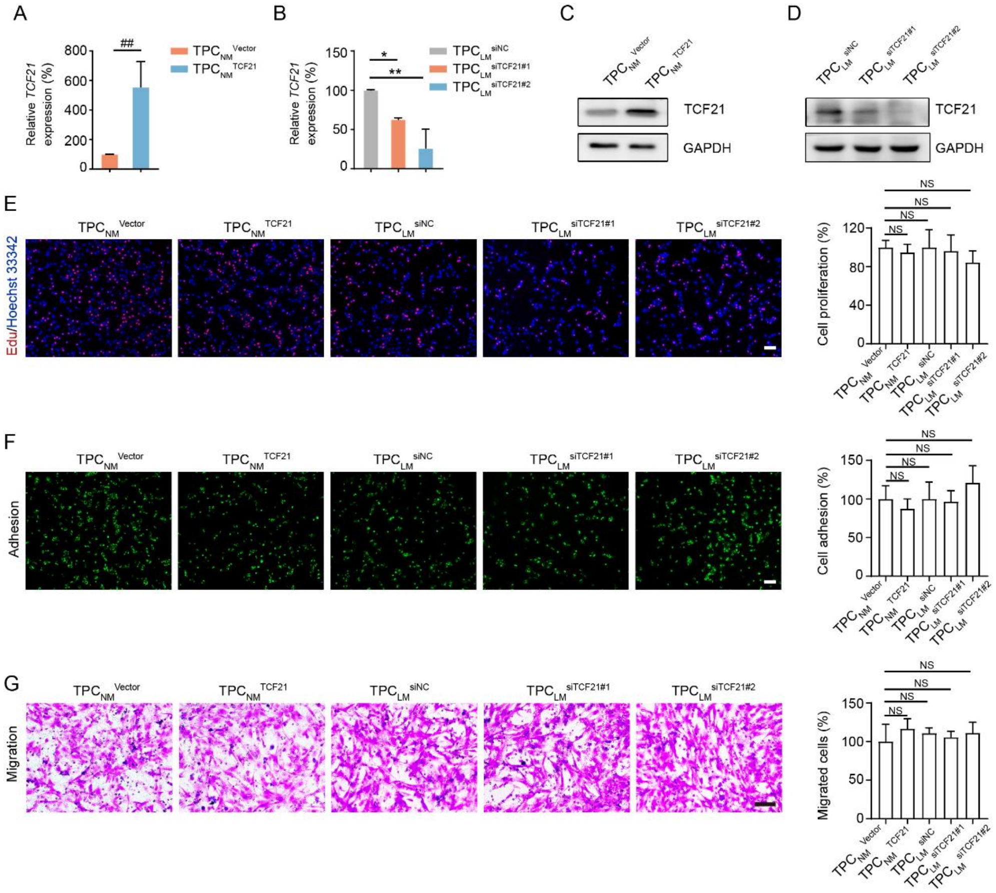
Supplemental Figure 2. Distribution of the matrix-pericyte-related genes in all subsets of TPCs. t-SNE visualization of gene distribution of *A2M*, *FBLN1*, *FBLN2*, *c11orf96*, *MAF*, *IGFBP5*, *CDKN1C*, *SFRP2* and *PTGDS* in all subsets of TPCs.



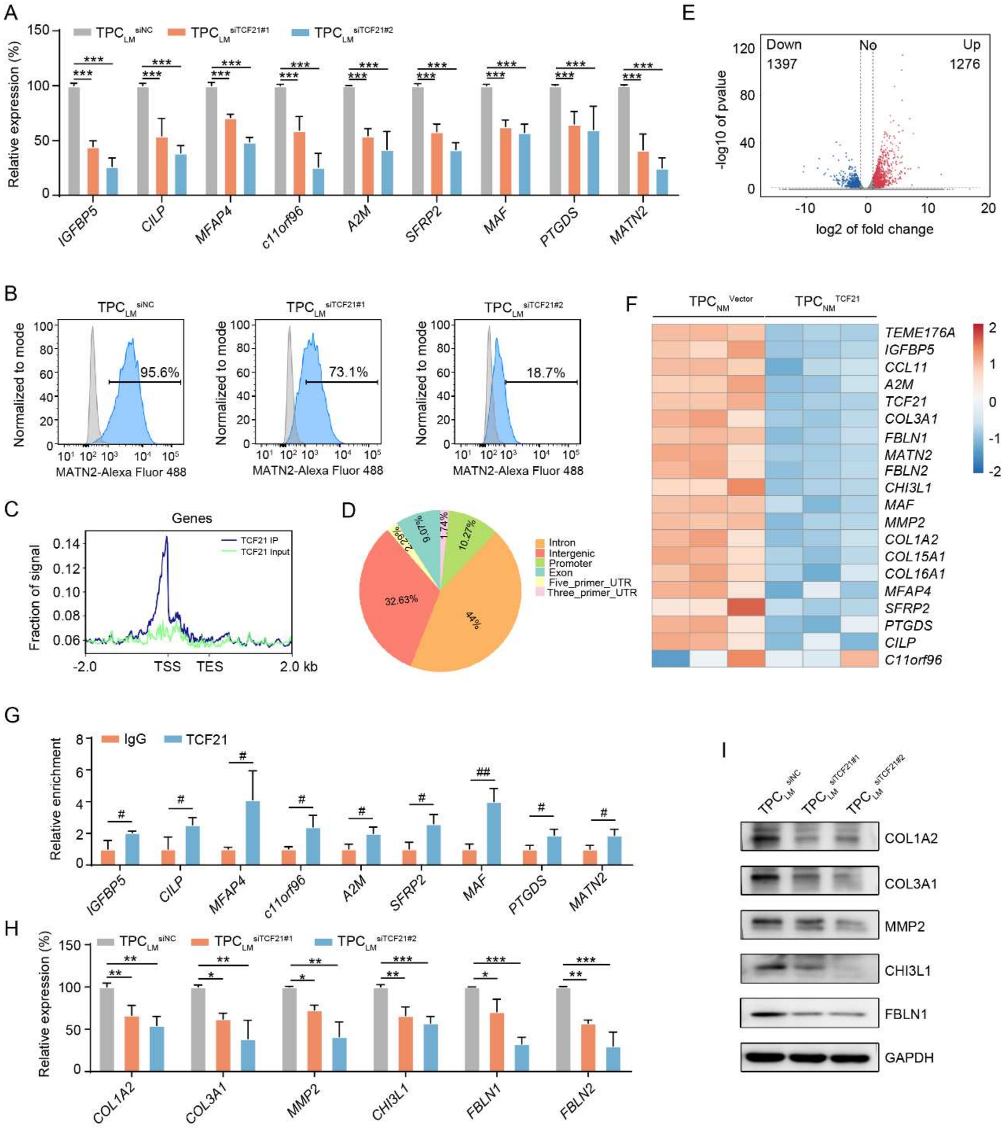
Supplemental Figure 3. Determination of TCF21 in TPCs in the hepatic metastatic tumors from CRC patients. Representative images of TCF21 staining (red) in TPCs (αSMA$^+$, green) in the hepatic metastatic tumors from CRC patients (n = 20). Scale bar, 20 μm.



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).

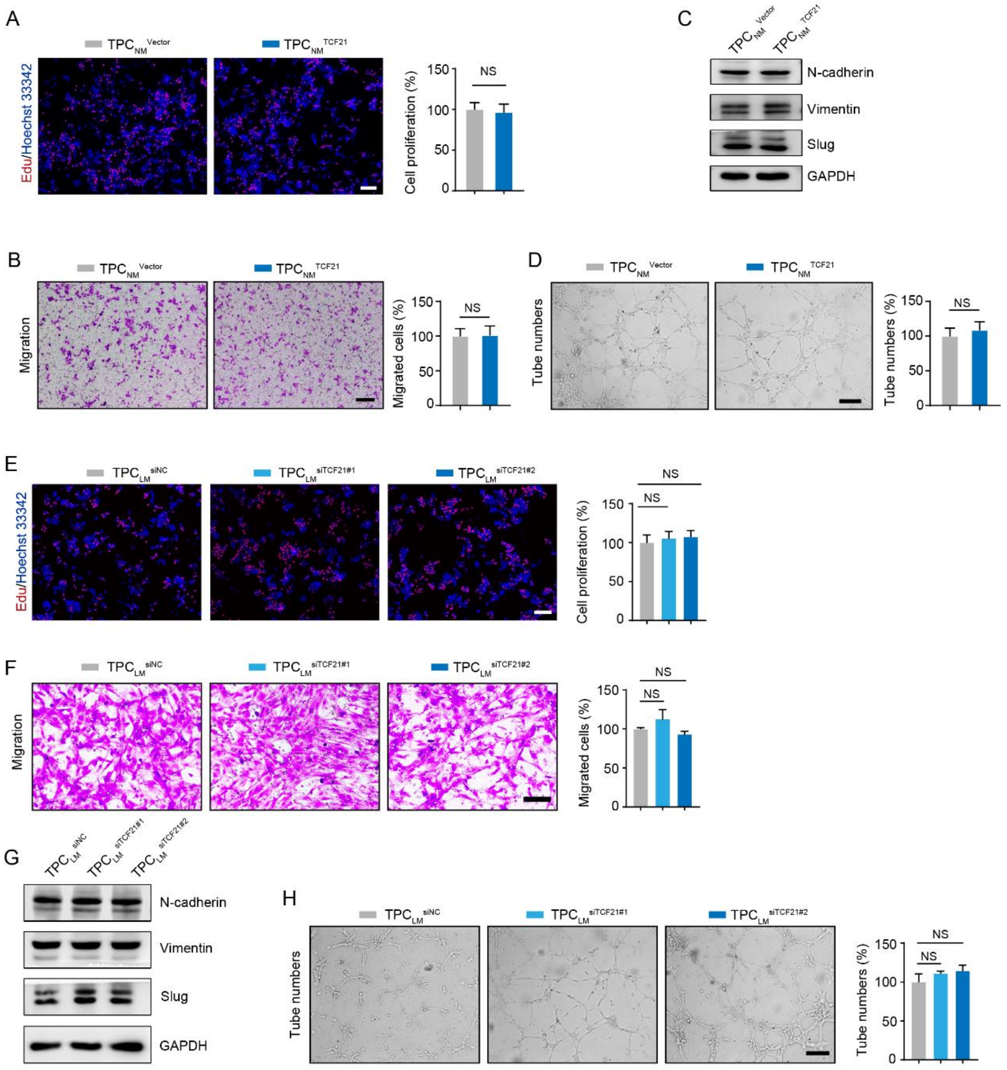


Supplemental Figure 5. Effects of TCF21 on the proliferation, adhesion, and migration of TPCs. (A, B) RT-qPCR analysis of *TCF21* mRNA levels in TCF21-overexpressing **(A)** or -knockdown **(B)** TPCs (n = 3). **(C, D)** Western blotting analysis of TCF21 in TCF21-overexpressing **(C)** or -knockdown **(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 TPCs adhesive to fibronectin (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.

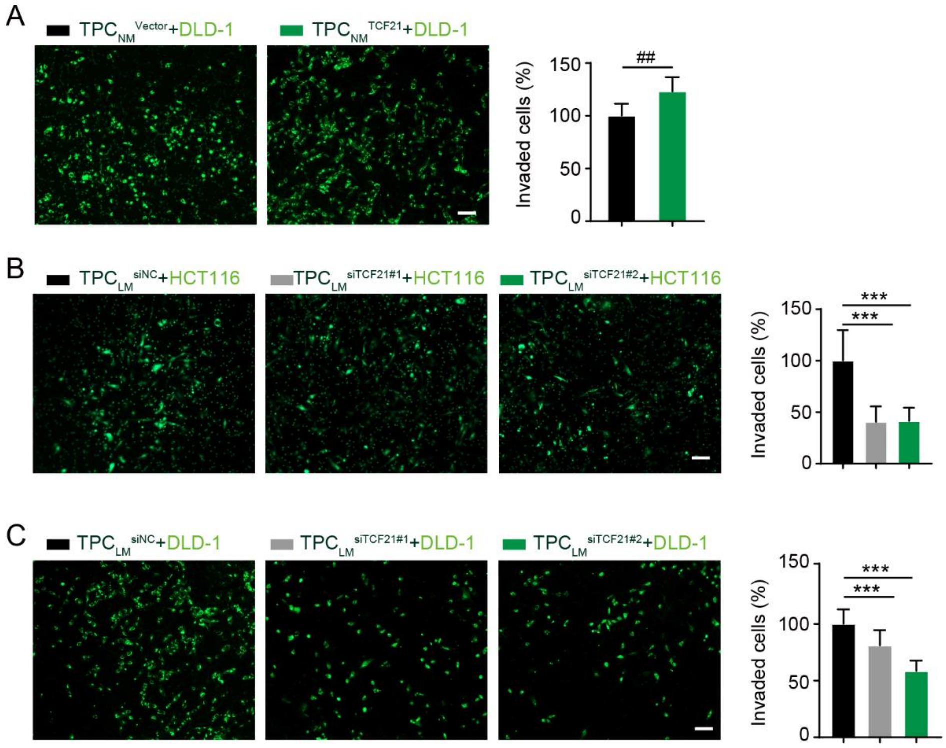


Supplemental Figure 6. TCF21 stimulates the generation of matrix-pericytes and induces ECM remodeling. (A) RT-qPCR analysis of matrix-pericyte-specific genes in TPC$_{LM}$ transfected with siRNA targeting TCF21 (TPC$_{LM}$^{siTCF21}) or negative control (TPC$_{LM}$^{siNC}) (n = 3). **(B)** FCM analysis 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

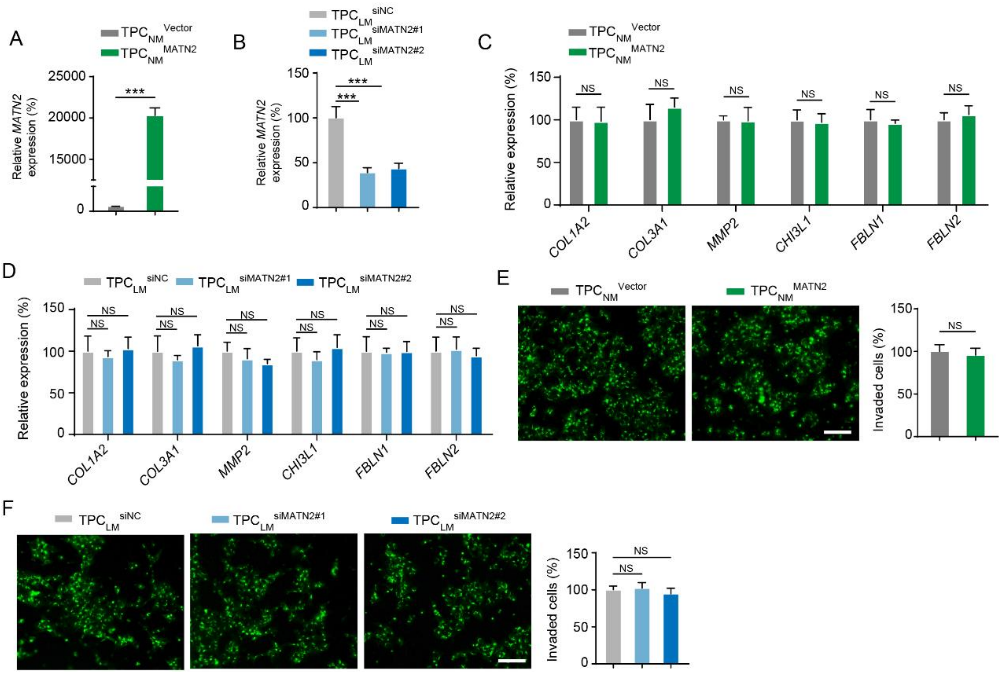
= 3); Red dots represent the up-regulated genes and blue dots represent the down-regulated genes. **(F)** Heat maps of the differentially expressed genes between TPC$_{NM}^{Vector}$ and TPC$_{NM}^{TCF21}$ (n = 3). **(G)** ChIP-qPCR analysis of TCF21 binding at the promoter of indicated genes in TPC$_{NM}^{TCF21}$ (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.



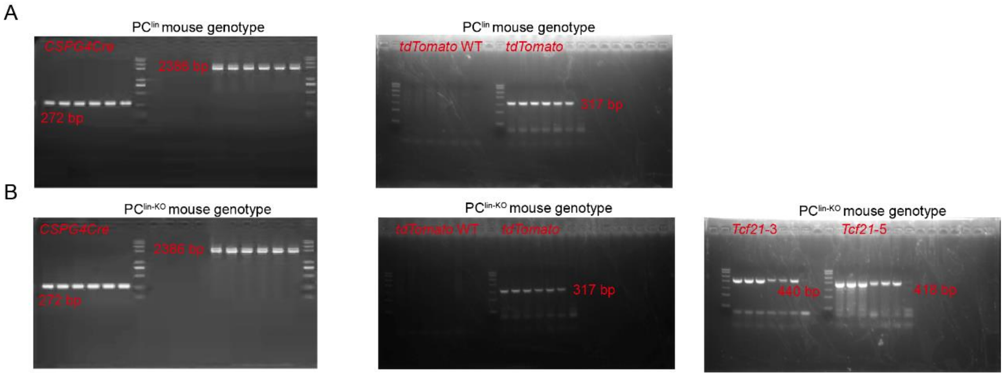
Supplemental Figure 7. TCF21 in TPCs has negligible effects on cell migration and angiogenesis. **(A)** EdU assay for the proliferation of HCT116 cells primed with conditioned medium from TPC_{NM}^{Vector} and TPC_{NM}^{TCF21} for 48 h (n = 3)**.** Scale bar, 100 µm. **(B)** Transwell assay for cell 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**).



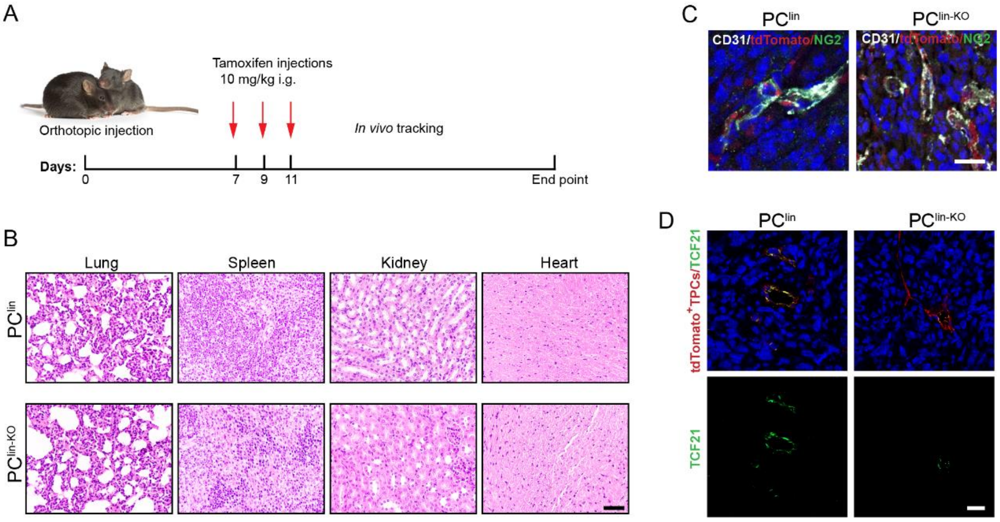
Supplemental Figure 8. TCF21 in TPCs promotes invasion of CRC cells. (A) Representative images and quantification of invaded DLD-1 cells (green). DLD-1 cells mixed with TPC$_{NM}$^{Vector} or TPC$_{NM}$^{TCF21} were seeded into the Matrigel-coated transwell. The invaded DLD-1 cells were photographed 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



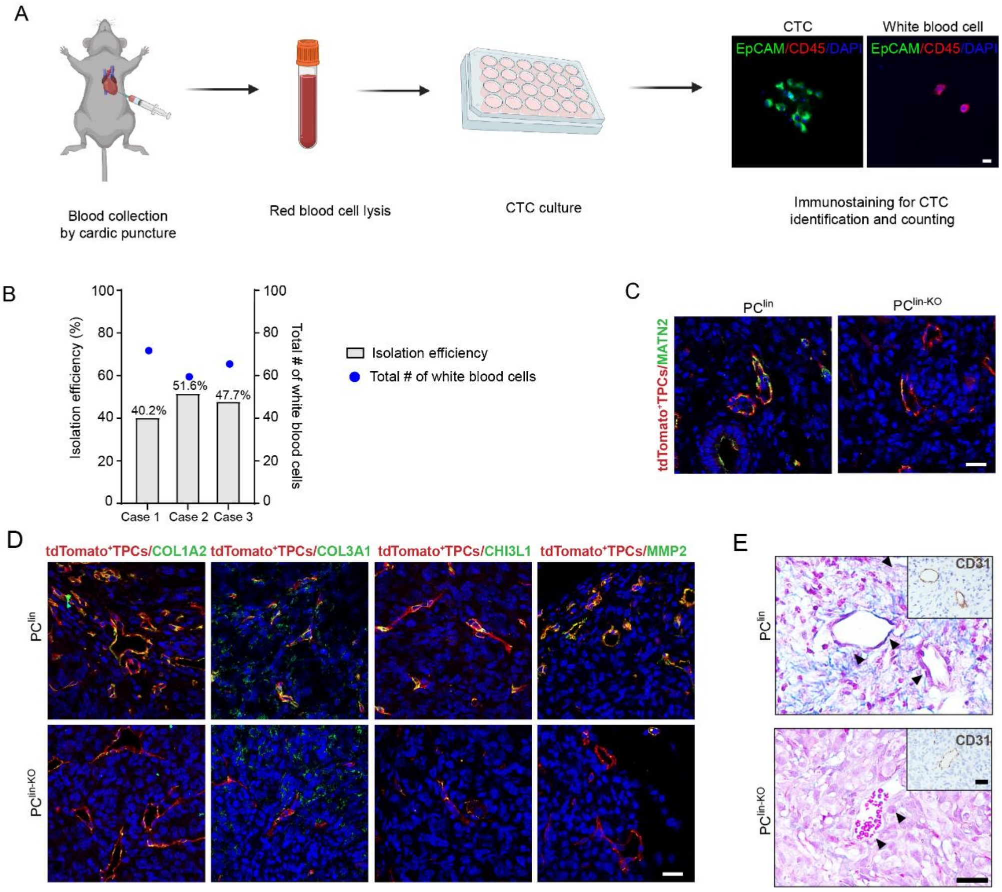
Supplemental Figure 9. MATN2 has negligible effects on ECM remodeling and CRC metastasis. (A, B) RT-qPCR analysis of *MATN2* in MATN2-overexpressing **(A)** or -knockdown (**B)** TPCs (n = 3). **(C)** RT-qPCR analysis of the ECM-related genes in $\text{TPC}_{\text{NM}}^{\text{Vector}}$ and $\text{TPC}_{\text{NM}}^{\text{MATN2}}$ (n = 3). **(D)** RT-qPCR analysis of the indicated genes in $\text{TPC}_{\text{LM}}^{\text{siNC}}$ and $\text{TPC}_{\text{LM}}^{\text{siMATN2}}$ (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**).



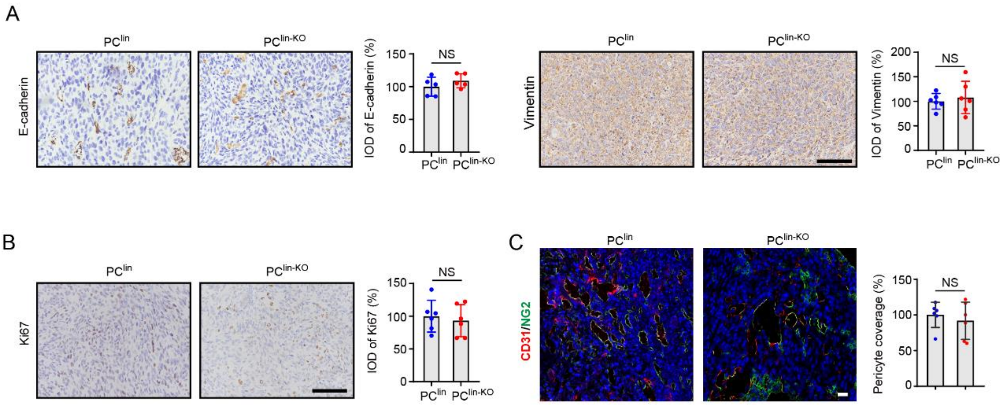
Supplemental Figure 10. Mouse genotyping. (A) PCR analysis of the genotype of PC^{lin} mouse (n = 6). **(B)** PCR analysis of the genotype of PC$^{lin\text{-}KO}$ mice (n = 6). All mice were analyzed by PCR genotyping. PCR analysis of *Cspg4Cre* showed two bands identifying homozygous *Cspg4Cre* at 2386 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.



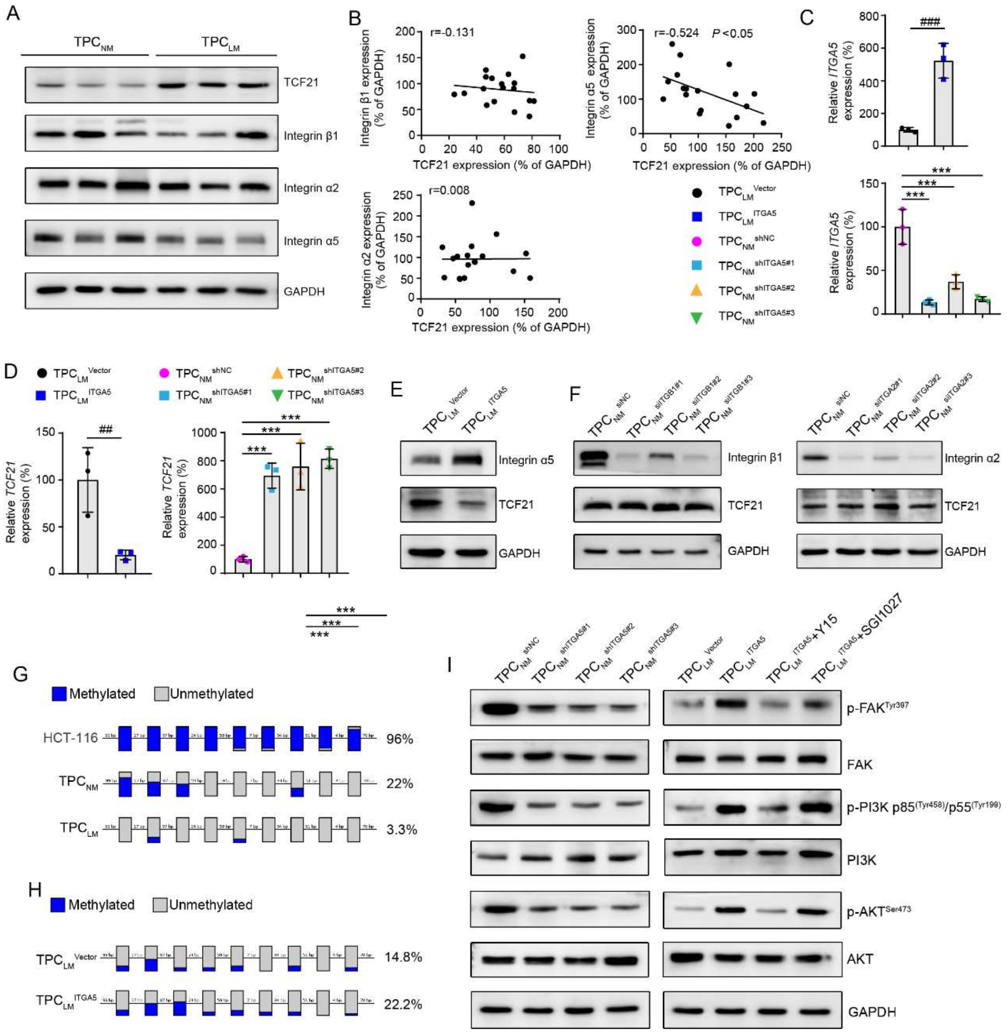
Supplemental Figure 11. Characterization of PC^{lin} and PC$^{lin\text{-}KO}$ mice. (A) Schematic diagram describing the experimental design of *in vivo* experiments. PC^{lin} mice and PC$^{lin\text{-}KO}$ mice were orthotopically injected with MC38-luc-LM3 cells. After 7 days, all mice were treated with tamoxifen (10 mg/kg) through intragastric administration every other day for three times. *In vivo* tracking was performed to detect tumor liver metastasis. **(B)** Representative images of H&E staining of lung, spleen, kidney, heart, and liver derived from PC^{lin} and $PC^{lin\text{-}KO}$ mice (n = 6). Scale bar, 50 μm. **(C)** Immunofluorescence analysis of TPCs (tdTomato) in tumor sections by NG2 (green) staining (n = 6). Scale bar, 20 μm. **(D)** Immunofluorescence analysis of TCF21 (green) in TPCs (tdTomato) from PC^{lin} mice and $PC^{lin\text{-}KO}$ mice (n = 6). Scale bar, 20 μm.



Supplemental Figure 12. Pericyte-specific knockout of *Tcf21* inhibits perivascular ECM remodeling and CRCLM. (A) Schematic diagram of the isolation and 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 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.

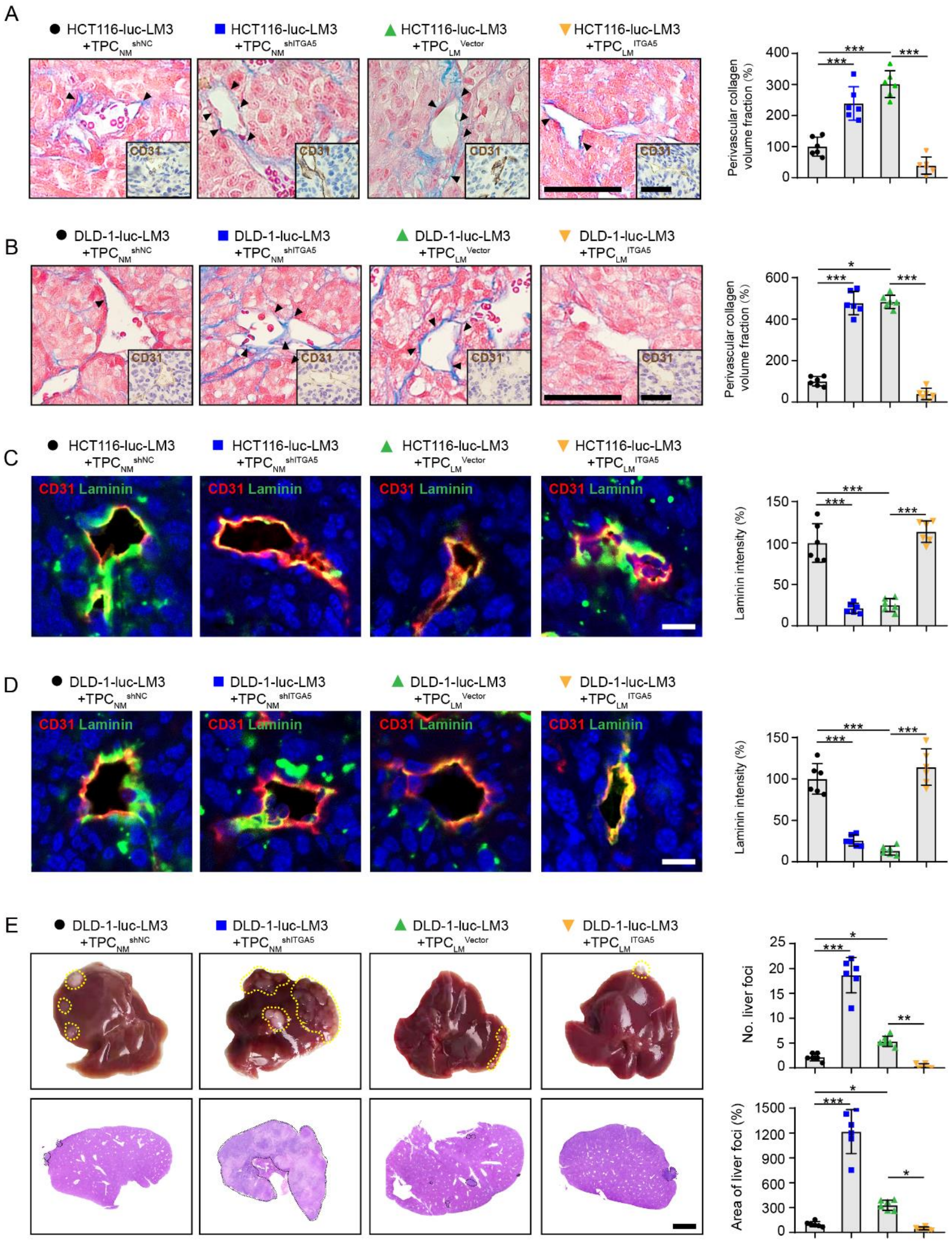


Supplemental Figure 13. Pericyte-specific knockout of *Tcf21* has negligible effects on EMT and proliferation of CRC cells. (A) Immunohistochemical staining and quantification of E-cadherin and vimentin in orthotopic MC38 tumor tissues (n = 6). Scale bar, 50 µm. **(B)** Representative images and quantification 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.

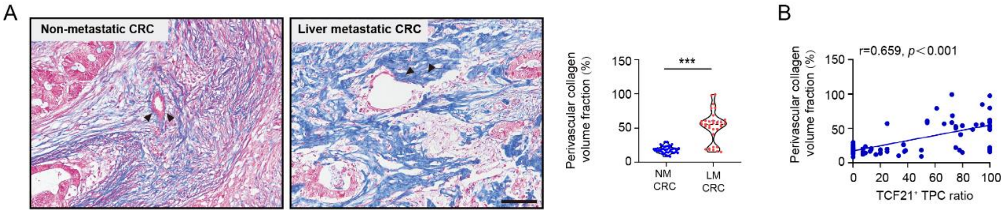


Supplemental Figure 14. Effects of integrin α2 and integrin β1 on TCF21 in TPCs. **(A)** Western blotting analysis of TCF21, integrin β1, integrin α2, and integrin α5 in TPC$_{NM}$ and TPC$_{LM}$ (n = 3). **(B)** Pearson's correlation analysis of the expression of integrin β1, integrin α2 and integrin 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)**

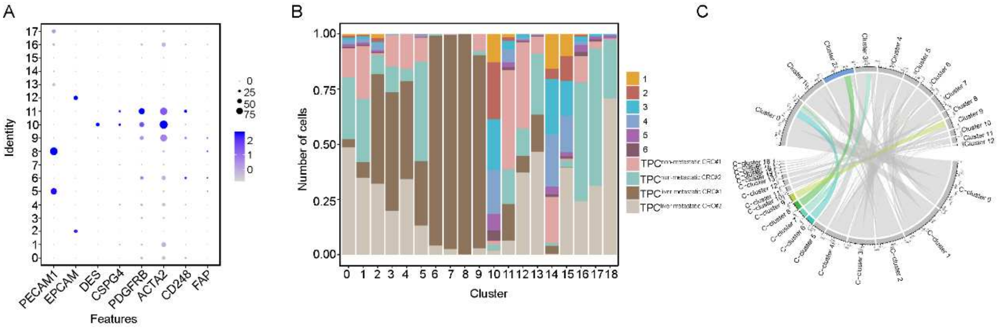
Bisulfite DNA sequencing analysis of *TCF21* promoter region in TPCs or HCT116 cells. Blue and gray circles represent methylated and unmethylated CpGs, respectively. The percentage of total methylated CpGs is given on right of each dataset (n = 3). **(H)** Bisulfite DNA sequencing analysis of *TCF21* 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.



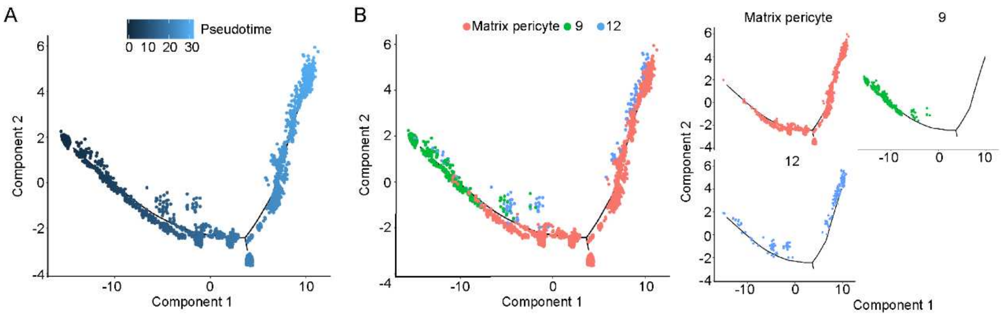
Supplemental Figure 15. Loss of integrin α5 in TPCs promotes perivascular ECM remodeling and CRCLM. (A, B) Representative images of Masson and CD31 staining in primary tumor sections (n = 6). Black arrows indicate the perivascular collagen fibers. Scale bar, 50 μm. The quantification of perivascular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

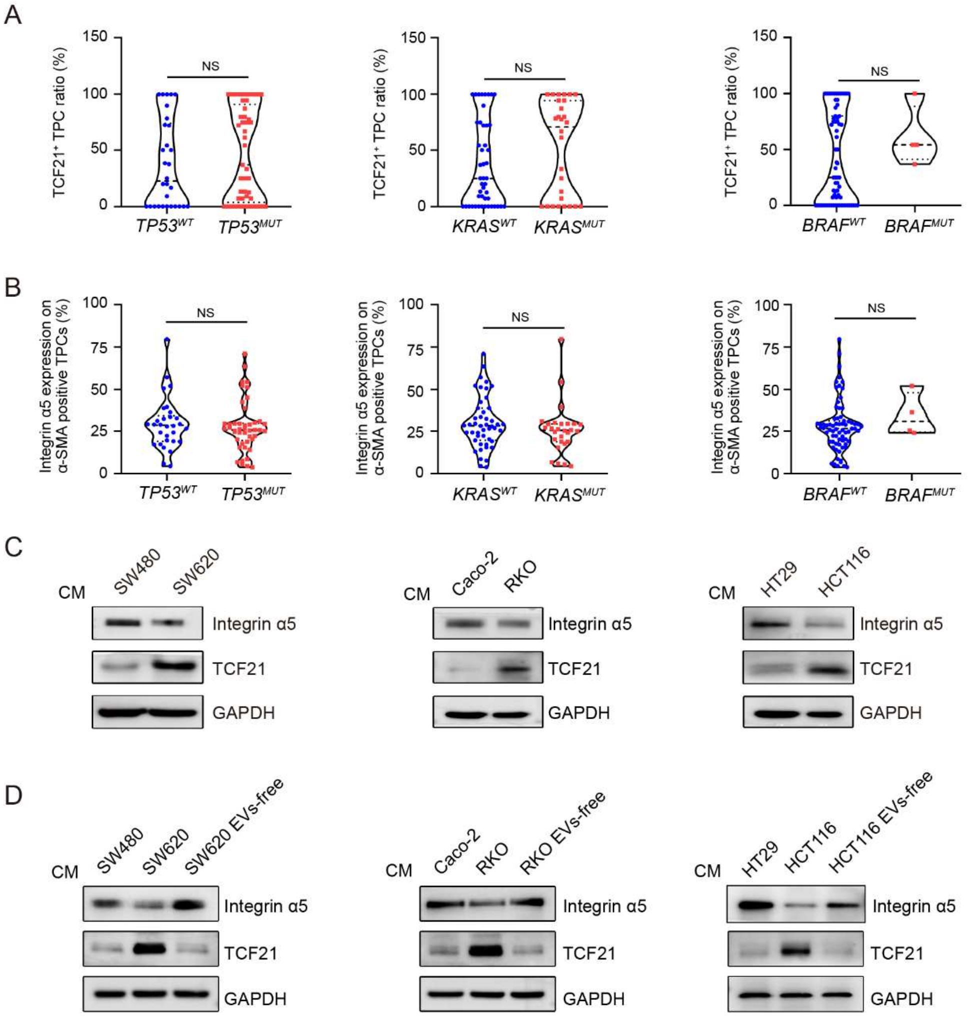


Supplemental Figure 17. Comparison of the scRNA-seq data derived from TPCs and the previous published data. (A) Dot plots for gene expressions in pericytes acquired from previous studies. Raw data of pericyte scRNA-seq were collected from six samples derived from previous studies (GEO accession GSM3140596, 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).



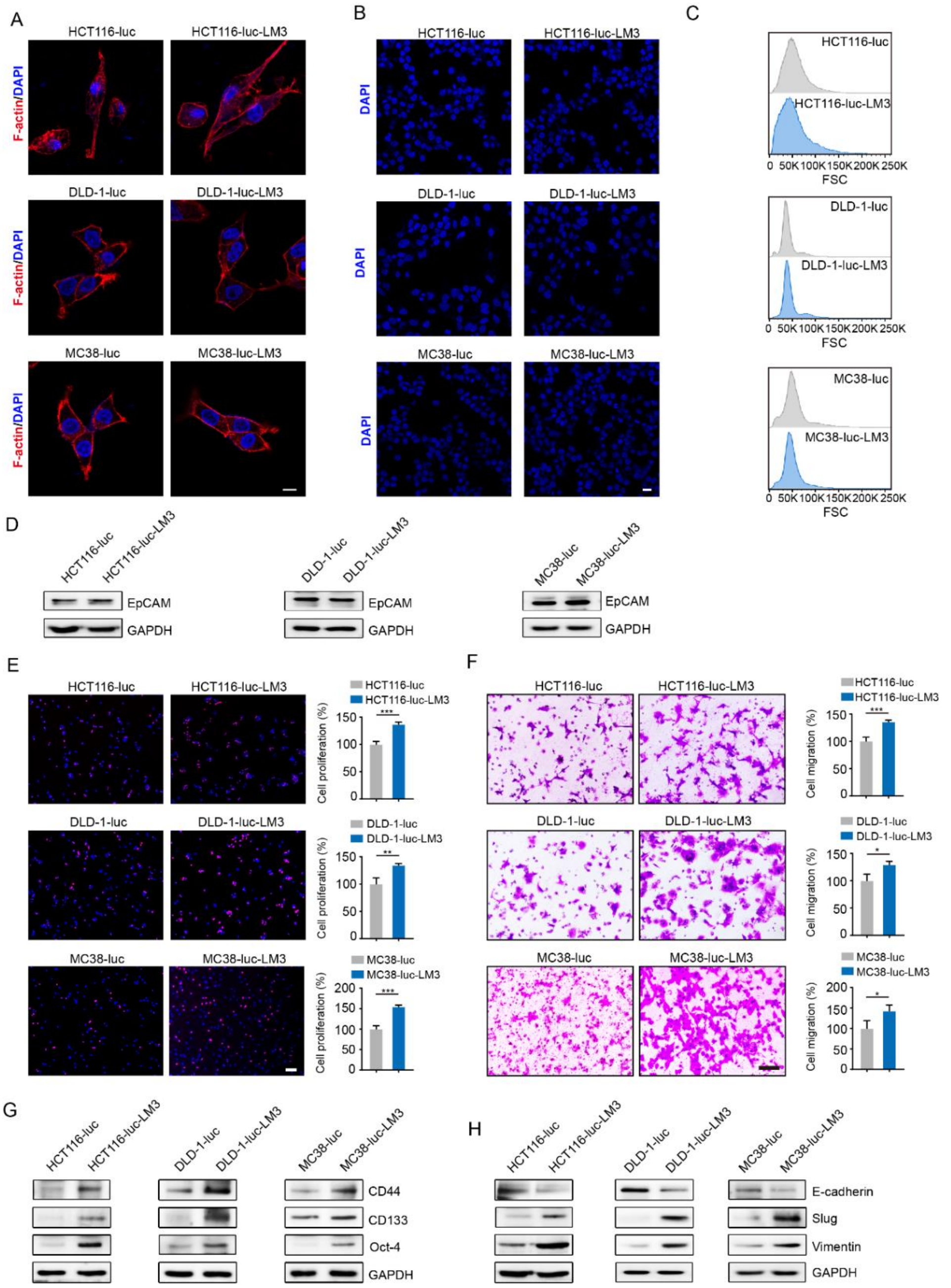
Supplemental Figure 18. Pseudo-time trajectory for dynamic changes in matrix-

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.

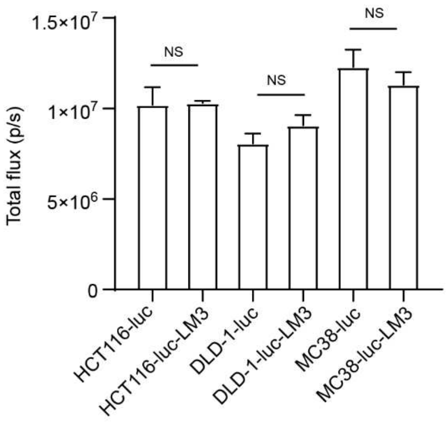


Supplemental Figure 19. Effects of metastatic CRC cells on the expressions of integrin α5 and TCF21 in TPCs. (A) Quantification of TCF21$^+$ TPC ratio in CRC patients with or without *TP53*, *BRAF* and *KRAS* mutation (n = 75). **(B)** Quantification of integrin α5 expression in TPCs derived from CRC 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 weakly- or highly-metastatic CRC cells (n = 3). The EVs-free CM of highly-metastatic CRC cells were generated by centrifugation at 1×10^5 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*.



Supplemental Figure 20. Comparison of parental cells with LM3 cells. (A) Immunofluorescence analysis of cellular morphology (n = 3). Phalloidin-rhodamine was used to identify F-actin. Scale bar, 10 μm. **(B)** Immunofluorescence analysis of nucleus size (n = 3). Cell nucleus was measured after DAPI staining of fixed cells. Scale 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.

Supplemental Table 1. Clinical characteristics of CRC specimens.

Character	Overall population
Gender, number of patients (%)	
Male	33 (44.0)
Female	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)

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 Fl: 418bp	T013083 *Tcf21*-flox

	JS04431-Tcf21-5wt-tR1	GTTTGCTAACTTGCTGCCACACAC		
4	XM003792- Cspg4-TF1	AAATCTAAGCGCGGGTCTGGC	WT:0bp KI:272bp	T006187 *Cspg4*-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

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

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
FBLN2	AB_11153545	Invitrogen
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

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™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

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*$^{tm1(CAG\text{-}LSL\text{-}Cas9\text{-}tdTomato)}$/ Gpt; Cat. T002249), *Cspg4*-CreERT2 mice (B6/JGpt-*Cspg4*$^{em1Cin(CreERT2\text{-}P2A)/\mathrm{Gpt}}$; 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$^{\mathrm{lin}}$) were generated by crossing mice carrying a tamoxifen-inducible Cre recombinase driven by the pericyte-specific *Cspg4* promoter (Tg$^{\mathrm{Cspg4\text{-}CreERT2}}$) with mice carrying a Cre-responsive reporter gene (tandem dimer Tomato (*tdT*)) inserted at the *ROSA26* locus (*ROSA*$^{\mathrm{tdT/+}}$). The PC$^{\mathrm{lin}}$ 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$^{\mathrm{lin\text{-}KO}}$). 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 ≤ 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 (1×10^5) 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 (1×10^5) 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 (4×10^5) were premixed with TPCs (1.6×10^6) 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 (2×10^5) 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^{TM} X-100, blocked with QuickBlock^{TM} 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^{TM} 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 (3×10^4 HCT116 cells or 2×10^4 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 (5×10^3) mixed with TPCs (2.5×10^4) 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 5×10^6 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 (5×10^3) 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 (2×10^4) were seeded in a 96-well plate for 2 h. Then, the media was removed, and cells were washed with

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 (2×10^4) 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 with Donkey anti-Goat IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor™ 488 (RRID: AB_2534102, Invitrogen) and immersed in PBS for SHG/TPEF microscopy (Nikon, Tokyo, Japan). TPEF was utilized for visualization of tumor vessels stained by CD31 (red) and SHG was used to visualize collagen structure (green) at 790 nm excitation light.

Atomic Force Microscopy (AFM) measurement

For collagen organization, coatings of isotropic collagen I and AFM detection were performed according to previous report[13]. *In vitro* fibrillogenesis of collagen-I (Cat. 08-115, Merck; 4.4 mg/mL) was initiated by diluting the collagen I solution tenfold in fibrillogenic buffer (50 mM glycine, 200 mM KCl, pH 9.0). After mixture for 10 min at room temperature, 60 μL of the diluted solution were added to the 22 mm-silicide coverslips and incubated overnight at 37 °C. The next day, coverslips were washed with PBS twice and plated in 12-well plate. Then, TPCs (1×10^4) were seeded on the Collagen I-coated coverslip and cultured for 5 days. AFM Imaging was performed with a NanoWizard II AFM (JPK-Instruments, Berlin, Germany) mounted on an inverted microscope (Axiovert 200, Zeiss, Jena, Germany). Scanning of samples was performed at a scan rate of 0.25 Hz and five fields were recorded for each sample. For perivascular stiffness measurements, AFM was performed according to modified published procedures[14]. Tissues were acquired and embedded within OCT. Then, the frozen tissue blocks were sectioned at a thickness of 20 μm and immersed in proteinase inhibitor-contained PBS at room temperature. The next day, tumor sections were applied for AFM quantification of Young's modulus (Bruker, USA). Briefly, silicon nitride cantilevers with a spring constant of 0.15 N/m were attached by a borosilicate glass 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 (2×10^4) supplemented with 100 µL ECM were seeded in the upper chamber and incubated overnight. The next day, ECM was removed, and TPCs (5×10^5) 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 (5×10^5) 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 (3×10^4) 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$^{\mathrm{hi}}$ 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.

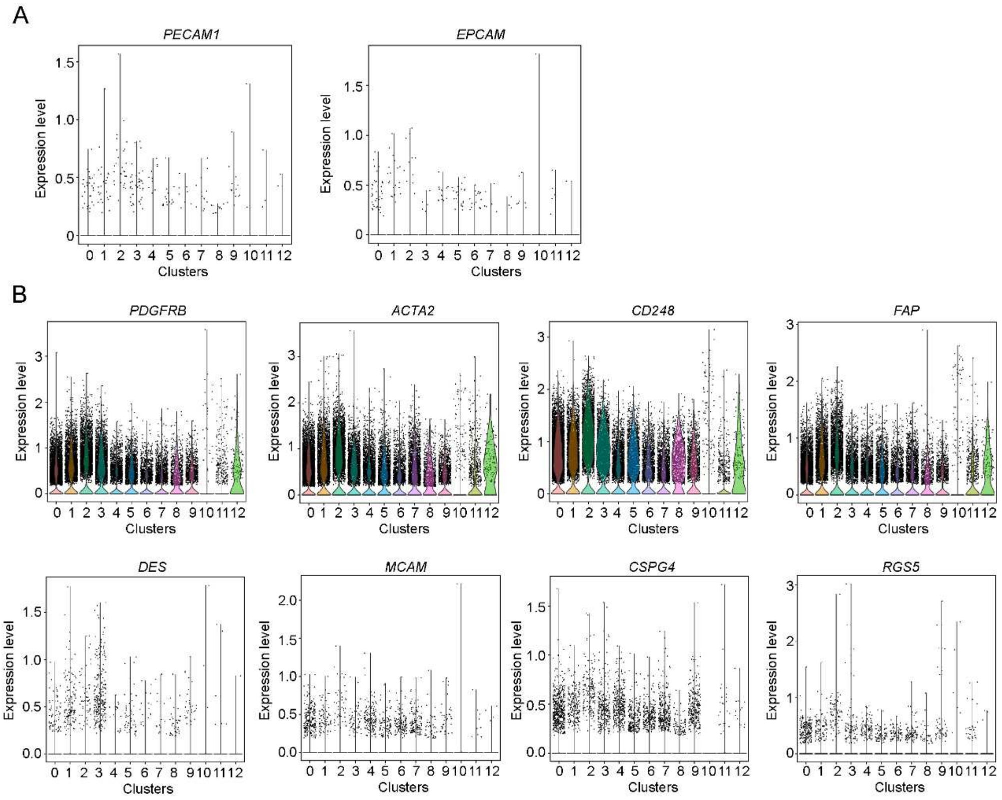
13. Kirmse R, Otto H, Ludwig T. Interdependency of cell adhesion, force generation and extracellular proteolysis in matrix remodeling. *J Cell Sci* 2011; 124: 1857-66.

14. Laklai H, Miroshnikova YA, Pickup MW, *et al*. Genotype tunes pancreatic ductal adenocarcinoma tissue tension to induce matricellular fibrosis and tumor progression.

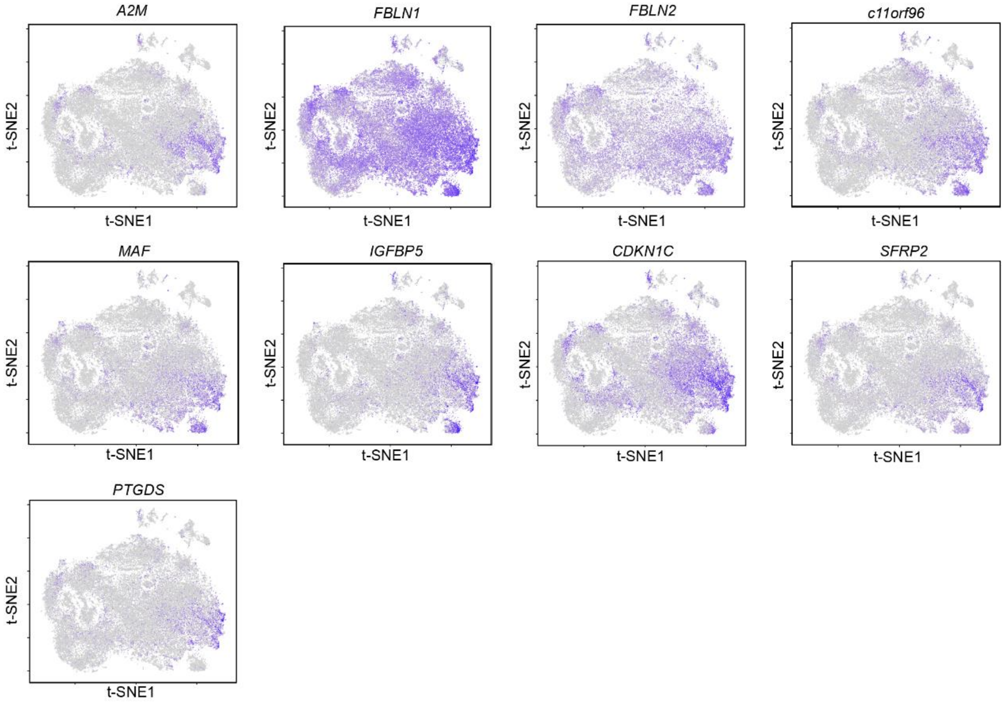
Nat Med 2016; 22: 497-505.

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

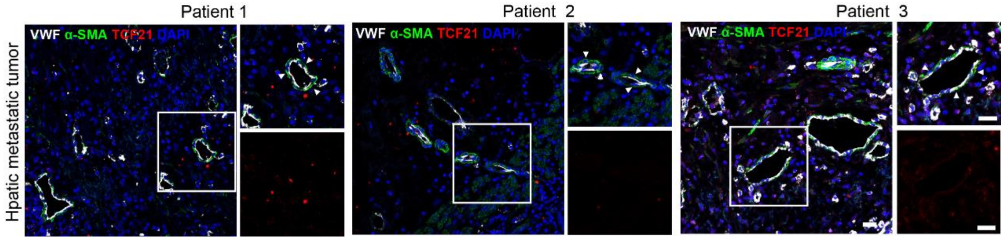
Supplemental figures and figure legends



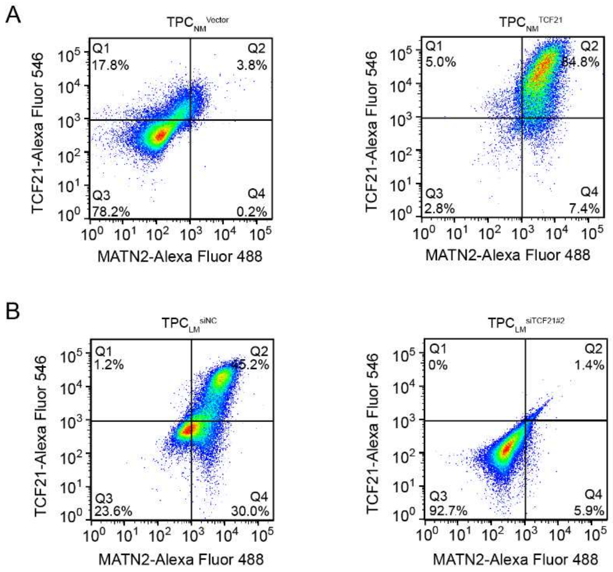
Supplemental Figure 1. Transcriptomic characterization of 13 subsets of TPCs. (A) Gene expression profiles of *PECAM1* and *EPCAM* in distinct subsets of TPCs. **(B)** Gene expression profiles of *PDGFRB*, *ACTA2*, *CD248*, *FAP*, *DES*, *MCAM*, *CSPG4,* and *RGS5* in distinct subsets of TPCs.



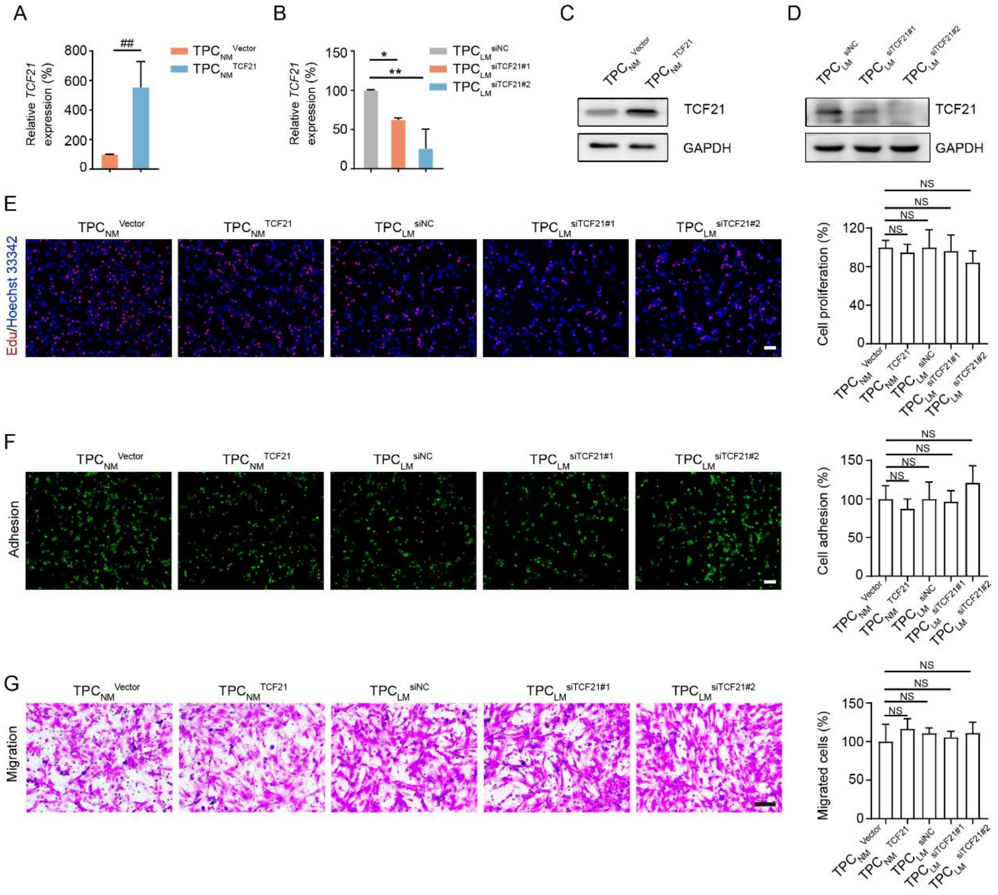
Supplemental Figure 2. Distribution of the matrix-pericyte-related genes in all subsets of TPCs. t-SNE visualization of gene distribution of *A2M*, *FBLN1*, *FBLN2*, *c11orf96*, *MAF*, *IGFBP5*, *CDKN1C*, *SFRP2* and *PTGDS* in all subsets of TPCs.



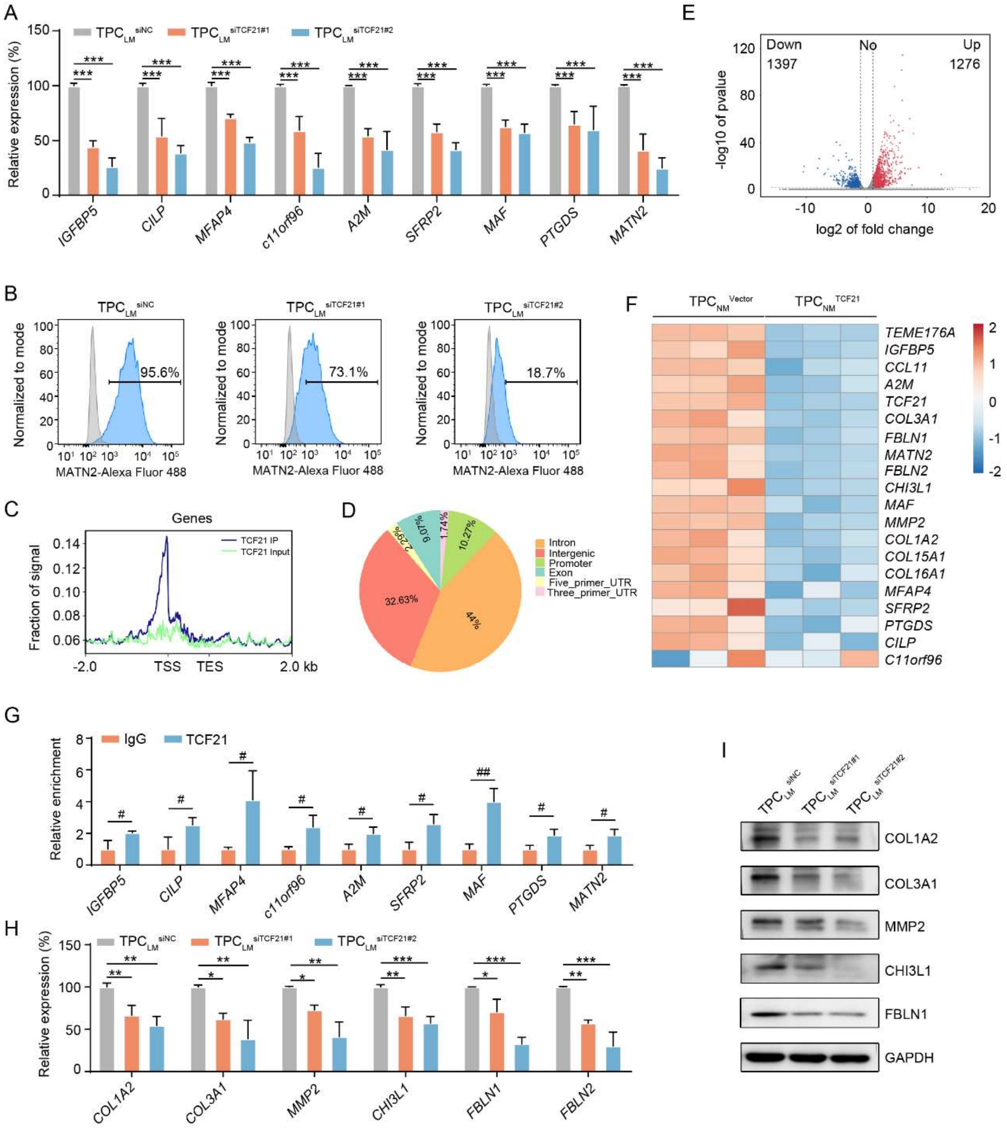
Supplemental Figure 3. Determination of TCF21 in TPCs in the hepatic metastatic tumors from CRC patients. Representative images of TCF21 staining (red) in TPCs (αSMA$^+$, green) in the hepatic metastatic tumors from CRC patients (n = 20). Scale bar, 20 μm.



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).

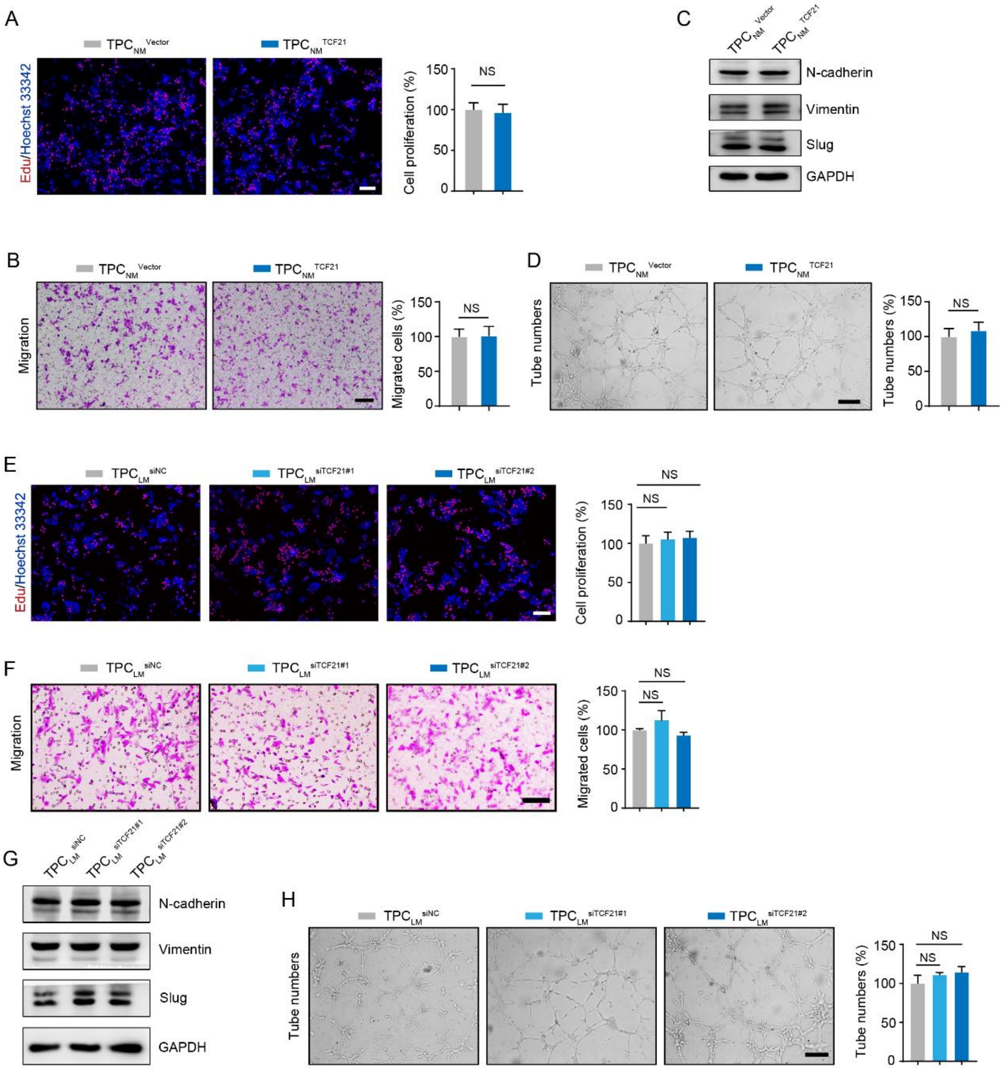


Supplemental Figure 5. Effects of TCF21 on the proliferation, adhesion, and migration of TPCs. (A, B) RT-qPCR analysis of *TCF21* mRNA levels in TCF21-overexpressing **(A)** or -knockdown **(B)** TPCs (n = 3). **(C, D)** Western blotting analysis of TCF21 in TCF21-overexpressing **(C)** or -knockdown **(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.

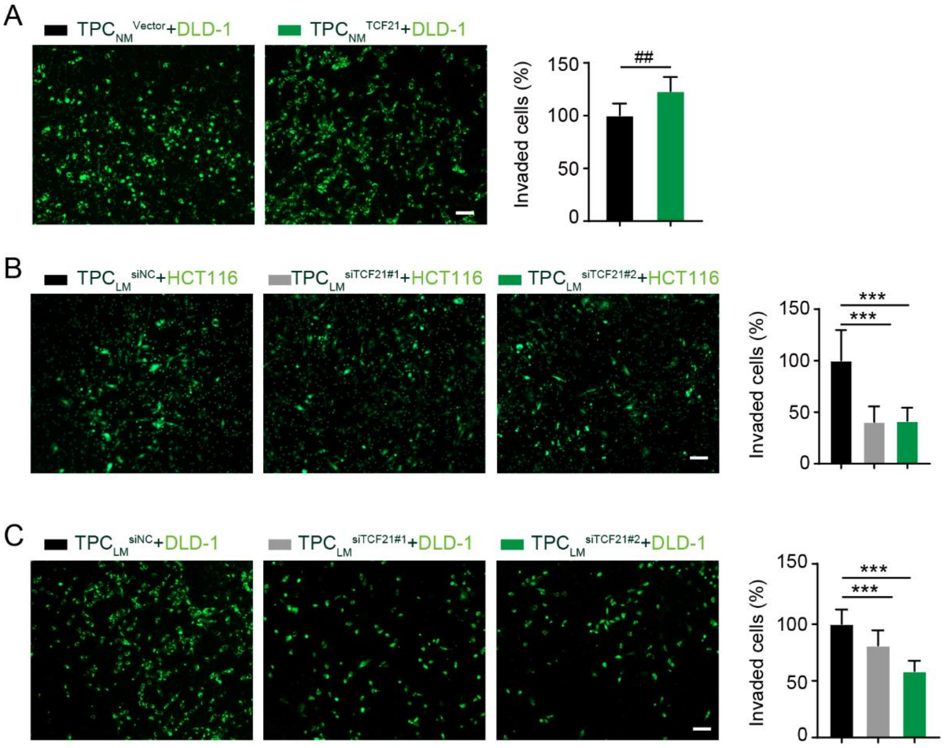


Supplemental Figure 6. TCF21 stimulates the generation of matrix-pericytes and induces ECM remodeling. (A) RT-qPCR analysis of matrix-pericyte-specific genes in TPC_{LM} transfected with siRNA targeting TCF21 ($TPC_{LM}^{siTCF21}$) or negative control (TPC_{LM}^{siNC}) (n = 3). **(B)** FCM analysis of 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

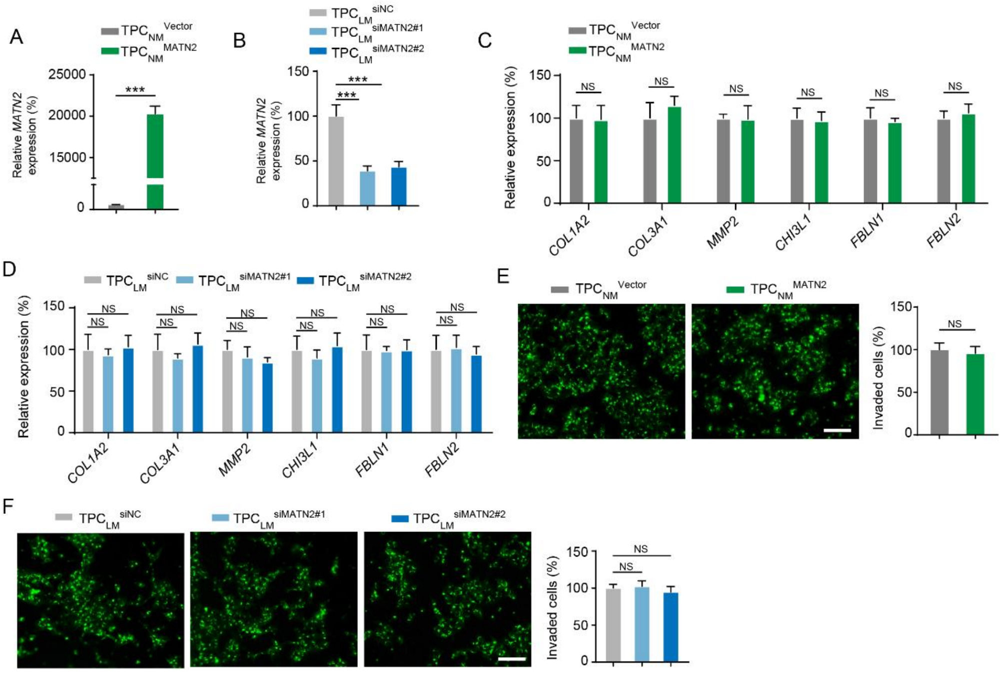
= 3); Red dots represent the up-regulated genes and blue dots represent the down-regulated genes. **(F)** Heat maps of the differentially expressed genes between $\text{TPC}_{\text{NM}}^{\text{Vector}}$ and $\text{TPC}_{\text{NM}}^{\text{TCF21}}$ (n = 3). **(G)** ChIP-qPCR analysis of TCF21 binding at the promoter of indicated genes in $\text{TPC}_{\text{NM}}^{\text{TCF21}}$ (n = 3). **(H)** RT-qPCR analysis of differentially expressed genes in $\text{TPC}_{\text{LM}}^{\text{siNC}}$ and $\text{TPC}_{\text{LM}}^{\text{siTCF21}}$. (**I**) Western blotting analysis of COL1A2, COL3A1, MMP2, CHI3L1, and FBLN1 in $\text{TPC}_{\text{LM}}^{\text{siNC}}$ and $\text{TPC}_{\text{LM}}^{\text{siTCF21}}$ (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.



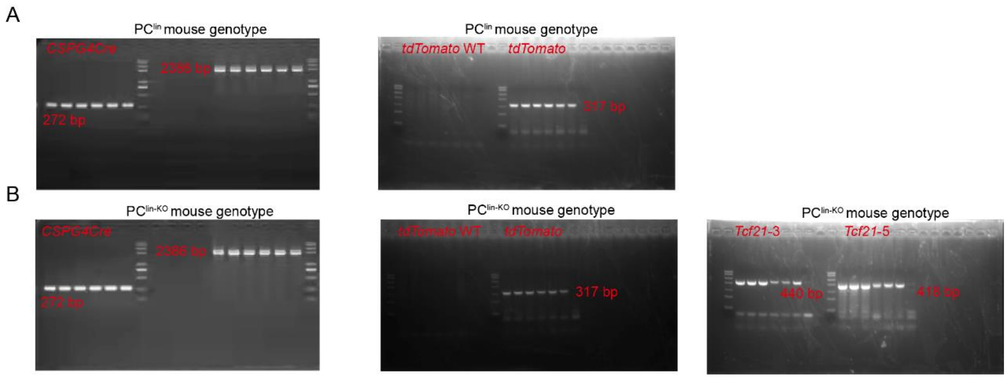
Supplemental Figure 7. TCF21 in TPCs has negligible effects on cell migration and angiogenesis. **(A)** EdU assay for the proliferation of HCT116 cells primed with conditioned medium from TPC_{NM}^{Vector} and TPC_{NM}^{TCF21} for 48 h (n = 3)**.** Scale bar, 100 µm. **(B)** Transwell assay for cell 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**).



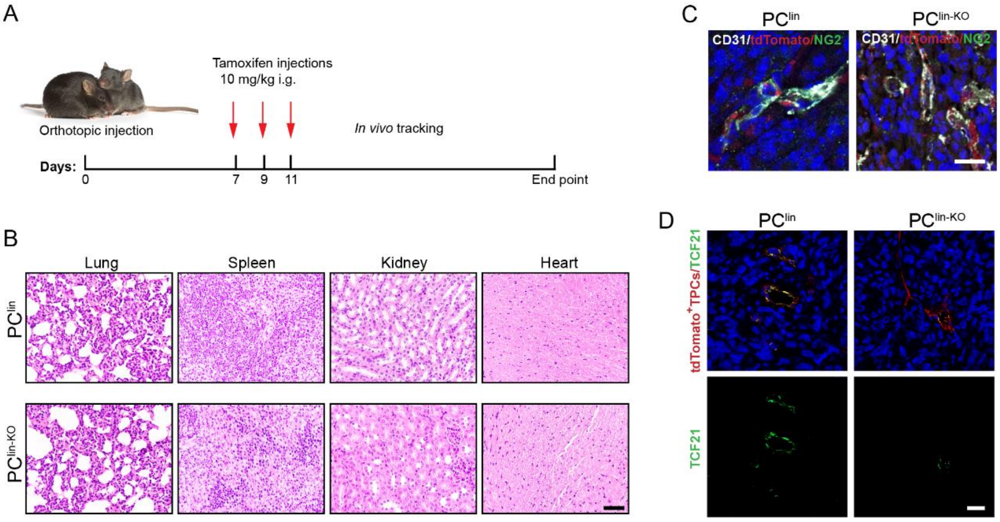
Supplemental Figure 8. TCF21 in TPCs promotes invasion of CRC cells. (A) Representative images and quantification of invaded DLD-1 cells (green). DLD-1 cells mixed with $\text{TPC}_{\text{NM}}^{\text{Vector}}$ or $\text{TPC}_{\text{NM}}^{\text{TCF21}}$ were seeded into the Matrigel-coated transwell. 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 $\text{TPC}_{\text{LM}}^{\text{siNC}}$ or $\text{TPC}_{\text{LM}}^{\text{siTCF21}}$ 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



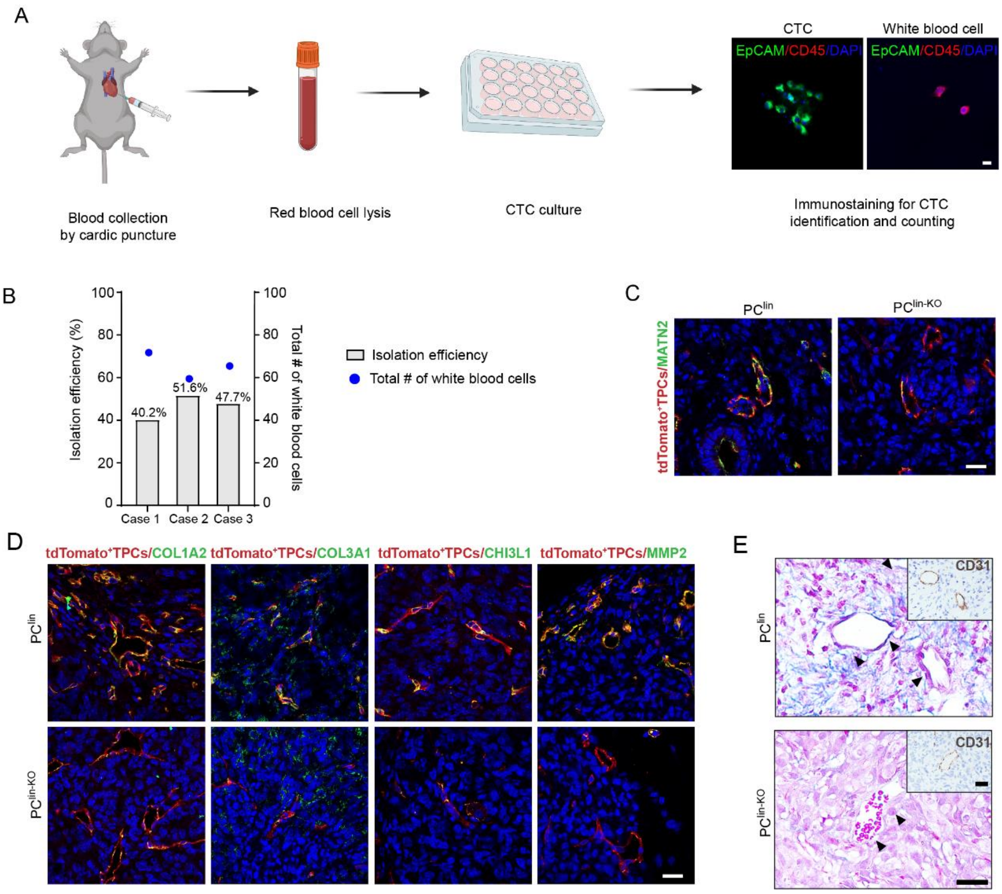
Supplemental Figure 9. MATN2 has negligible effects on ECM remodeling and CRC metastasis. (A, B) RT-qPCR analysis of *MATN2* in MATN2-overexpressing **(A)** or -knockdown (**B)** TPCs (n = 3). **(C)** RT-qPCR analysis of the ECM-related genes in $\text{TPC}_{\text{NM}}^{\text{Vector}}$ and $\text{TPC}_{\text{NM}}^{\text{MATN2}}$ (n = 3). **(D)** RT-qPCR analysis of the indicated genes in $\text{TPC}_{\text{LM}}^{\text{siNC}}$ and $\text{TPC}_{\text{LM}}^{\text{siMATN2}}$ (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**).



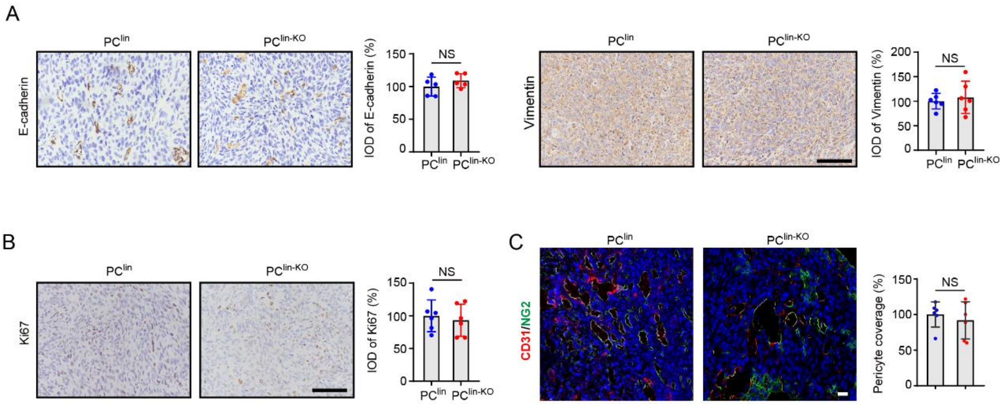
Supplemental Figure 10. Mouse genotyping. (A) PCR analysis of the genotype of PC^{lin} mouse (n = 6). **(B)** PCR analysis of the genotype of PC$^{lin\text{-}KO}$ mice (n = 6). All mice were analyzed by PCR genotyping. PCR analysis of *Cspg4Cre* showed two bands identifying homozygous *Cspg4Cre* at 2386 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.



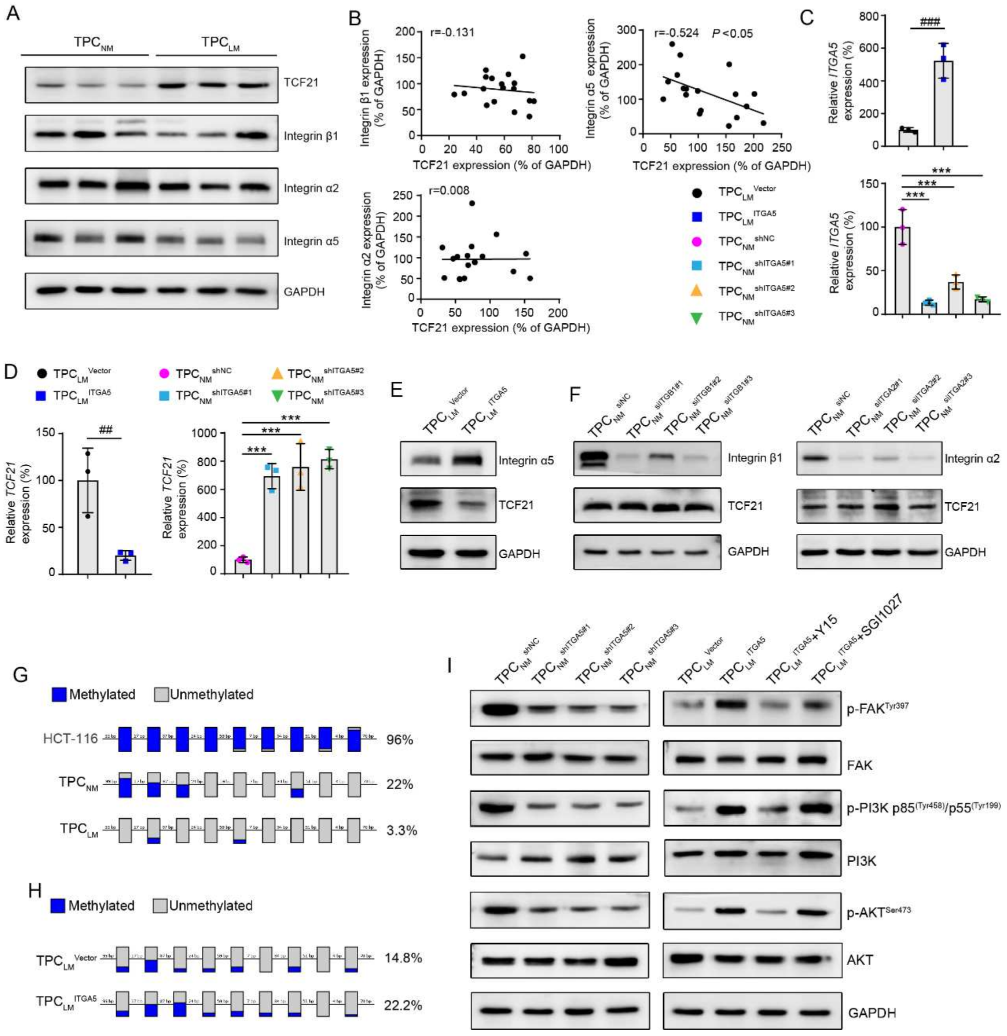
Supplemental Figure 11. Characterization of PC^{lin} and PC$^{lin\text{-}KO}$ mice. (A) Schematic diagram describing the experimental design of *in vivo* experiments. PC^{lin} mice and PC$^{lin\text{-}KO}$ mice were orthotopically injected with MC38-luc-LM3 cells. After 7 days, all mice were treated with tamoxifen (10 mg/kg) through intragastric administration every other day for three times. *In vivo* tracking was performed to detect tumor liver metastasis. **(B)** Representative images of H&E staining of lung, spleen, kidney, heart, and liver derived from PC^{lin} and PC$^{lin\text{-}KO}$ mice (n = 6). Scale bar, 50 μm. **(C)** Immunofluorescence analysis of TPCs (tdTomato) in tumor sections by NG2 (green) staining (n = 6). Scale bar, 20 μm. **(D)** Immunofluorescence analysis of TCF21 (green) in TPCs (tdTomato) from PC^{lin} mice and PC$^{lin\text{-}KO}$ mice (n = 6). Scale bar, 20 μm.



Supplemental Figure 12. Pericyte-specific knockout of *Tcf21* inhibits perivascular ECM remodeling and CRCLM. (A) Schematic diagram of the isolation and 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.

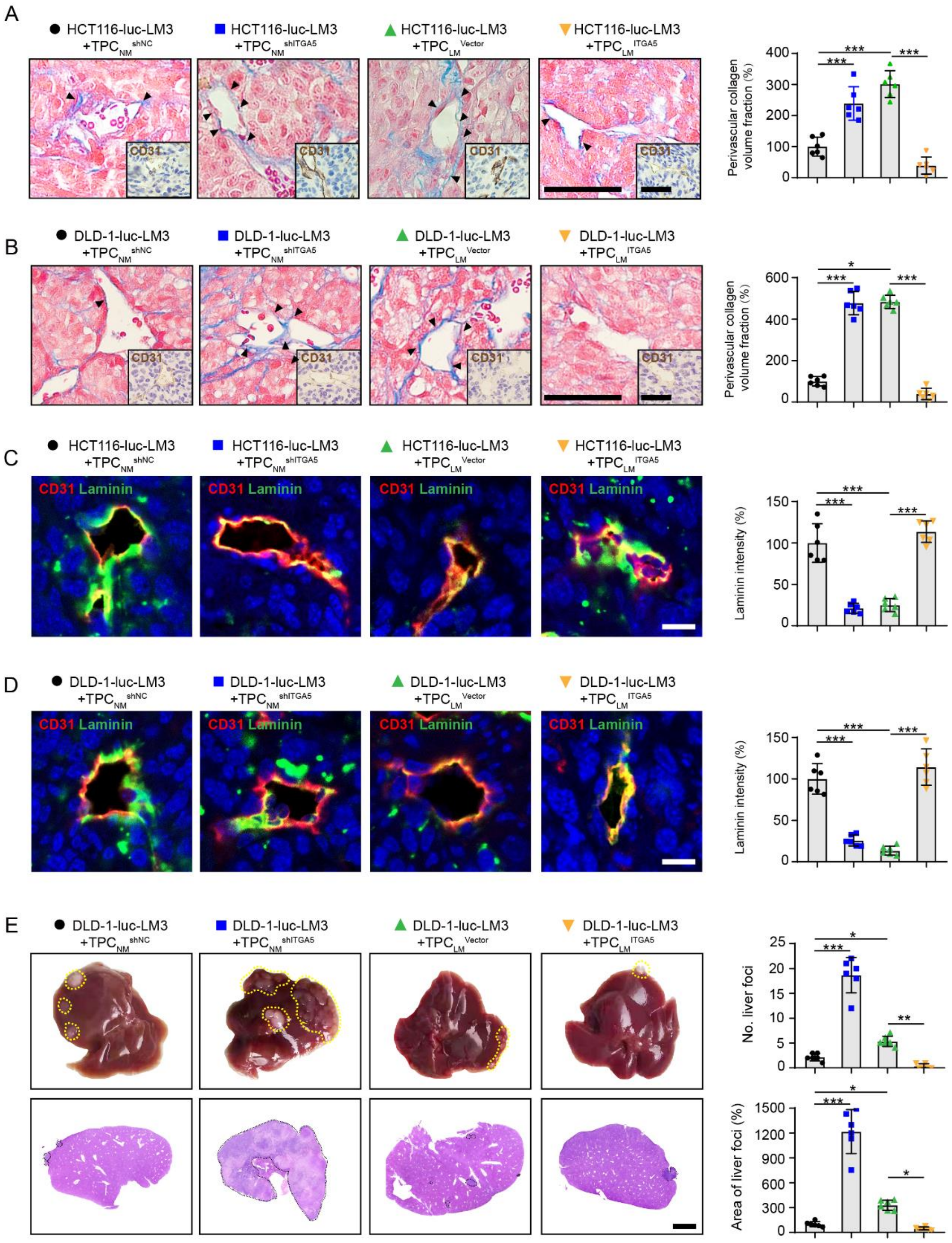


Supplemental Figure 13. Pericyte-specific knockout of *Tcf21* has negligible effects on EMT and proliferation of CRC cells. (A) Immunohistochemical staining and quantification of E-cadherin and Vimentin in orthotopic MC38 tumor tissues (n = 6). Scale bar, 50 µm. **(B)** Representative images and quantification 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.

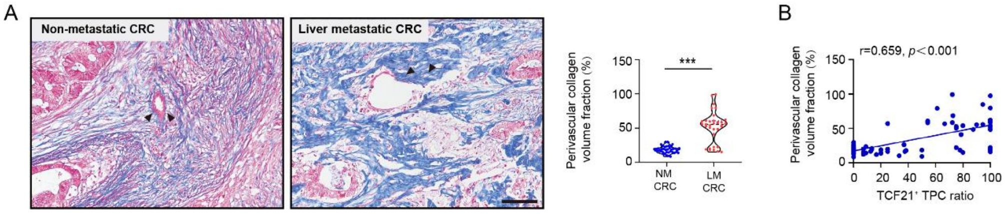


Supplemental Figure 14. Effects of integrin α2 and integrin β1 on TCF21 in TPCs. **(A)** Western blotting analysis of TCF21, integrin β1, integrin α2, and integrin α5 in TPC$_{NM}$ and TPC$_{LM}$ (n = 3). **(B)** Pearson's correlation analysis of the expression of integrin β1, integrin α2 and integrin 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)**

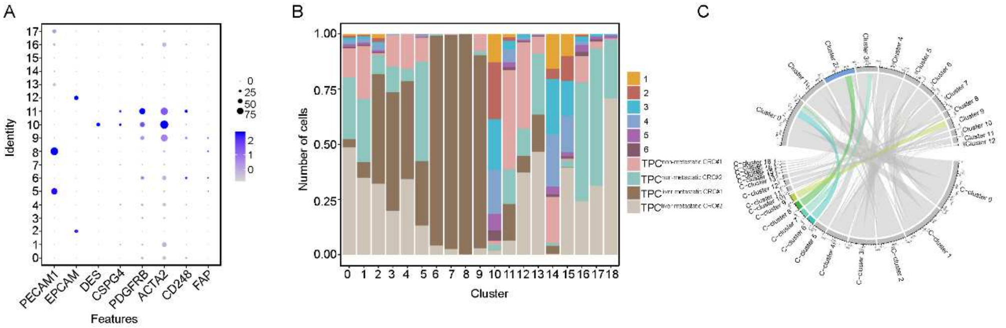
Bisulfite DNA sequencing analysis of *TCF21* promoter region in TPCs or HCT116 cells. Blue and gray circles represent methylated and unmethylated CpGs, respectively. The percentage of total methylated CpGs is given on right of each dataset (n = 3). **(H)** Bisulfite DNA sequencing analysis of *TCF21* 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.



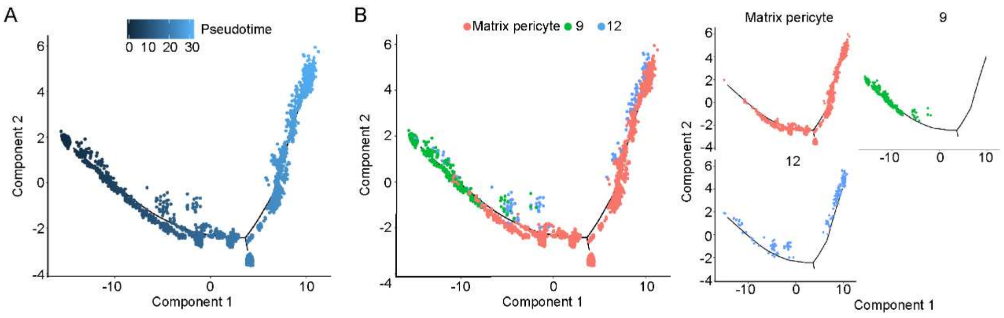
Supplemental Figure 15. Loss of integrin α5 in TPCs promotes perivascular ECM remodeling and CRCLM. (A, B) Representative images of Masson and CD31 staining in primary tumor sections (n = 6). Black arrows indicate the perivascular collagen fibers. Scale bar, 50 μm. The quantification of perivascular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

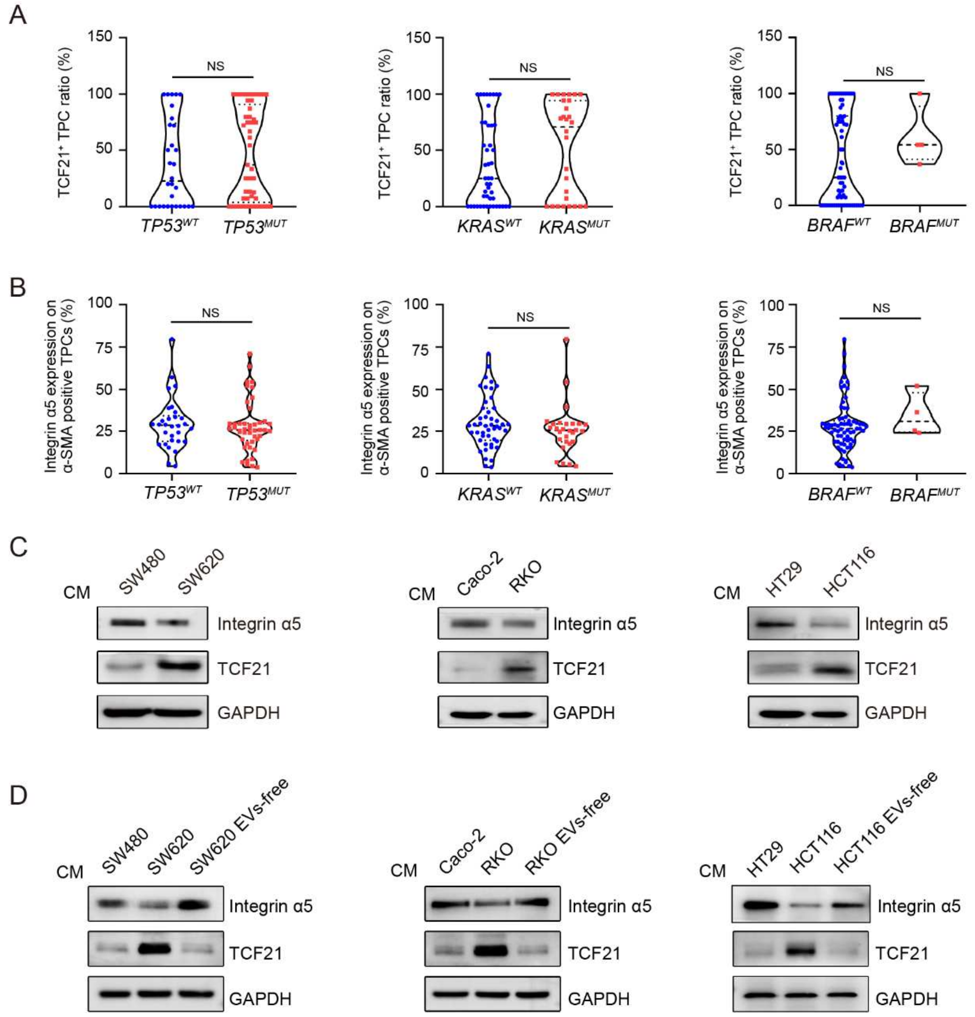


Supplemental Figure 17. Comparison of the scRNA-seq data derived from TPCs and the previous published data. (A) Dot plots for gene expressions in pericytes acquired from previous studies. Raw data of pericyte scRNA-seq were collected from six samples derived from previous studies (GEO accession GSM3140596, 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).



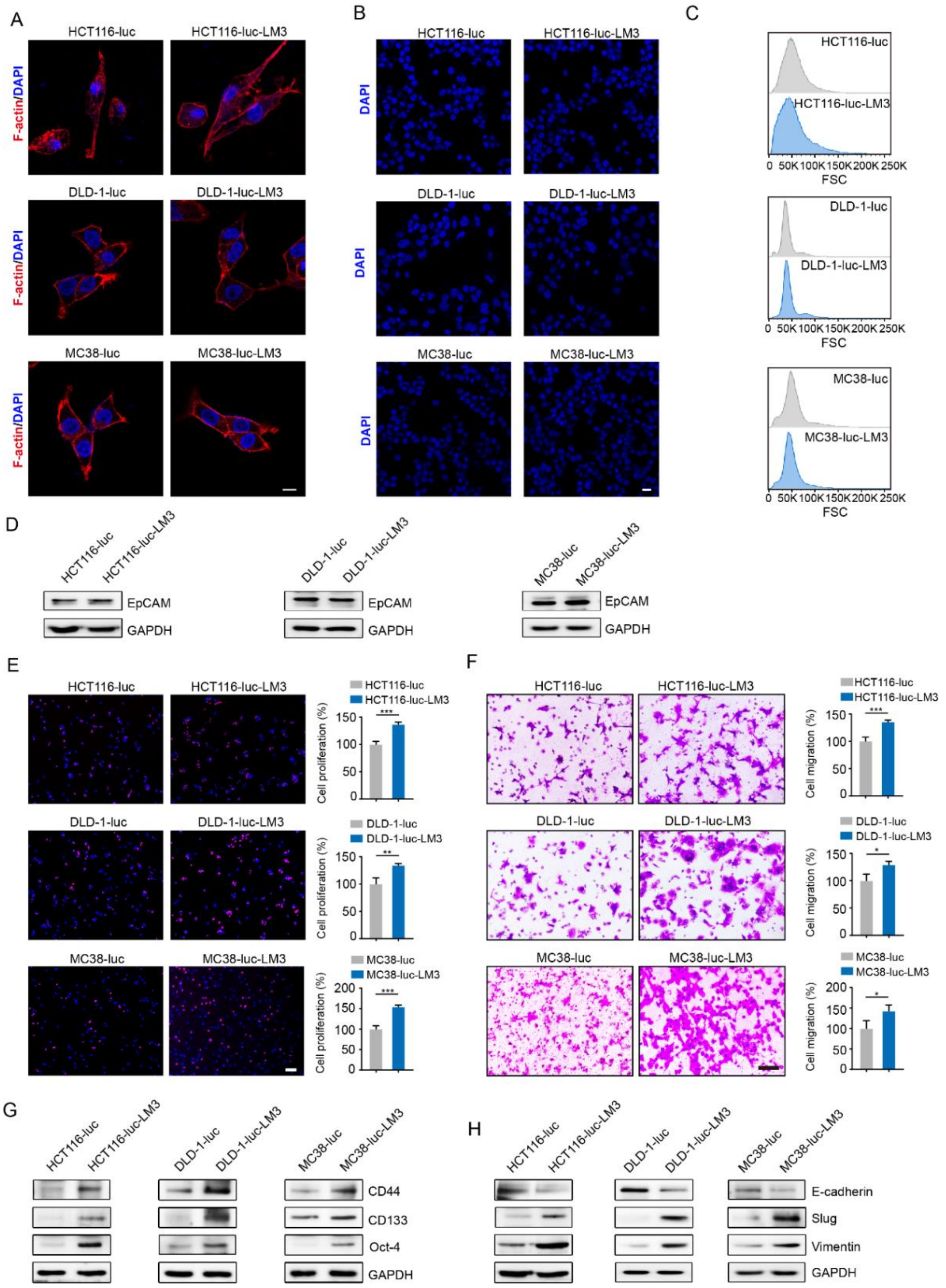
Supplemental Figure 18. Pseudo-time trajectory for dynamic changes in matrix-

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.

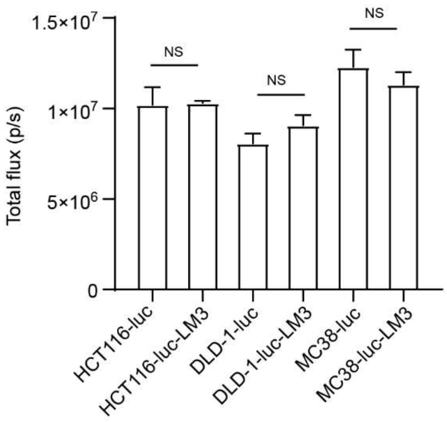


Supplemental Figure 19. Effects of metastatic CRC cells on the expressions of integrin α5 and TCF21 in TPCs. (A) Quantification of TCF21$^+$ TPC ratio in CRC patients with or without *TP53*, *BRAF* and *KRAS* mutation (n = 75). **(B)** Quantification of integrin α5 expression in TPCs derived from CRC 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 1×10^5 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*.



Supplemental Figure 20. Comparison of parental cells with LM3 cells. (A) Immunofluorescence analysis of cellular morphology (n = 3). Phalloidin-rhodamine was used to identify F-actin. Scale bar, 10 µm. **(B)** Immunofluorescence analysis of nucleus size (n = 3). Cell nucleus was measured after DAPI staining of fixed cells. Scale 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.

Supplemental Table 1. Clinical characteristics of CRC specimens.

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)

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 Fl: 418bp	T013083 *Tcf21*-flox

	JS04431-Tcf21-5wt-tR1	GTTTGCTAACTTGCTGCCACACAC		
4	XM003792- Cspg4-TF1	AAATCTAAGCGCGGGTCTGGC	WT:0bp KI:272bp	T006187 *Cspg4*-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

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

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