Single cell analysis of colorectal cancer identifies mitogen-activated protein kinase as a key driver of tumor cell plasticity

Florian Uhlitz[1,2,3,*], Philip Bischoff[1,4*], Stefan Peidli[1,2], Anja Sieber[1,2], Benedikt Obermayer[5], Eric Blanc[5], Alexandra Trinks[1,4], Mareen Lüthen[1,3], Yana Ruchiy[1], Soulafa Mamlouk[1,3], Roberto Arsie[6], Tzu-Ting Wei[6], , Kathleen Klotz-Noack[1,7], Roland F Schwarz[3,6], Birgit Sawitzki[7], Carsten Kamphues[3,8], Dieter Beule[5], Markus Landthaler[4,6], Christine Sers[1,3,4], David Horst[1,3,4], Nils Blüthgen[1,2,3,4,#] and Markus Morkel[1,3,4,#]

1 Charité Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Institute of Pathology, Charitéplatz 1, 10117 Berlin, Germany

2 IRI Life Sciences, Humboldt University of Berlin, Philippstrasse 13, 10115 Berlin, Germany

3 German Cancer Consortium (DKTK) Partner Site Berlin, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany

4 Berlin Institute of Health (BIH), Anna-Louisa-Karsch-Straße 2, 10178 Berlin, Germany

5 Charité Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Core Unit Bioinformatics (CUBI), Charitéplatz 1, 10117 Berlin, Germany

6 Berlin Institute for Medical Systems Biology (BIMSB), Max Delbrück Center for Molecular Medicine, Hannoversche Strasse 28, 10117 Berlin, Germany

7 Charité Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Institute of Medical Immunology, Augustenburger Platz 1, 13353 Berlin

8 Charité Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Department of Surgery, Hindenburgdamm 30, 12203 Berlin

[*] joint first authors

[#]corresponding authors: markus.morkel@charite.de, Tel: ++49-30-450 536 107; nils.bluethgen@charite.de, Tel: ++49-30-2093 92 390

Author contributions: PB, FU, SP, MM, AS, MLu, AT, RA, YR, SM, KKN conducted and analyzed experiments; FU, SP, NB, BO, EB, TTW performed bioinformatic analyses; MM, NB, PB, CS, DB, BS, DH, MLa, RFS conceived, designed, interpreted experiments and/or supervised parts of the study; PB, MM, DH, CK, AT contributed to clinical sample acquisition and preparation; MM wrote the manuscript; all authors provided critical feedback and helped shaping the research, analysis, and manuscript.

Competing Interests: The authors declare no competing interest.

Abstract

In colorectal cancer, oncogenic mutations transform a hierarchically organized and homeostatic epithelium into invasive cancer tissue lacking visible organization. We sought to identify differences in cellular composition between normal colon and colorectal cancer, and to define signals controlling cancer cell development.

We used single cell RNA and protein profiling to analyze tumors and matched normal tissues of twelve colorectal cancer patients. RNA metabolic labelling followed by single cell RNA sequencing in patient-derived normal colon and colorectal cancer organoids was employed to define colorectal cancer cell developmental trajectories.

We find that colorectal cancer tissues exhibited consistent changes in cellular composition in the epithelial, immune and stromal compartments across patients compared to normal colon. Tumor epithelial cells displayed patient-specific gene expression often correlating with somatic copy number alterations, but mainly organized into patient-overarching clusters. These clusters were defined by cell type-specific transcriptional programs related to stem, transient-amplifying and immature goblet cells, and showed differential expression for signatures of oncogenic traits such as replication stress. Patient-derived colorectal cancer organoids exhibited developmental trajectories forming along a gradient of mitogen-activated protein kinase activity that were Wnt-independent. Likewise, colorectal cancer cell types of patient samples were organized by mitogen-activated protein kinase activity.

Our single-cell analyses provide a subtyping system for colorectal cancer cells based on transcriptional readout of morphogenetic signals and oncogenic traits. We provide evidence that mitogen-activated protein kinase signaling, a key pathway for targeted therapy, is a main driver of colorectal cancer cell plasticity.

Keywords: Single Cell RNA Sequencing, Transcriptomics, Cancer Profiling, Organoids

Introduction

Cells in the human body develop along trajectories controlled by intrinsic and extrinsic signals to ensure tissue homeostasis. Cancer cells are compromised in their ability to maintain homeostasis, as oncogenic mutations activate signaling pathways cell-intrinsically and render cancer cells less reliant on paracrine signals[1]. Furthermore, cancer cells induce remodeling of neighboring tissues, for instance by secreting growth factors not found in the normal cellular environment[2]. Cancer cells are also immunogenic or associated with inflammation, and therefore attract immune cells[3]. These processes intersect and result in the emergence of a qualitatively and quantitatively unbalanced cellular ecosystem in cancer tissue. Cell interactions and cancer cell-intrinsic signal processing are critical for tumor progression and therapy response[4,5].

Colorectal cancer (CRC) commonly initiates via mutations activating Wnt/β-catenin signaling that maintains stem cells in the normal colon epithelium, while subsequent mutations deregulate further signaling pathways such as the RAS-RAF-MEK-ERK (also known as mitogen-activated protein kinase; MAPK) signaling cascade[6]. Less frequently, CRC initiates via BRAF mutations, or from chronic inflammation increasing the mutation rate in the tissue[7,8]. Genetic CRC drivers have direct and indirect effects on cancer cell development and the cellular composition of CRC and its microenvironment. There is substantial evidence for the existence of tumor cell subpopulations in CRC, as so-called cancer stem cells can be distinguished by surface proteins like CD133, EPHB2 or LGR5[9–12] and cells at the invasive front expressing genes such as the matrix metalloproteinase MMP7 contribute disproportionally to metastasis[13,14]. Single cell technologies allow to determine cellular constituents of CRC, and previous studies have mainly focused on the immune and stroma compartments[15–17]. However, it has not been investigated systematically how CRC cells organize into distinct states along developmental trajectories, and whether CRC cell types are distinguished by oncogenic traits relevant for therapy.

Here, we use droplet-based single-cell RNA sequencing to profile cell types and their differentiation states in normal colon and tumor tissues of twelve CRC patients, and in patient-matched CRC organoids. We identify tumor cell clusters across patients that differ in gene expression signatures related to oncogenic signaling and DNA replication. We show that CRC cell differentiation states are informed by MAPK activity.

Methods

Collection and processing of clinical specimens

Fresh normal colon and colorectal cancer tissues were acquired during the intraoperative pathologist´s examination at Charité University Hospital Berlin. Tissues (approx. 0.1-0.4g) were minced using scalpels, processed using the Miltenyi Human Tumor Dissociation Kit (Miltenyi, #130-095-929) and a Miltenyi gentleMACS Tissue Dissociator (Miltenyi, #130-096-427), using program 37C_h_TDK_1 for 30-45min. For three tumors, we also used digestion with the cold active protease from Bacillus licheniformis (Sigma P5380) at approx. 6°C for 45min. with frequent agitation, following a published protocol [18] (see Supplementary Fig 1). Cell suspensions were filtered using 100µm filters, pelleted by centrifugation, and treated with 1ml ACK erythrocyte lysis buffer, washed and resuspended in ice-cold PBS, and filtered using 20µm filters. Debris was removed using the Debris Removal Solution (Miltenyi #130-109-398). Cell suspensions were analyzed for cell viability >75% using LIVE/DEAD Fixable Dead Cell Stain Kit (488nm; Thermo Fisher) and a BD Accuri cytometer.

Single-cell proteogenomic and immune repertoire profiling

Fresh normal colon and colorectal cancer tissue were minced and digested in 10ml HBSS containing 2% FCS, 0.6% BSA, 0.5mg/ml Collagenase Type IV (Sigma-Aldrich, #C5138) and 30U/ml DNase I (Sigma-Aldrich, #AMPD1-1KT) for 30min at RT. Cell suspensions were filtered using 100µm filters, washed twice with HBSS and enriched for CD45+ cells using CD45-microbeads, LS-columns and magnet separators according to manufactures instruction (Miltenyi, #130-045-801). CD45+ positive cells were washed in PBS/2% FCS, FcR-blocked with Beriglobin (3.2mg/ml) and 2% FCS and stained with pre-mixed Total seq-C antibodies antibodies (Biolegend; anti-human CD20 #302363; anti-human CD19 #302265; anti-human CD184 #306533; anti-human HLA-DR #307663; anti-human CD40 #334348; anti-human CD38 #303543; anti-human CD27 #302853; anti-human CD3 #300479; anti-human CD45RA 304163; anti-human CD45RO #304259; anti-human CD28 #302963; anti-human CD11a #350617; anti-human CD4 #300567; anti-human CD25 #302649; anti-human CD127 #351356; anti-human TIGIT #372729; anti-human CD152 #369621; anti-human CD137 #309839; anti-human/mouse/rat CD278 #313553; anti-human CD95 #305651; anti-human CD134 #350035; anti-human CD8a #301071; anti-human CD69 #310951; anti-human CD103 #350233; anti-human CD33 #366633; anti-human CD11b #301359; anti-human CD14 #301859; anti-human CD16 #302065; anti-human CD64 #305045) for 20min. at room temperature.

Single-cell RNA sequencing

10^4 single cells were used for single-cell library production, using the Chromium Single Cell 3´Reagent Kits v3 and the Chromium Controller (10x Genomics). Libraries were sequenced on a HiSeq 4000

Sequencer (Illumina) at 200-400mio. reads per library to a mean library saturation of 50%. This resulted in 35.000 to 120.000 reads per cell.

DNA Sequencing

For panel sequencing, DNA was extracted from FFPE tumor tissue using the Maxwell RSC DNA FFPE Kit (Promega) or the GeneRead DNA FFPE kit (Qiagen) and sequenced using a CRC panel[19], and/or the Ion AmpliSeq Cancer Hotspot Panel (CHP) v2 and an IonTorrent sequencer (ThermoFisher). Variant calling was performed using Sequence Pilot (Version 4.4.0, JSI Medical Systems) or SoFIA[19]. For exome sequencing, DNA was isolated from fresh frozen tumor tissue using the DNeasy Blood and Tissue Kit (Qiagen). Exomes were sequenced using the AllExon Human SureSelect v7 Kit (Agilent).

Histology and immunostaining

3-5 µm tissue sections of formalin-fixed and paraffin-embedded (FFPE) tissue were used. Immunostainings of FFPE tissue sections were performed on the BenchMark XT immunostainer (Ventana Medical Systems),using CC1 mild buffer or Ultra CC1 buffer (Ventana Medical Systems) for 30 min at 100°C for antigen retrieval. The following primary antibodies were used: rabbit anti-TFF3 (1:250, Abcam, ab108599), mouse anti-FABP1 (1:1000, Abcam, ab7366), rabbit anti-OLFM4 (1:100, Atlas Antibodies, HPA077718), mouse anti-EPCAM (1:100, ThermoFischer Scientific, MS-144-P1), rabbit anti-Ki67 (1:400, Abcam, ab16667), mouse anti-Ki67 (1:50, Dako, M7240), rabbit anti-LYZ (1:1500, Abcam, ab108508), rabbit anti-EREG (1:50, ThermoFischer Scientific, PA5-24727), anti-PARP1, mouse anti-MUC2 (1:50, Leica, NCL-MUC-2), mouse anti-CK17 (1:10, Dako, M7046), mouse anti-MMP7 (1:100, ThermoFisher Scientific, MA5-14215). Images were taken using an Axio Vert.A1 fluorescence microscope (Zeiss) equipped with an Axiocam 506 color camera (Zeiss) or a CQ1 Yokogawa Benchtop Analysis System. Hematoxylin-and-eosin and immunohistochemical images were taken using a Pannoramic SCAN 150 slide scanner (3DHISTECH).

Organoid culture and metabolic labelling

Cells from patient-derived tumor tissues were washed in Advanced DMEM/F12 medium (Gibco), embedded in Matrigel, and cultured in 24-well plates, as published[20]. Wnt3 and R-Spondin3 were prepared as conditioned media[20]. For single-cell sequencing, organoids were dissociated completely using TrypLE and DNAseI, and filtered via a 20µm filter. For single cell SLAM-seq, NCO, P009T and P013T replicate cultures were cultured in media with and without Wnt/R-Spondin. Organoids were metabolically labelled in culture using 200µM 4-thio-uridine for 3 hours[21], harvested, disaggregated to single cells by TrypLE, and fixed in fixation buffer (80% methanol/20% DPBS) at ≥ -20 °. Samples were warmed to room temperature and incubated with 10mM iodoacetamide. Alkylation was carried out overnight, in the dark, with gentle rotation, followed by two washes with cold fixation buffer. Single

cell suspensions were rehydrated and incubated 10 minutes at room temperature in 100mM DTT. Samples were resuspended in fixation buffer and conserved at -80 °C.

Single-cell RNA-seq data analysis

UMIs were quantified using cellranger 3.0.2 with reference transcriptome GRCh38. Spliced, unspliced and ambiguous UMIs were quantified with velocyto[22] (mode: run10x, default parameters). Quality control filters were set to include cells with 500 to 5000 genes detected, 1000 to 50000 UMIs counted, fraction of mitochondrial reads ranging between 0 and 0.8, fraction of spliced reads ranging between 0.3 and 0.9, fraction of unspliced reads ranging between 0.1 and 0.7 and fraction of ambiguous reads ranging between 0 and 0.2. After filtering, UMI counts were variance-stabilized using scTransform[23] with 3000 variable features, while regressing out fraction of mitochondrial reads and differences in S-Phase and G2M-Phase scored with Seurat v3[24]. The top ten principal components were used to construct shared nearest neighbor (SNN) graph and UMAP embedding as implemented in Seurat v3. Next, main cell types (epithelium, stromal, and immune cells) were identified by scoring cell type markers across Louvain clusters for each sample (resolution = 1). Cell type markers used to score epithelium, stromal, and immune cells were adapted from Similie et al. [25] and are listed in Supplementary table 1. Sample-wise quality control assessments and subsettings into main cell types are documented at sys-bio.net/sccrc/. Normalized subsets were merged for each main cell type of normal and tumor samples without further batch correction. SNN graph, Louvain clusters and UMAP embeddings were recomputed for each subset based on top ten principal components. Louvain cluster-specific marker genes of merged normal and tumor samples were used to identify sub cell types among epithelial, stromal and immune subsets. Here, marker genes were determined with Seurat (wilcox text) at a minimum log fold change threshold of 0.25. Gene expression sets were taken from the hallmark signature collection of the Broad institute[26], unless otherwise referenced in the main text, and were scored as implemented in the progeny R package and Seurat v3, respectively.

To compute ligand-receptor connectivity between cell type clusters, UMI counts were summed for all ligands of the same pathway in each stromal or immune cell type of normal or tumor samples. Summed ligand counts were scaled to range between zero and one for each pathway. The fraction of normal and tumor proliferative epithelial cells expressing a given receptor was calculated and fractions were averaged across receptors for each pathway and cell type. Averaged fractions of cells expressing receptors were likewise scaled to range between zero and one for each pathway. Connectivity between stroma or immune ligand expression and epithelial receptor expression was calculated as the product of scaled ligand counts and scaled receptor expression fractions and, accordingly, also ranged between zero and one. Ligands and receptors used for each pathway are listed in Supplementary table 2.

Single cell SLAM sequencing data was pre-processed using Seurat, and counted using the resulting alignments (as BAM files), using a custom pipeline utilizing Snakemake[27], SeqAn[28], R. For each read, the numbers of T nucleotides and T-to-C conversions was counted, leaving out positions with common SNPs (using the dbSNP build 151 as available as track from UCSC genome browser). For each molecule as identified by cell barcode and UMI, positions with discordant nucleotides were excluded. Subsequently, molecules were counted as nascent RNA if they contained a T-to-C conversion, and old RNA otherwise.

Total-seq data and the associated 5' sc-RNA-seq data was integrated with the 3' sc-RNA-seq data by first integrating the two data sets from the same tumor using IntegrateData of Seurat, and a common shared nearest neighbor (SNN) graph was constructed using FindNeighbors. Subsequently, only immune cells of the 3' data each cell of the Total-seq were considered, and the cells of the Total-Seq run were assigned to the cluster with which it had most neighbors in the SNN graph. The Total-Seq signals were normalized and scaled using Seurat's default parameters, and average expression of Total-Seq markers in each cluster was used to display cluster-specific expression.

scRNAseq and scSLAM-seq data for organoids was analyzed using scanpy[29] and scvelo[30]. For diffusion map analysis and RNA velocity, cells were first filtered by the number of genes (between 2000 and 5000) and the percent mitochondrial reads (between 0.075 and 0.2) and normalized, using scvelo standard settings. Cell cycle was scored according to the scanpy standards, and S_score, G2M_score, percent mitochondrial reads and UMI counts per cell were regressed out. The diffusion map was calculated on the top 10 principal components and using a neighborhood graph with 50 neighbors and calculated on all genes. Moments were calculated on 30 principal components and 30 neighbors, and velocity was calculated using the stochastic model.

Calling of SCNAs from single cells and whole exome data

InferCNV v1.3.3 was used with default parameters to estimate copy number changes per single cell and genomic location from scRNA seq data. To detect copy number aberrant clones in tumor samples, inferCNV dendrograms were cut at k = 2 for each patient. For the resulting two clones per patient and for each normal sample, a clone-wise SCNA score was computed by calculating the average standard deviation in inferCNV expression of all cells in a given clone or normal sample and divided by the average standard deviation in inferCNV expression of all normal samples taken together. Clones with a SCNA score greater than the highest observed score for normal samples were considered copy number aberrant clones.

To validate SCNA calls from single cells we performed allele-specific SCNA calling for each patient from bulk whole exome data. Germline variants were discovered *de-novo* and read counts were

accumulated for each allele at heterozygous germline variants using the bcftools (v1.9) [31] multi-allelic caller. Discovered variants and read counts were passed to Sequenza (v3.0.0) [32] for segmentation and calling of allele-specific SCNAs.

Ethics permission

All patients were aware of the planned research and agreed to the use of tissue. Research was approved by vote EA4/164/19 of the ethic´s commission of Charité Universitätsmedizin Berlin.

Results

Single Cell RNA sequencing identifies key features of CRC

To capture the cellular diversity in CRC and track changes from normalcy to disease, we performed single-cell transcriptome analysis of twelve previously untreated CRC patients (Fig. 1A). We utilized tumor tissue samples that included the invasive tumor front and matched normal tissues (Supplementary Fig. 2). Tumors encompass stages pTis (Tumor *in situ*) to pT4, that is, from cancer confined within the lamina propria to invasive through the visceral peritoneum, with or without metastasis, and with various locations along the cephalocaudal axis of the colon. Genetic analysis revealed mutational patterns characteristic for canonical CRC progression in most tumors; however, tumors from patients P007, P014 and P026 contained the $BRAF^{V600E}$ mutation often associated with the serrated progression pathway and tumor P008 was colitis-associated and harbored a TP53 mutation. Eleven patients were diagnosed with microsatellite-stable (MSS) CRC, while patient P026 was microsatellite-instable (MSI).

We enzymatically dissociated the fresh normal and cancer tissues to single cells, produced transcriptome libraries using a commercial droplet-based system, and sequenced the libraries to obtain transcriptomes covering 500 to 5000 genes per cell. For each library, single cell profiles were split into epithelial, immune or stromal subsets, using known marker genes[25]. Subsets were merged into three separate supersets with >30000 epithelial or immune cells each or 3003 stromal cells (see Supplementary Fig. 3 for quality controls and initial clustering). We observed that quality parameters, such as fraction of mitochondrial reads, were more variable in epithelial cells, in line with previous studies[25,33].

When visualized in a standard UMAP embedding employing ten UMAP components[34], transcriptomes clustered by cell lineage (Fig. 1B). Cells from normal versus tumor samples distributed similarly across many clusters, but we also identified clusters which were strongly enriched for either normal or tumor tissues (Fig. 1C). These features of the UMAP embedding indicate that our single cell data are largely free from sample-specific bias, but instead reflect intrinsic differences between normal and tumor cell transcriptomes.

Cell type census in CRC versus normal colon

We used cell type-specific signatures and marker genes to annotate cell clusters[25] (Fig. 2A, B). In the normal epithelium, we identified clusters of undifferentiated cells by transcripts of stem cell markers such as *OLFM4*. Neighboring clusters were annotated as enterocyte progenitors or mature enterocytes

by expression of absorptive lineage markers such as *KRT20* and *FABP1* (Supplementary Fig. 4A, Supplementary table 3). *BEST4*- and *OTOP2*-expressing enterocytes formed a discrete cluster (Fig. 2A), as observed previously[35]. Further separate epithelial clusters were identified as immature and mature secretory goblet cells expressing *MUC2* and *TFF3,* and as tuft cells expressing *TRPM5.* In tumor tissue, we additionally identified four tumor-specific clusters, TC1-TC4, expressing high levels of stem cell markers such as *OLFM4*, *CD44* and *EPHB2* (Fig. 2A-B; Supplementary Fig. 5 for cluster compositions per patient). Clusters of differentiated absorptive and secretory cells were smaller in tumors, and profiles representing tuft cells and *BEST4*/*OTOP2*-positive enterocytes were vastly underrepresented. *MMP7* was among the few genes expressed exclusively in the tumor, but not normal epithelium (Supplementary Fig. 4A).

We annotated immune cell clusters by lineage-specific marker genes (Fig. 2C, Supplementary Fig. 4B, and Supplementary table 4), and furthermore validated cluster identities by staining of immune cells isolated from the MSI CRC patient P026 with a panel of nucleic-acid tagged antibodies (Fig. 2D). Immune cell clusters could be grouped into two B cell, one germinal cell, nine plasma cell, six T cell, two myeloid and one mast cell cluster.

B cell clusters were defined by expression of *CD19, MS4A1* (encoding CD20), *CD79A*, *CD79B*, *IGD*, *IGM* and strong staining by the tagged CD19 and CD20 antibodies. High expression of *CXCR5* and *BCL6* was characteristic for the germinal center cluster, whereas the plasma cell cluster could be defined based on their expression of *SDC1*, *CD38, CD27* and lack of *MS4A1* expression. As the individual plasma cell clusters varied according to lambda and kappa light chain usage (*IGLC2, IGLC3, IGKC*) we termed the plasma cell clusters L1-2 and K1-7, respectively. The plasma cell clusters also differed in proportion of IgG (*IGHG1 -4*)-producing cells (Figure 2C).

T cell clusters, identified by *CD3D* expression, could be further grouped into four *CD8A*-positive cytotoxic T cell and three *CD4*-positive T helper cell clusters (Fig. 2C). One T helper cell cluster was characterized by high transcription of *IL2RA, FOXP3, TIGIT, CLTA4, TNFRSF18* indicative for regulatory T cells (thus termed $CD4^{+}CD25^{+}$). Strong immunoreactivity for CD25 confirmed enrichment for regulatory T cells, although a proportion of cells also transcribed *IL7R* and stained positive for CD127 which suggests inclusion of non-regulatory conventional T cells (Fig. 2D). The remaining two T helper cell clusters (Tconv1, Tconv2) were characterized by high transcription of the activation marker *CD40LG*, with the cluster Tconv1 expressing more *CD69,* potentially marking tissue-resident memory T cells. Cells contained within the two main $CD8^{+}$ cytotoxic T cell cluster ($CD8^{+}$ T1 & T2) showed high transcription of effector genes such *PRF1*, *GZMB* and *GZMK* in one cluster, whereas the other cluster was enriched for proliferating *MKI67*-expressing cells.

Myeloid cell clusters predominately contained *CD68*-expressing macrophages (Figure 2C). Myeloid cell cluster 1 showed a more inflammatory profile as the cells were characterized by high expression of the antibacterial proteins *S100A8* and *S100A9* as well as *IL1B*. In contrast, cells belonging to the myeloid cell cluster 2 displayed high expression of the M2 marker *CD163* as well as *FCGR3A* transcription and CD16 protein reactivity (Fig. 2C, D).

Overall, we observed an increase in the numbers of T cells and macrophages in the tumors as compared to normal tissue. In contrast, we observed a decrease in total B and plasma cell numbers in the tumor microenvironment compared to the adjacent normal tissue counterparts (Fig. 2B), which is in line with a recent study[17]. Three immune cell clusters were larger in tumors compared to normal tissue across all patients analyzed: these were the plasma cell clusters L2 and K2, which were expanded in the tumor in contrast to the overall trend for less plasma cells in tumor tissue, as well as the CD8+ T4 T cell cluster. Plasma cell clusters L2 and K2 contain more IgG producing cells, which confirms the previously observed loss of IgA^+ but gain of IgG^+ plasma cells[17]. $CD8^+$ T4 is characterized by proliferating cells with reduced effector gene expression potentially indicating reduced anti-tumor activity.

Among stromal cells, we annotated an interconnected supercluster of fibroblasts (Fig. 2A, Supplementary Fig. 4C, and Supplementary table 5). Strikingly, one fibroblast cluster was confined to the tumor samples and was therefore designated as cancer-associated fibroblasts (CAFs). CAF transcriptomes were defined by high expression of collagen-encoding genes *COL1A1*, *COL1A2*, *COL3A1* and matrix metalloproteinase-encoding genes *MMP1*, *MMP11*, *MMP3* and *MMP2*, suggesting a role of these cells in the remodeling of the tumor extracellular matrix. The fibroblast supercluster, in addition, contained profiles of putative crypt base fibroblasts of the stem cell niche expressing the Wnt amplifier *RSPO3*[36,37], upper crypt fibroblasts expressing *BMP2* and *BMP4* that encode differentiation-associated growth factors[38] and a further small cluster of fibroblasts positive for various chemokine ligands and receptors including *CCL2*, *CCL8*, *CCL11*, *CCL13*, *CXCL1* and *CXCL14*. Further distinct clusters of stromal cells were composed of myofibroblasts, possibly intermingled with smooth muscle cells, expressing *DES* and the activated fibroblast marker *ACTA2*[39]. A pericyte cluster was marked by a combination of *MCAM* (encoding MUC18/CD146) and *STEAP4*. We also detected small numbers of endothelial cells and glial cells, respectively. Pericytes and endothelial cells were more frequent in the tumor samples, but, in contrast to CAFs, these cells were also present in normal tissue samples at low frequencies.

The tumor-specific TC1-4 epithelial cell clusters and expanded clusters of stromal and immune cells express multiple ligands and receptors of tumor-relevant signaling pathways, potentially allowing for novel paracrine interactions activating the Wnt/β-Catenin, EGFR and Met-HGF signaling pathways in the tumor microenvironment (Supplementary Fig. 6, 7).

Identification of tumor cells by somatic copy number aberrations

MSS CRC is defined by somatic copy-number aberrations (SCNAs). Thus, we next distinguished epithelial tumor cells from normal cells by inferring SCNAs from the single cell transcriptome data. We identified clusters of SCN-aberrant epithelial cells in ten out of the twelve tumors (Fig. 3A). P014 and P026 contained no cells with overt SCNAs. This was expected for tumor P026, which is MSI and thus defined by single nucleotide polymorphisms rather than SCNAs, but unexpected for P014, which was diagnosed as BRAF-mutant however MSS. Epithelial cells with SCNAs predominantly populated the TC1-4 clusters, along with substantial fractions of cells defined as stem cell/TA-like or immature goblet cells (Fig. 3B). In contrast, a majority of absorptive enterocytes and mature goblet cells were identified as copy-number normal, and therefore likely stem from non-cancerous tissue at the tumor margins (Supplementary Fig. 2). We used exome sequencing of tumors P007, P008 and P009 to validate SCNA calling from transcriptomes, showing that the procedure is robust for our single cell data (Supplementary Fig. 8A).

While most SCN-aberrant cells were rather undifferentiated, a fraction also showed high expression levels of the absorptive and secretory differentiation marker genes *FABP1* and *TFF3* (Fig. 3E). We used immunofluorescence to assess spatial distributions of proteins marking undifferentiated and differentiated epithelial cells (Fig. 3F). In normal colon sections, we detected the stem cell marker OLFM4 exclusively at crypt bases. In contrast, OLFM4 stained cells scattered throughout the tumor epithelium, as validated by co-staining with the epithelial marker EPCAM. Similarly, the proliferation marker MKI67 showed regionalized staining in lower crypt regions of the normal colon, but widespread staining throughout the tumor epithelium. The goblet cell and enterocyte differentiation markers TFF3 and FABP1 stained preferentially cells in the lower and upper crypt of the normal colon, respectively. In contrast, TFF3- and FABP1-positive tumor cell populations were not clearly organized in domains. TFF3-positive cells that were largely negative for MKI67 suggest the presence of secretory cell-like differentiated cells in CRC, in line with our finding of SCN aberrant tumor cell populations scoring as goblet cell-like in the single cell transcriptome analysis.

Patient-specific gene expression patterns often correlate with SCNAs

As individual CRCs are driven by specific combinations of oncogenic mutations and SCNAs, we were surprised by the lack of patient-specific epithelial cell clusters. We thus represented smaller differences between epithelial transcriptomes by using a higher number of top 50 components for UMAP embedding, instead of the initial top 10 components. At this resolution, the TC clusters and smaller parts of the immature goblet cell clusters clearly separated by patient (Fig. 4A). We re-clustered the

TC1-4 cells of the tumor samples into 13 subclusters (Fig. 4B). Subcluster 8 was removed from further analysis, as expression of immune-specific genes suggested misannotation of a few immune cell or epithelial-immune cell-doublet transcriptomes. The remaining subclusters were often patient-specific, such as subcluster 0 that was specific for patient P009 and subcluster 10 that was specific for patient P008 (Fig. 4C,D). Indeed, tumor P008 was selectively immunoreactive for cytokeratin 17 (KRT17), as predicted by transcriptome data (Fig. 4E). In contrast, subclusters 1, 2, 9 were represented in several patients. Accordingly, proteins encoded by genes in these subclusters such as *MUC2, EREG, LYZ* and *MMP7* were expressed in tumor cells of patient subgroups (Fig. 4D, E). We found that patient-specific gene expression clusters are often associated with SCNAs (Supplementary Fig. 8B). For example, subclusters 0 and 5 were enriched in genes located on chromosome 17q (p=1,41E-14 and p=3.9E-15, respectively, using Bonferroni-corrected hypergeometric distribution tests), and indeed SCNA profiles for patient P009 suggested amplification of this chromosome arm. Similar association exist for subcluster 4, chromosome 12q, patient P015 (p=1,77E-5), and subcluster 6, chromosome 20q, patient P013 (p=1,65E-33), respectively. However, not all putative amplifications manifested as transcriptome subclusters, and vice versa. Furthermore, proteins encoded by some subcluster genes showed regionalized expression, such as MMP7 that was immunoreactive predominantly for cells at the invasive front, suggesting that genetic/genomic aberrations and inductive signals can co-operate in directing CRC gene expression. In summary, our analysis indicates that patient-specific expression patterns in CRC are controlled partly by SCNA and are heterogeneous on single cell level. However, our cohort of 12 patients is too small to define patient-overarching subgroups based on single cell transcriptome patterns.

Epithelial tumor cell clusters differ by oncogenic traits and signals

We sought to define whether the tumor cells clustered as TC1-4, stem/TA-cell-like or goblet cell-like and defined by SCNAs were associated with oncogenic signaling or cancer-associated functional signatures (Fig. 5A, B, and Supplementary Fig. 9). We found a strong association of TC1 cells with the expression of hallmark signatures related to DNA repair and the G2/M replication checkpoint[26]. This indicates that TC1 cluster cancer cells experience high levels of replication stress, a therapy-relevant trait of many cancers, including CRC. Indeed, TC1 cluster cells were exclusively assigned to the S or G2/M cell cycle phases by gene expression (Fig. 5C), in line with cells under replication stress, as also seen by *XRCC2* expression (Supplementary Fig. 4A). The DNA damage-associated protein PARP stained many nuclei of the TC1-high CRC tissue P009 but not of TC1-low P008 (Fig.5D). TC2 transcriptomes were associated with high PI3K pathway activity, related to control of metabolism and apoptosis. Wnt/β-catenin target gene activity was high across all TC clusters. Wnt/β-catenin target gene

expression was not associated with specific tumor cell clusters, but stem/TA-cells had strongest expression of the LGR5-ISC signature that is Wnt activity-associated[12]; patient P012 tumor tissue showed the strongest overall Wnt/β-catenin activity. TC4 showed the strongest expression of YAP targets defined in intestinal tissue[40]. YAP transcriptional activity is linked to regenerative responses and tumor progression[41]. TC1 and TC4 were enriched for the expression of direct ERK targets that are activated by MAPK[42], across all tumors and in particular in P007. In summary, assessment of cell signaling and cell type signatures provides additional information on signaling pathway activities of tumor epithelial cell clusters, and specific features of individual tumors. The analyses suggest that assignment to the TC1-4 clusters reflects, at least partially, differential signaling activities and oncogene-induced functional traits of cancer cells.

RNA metabolic labeling defines MAPK-driven tumor cell hierarchies

To establish how cell type and signaling signatures are organized in tumor cell hierarchies, we established organoid lines of two tumor samples, P009, and P013 (Fig. 6A). We confirmed the identity of the organoids with the matched tumor tissue on a mutational level (Supplementary Table 6). To test how normal and tumor epithelial cell hierarchies differentially depend on paracrine signals, we cultured the P009T and P013T cancer organoids, as well as normal colon epithelial organoids (termed NCO), in medium containing Wnt, R-Spondin, and EGF (WRE medium). This medium maintains normal intestinal stem cells depending on paracrine Wnt signals. Alternatively, we cultured the lines for three days in medium lacking Wnt and R-Spondin (E medium). We first determined activities of cell type-specific signatures in the organoid transcriptomes (Fig. 6B). A transcriptional signature for intestinal stem cells expressing *LGR5*, termed LGR5-ISC[12], marked a fraction of cells of NCO organoids cultured in WRE medium, but expression was lost in E medium. Stem cell signature gene expression was much higher, and independent of Wnt/R-Spondin, in the CRC organoids. Signature genes for differentiated absorptive enterocytes or secretory goblet cells showed opposite patterns to stem cell-related gene expression in NCO but were expressed only at low levels in few P009 and P013 cells. Taken together, these expression patterns are in line with Wnt-dependent stem cell maintenance in normal tissue organoids, and Wnt-independent stem cell maintenance and block of terminal differentiation in cancer organoids with APC mutations. The data however do not show whether graded developmental trajectories exist in CRC.

We, therefore, metabolically labelled RNAs of the organoids by 4-thio-uridine, before dissociation and single cell sequencing (SLAM-scSeq; see Fig. 6C for experimental workflow)[21]. This allowed us to distinguish nascent labelled from older non-labelled mRNA and to order cells along inferred latent time

based on dynamic RNA expression (also known as RNA velocity). When cultured in WRE medium, developmental trajectories defined by RNA velocity initiated in areas of maximal LGR5-ISC signature scores and terminated in a region of apoptotic cells characterized by high proportions of mitochondrial reads (Fig. 6D). When cultured without Wnt/R-Spondin, NCO normal colon organoids lost uniform direction of RNA velocity, in line with the loss of stem cell-related gene expression observed above. In contrast, the P009T and P013T cancer organoids maintained strong transcriptional trajectories in E medium.

We ordered the organoid transcriptomes along latent time and assessed strengths of oncogenic signals (Fig 6E). In line with the key role of Wnt in stem cell maintenance, normal colon organoids showed a gradient of Wnt/β-catenin target gene activity along latent time when cultured in WRE medium. In contrast, P009T or P013T organoid transcriptomes showed no Wnt/β-catenin-related expression gradient, but a clear gradient of MAPK target gene activity along latent time. We found that *TFF3*, marking secretory differentiation, was graded in both CRC organoid lines along latent time, but *FABP1*, marking absorptive differentiation in the normal colon, was not (Fig. 6F). This suggests that *TFF3*, but not *FABP1*, is a valid marker for endpoints of CRC differentiation trajectories, in line with our prior finding of immature goblet cell-like CRC cells. The proliferation marker *MKI67* was confined to the beginning of the latent time trajectory of normal organoids in Wnt medium but showed extended gradients in CRC organoids. Unexpectedly, both CRC organoid lines displayed gradual loss of *MKI67* expression along latent time only in the absence of Wnt, and likewise, *MMP7* was expressed in a Wnt-dependent manner in both P009T and P013T organoids. In summary, our metabolic RNA labelling approach identifies MAPK signaling as a driving force of CRC developmental trajectories and suggests a role for Wnt as a potential paracrine signal influencing gene expression in APC-deficient CRC cells.

MAPK target gene expression defines CRC differentiation states

As MAPK signaling marked developmental trajectories in CRC organoids, we wanted to assess whether cell types and states are organized along a MAPK gradient in primary CRC. As a benchmark indicating developmental dynamics in the colon epithelium, we employed the LGR5-ISC stem cell signature[12]. We assigned SCN-aberrant tumor epithelial cell transcriptomes to 40 bins along a gradient of diminishing LGR5-ISC gene signature activity or along decreasing MAPK activity, and assessed expression of the stem cell markers *LGR5*, and *EPHB2*[12] along the gradients. As expected, both *LGR5* and *EPHB2* were graded with LGR5-ISC activity; but MAPK activity also ordered tumor cells in agreement with both cell hierarchy markers and performed even better in sorting cells along a gradient of *EPHB2* expression.

We conclude that cell hierarchies in CRC can be defined by the LGR5-ISC signature, but also by MAPK target gene expression.

We next investigated cell type distribution among SCN-aberrant tumor cells in the bins defined by an LGR5-ISC or the MAPK gradient. LGR5-ISC activity placed a higher proportion of stem/TA-like tumor cells at the beginning of the gradient, whereas tumor cells assigned as goblet cells aggregated in the lower end of the gradient, and TC1-4 cells were broadly distributed across the complete LGR5-ISC activity range (Fig. 7B). Sorting of tumor cells along a gradient of MAPK target gene activity placed a higher proportion of TC1 and TC4 cells at the start of the gradient (p=8,67E-23 and p=7,61E-20, respectively, adjusted Pearson´s chi-squared p-value), while stem/TA- and immature goblet cell-like tumor cells aggregated at the lower end of MAPK activity (p=1,32E-19 and p=1,11E-8, respectively).

Finally, we determined patient-specific differences among SCN-aberrant tumor cells (Fig. 7C). In line with the previous analyses, P007T showed the highest MAPK activity, and the highest proportion of TC4 cells. But TC4 cells, along with TC1 cells, had also the highest average MAPK activity in most other CRCs, including P008, P012, P013, and P016. Stem-/TA-like tumor cells were present in P009, P012, P013, P017, P021 and P025 and had relatively low expression of ERK/MAPK targets in all of these cancers. Stem/TA-like tumor cells of cluster 3 (dark blue) dominated P021 and P025 which had the lowest MAPK target activities. Immature goblet cell-like tumor cells were most enriched in the inflammation-associated tumor P008 and in P017 and showed relatively low MAPK activity. While our data show a clear association between MAPK activity and differentiation state of CRC tumor cells, there was no clear association between the presence of individual oncogenes, tumor stage or other clinical parameters with tumor cell differentiation or MAPK activity. In combination with our metabolic labeling experiments of CRC organoids, our analysis indicates that MAPK signaling is a key determinant of cell plasticity in CRC.

Discussion

Intratumor heterogeneity and cell type composition are a key determinants for treatment response of cancer. In this study, we used single cell omics to investigate the cellular composition and developmental trajectories of tumor cells in colorectal cancer ecosystems. We find that tumor cells, distinguished by somatic copy number aberrations, organize into clusters that either show tumor-specific expression patterns (termed here TC1-4) or resemble stem, TA or immature goblet cells. TC1-4 tumor cell clusters were distinguished by gene expression patterns related to replicative stress and oncogenic signaling pathway activities. We provide evidence for Wnt-independent intrinsic MAPK gradients as organizers of tumor cell plasticity, and Wnt-dependent expression of markers of the invasive front. We identify stromal and immune cell types enriched in the tumor microenvironment, including CAFs that were found specifically in the tumor samples. Our data suggest that MAPK, Wnt and other paracrine signals are defining factors shaping the CRC ecosystem.

Recent single cell studies defined features of the CRC immune microenvironment[16,17]; however, no consensus exists on how to subtype CRC on the single cell level. Here, we subgroup the epithelial tumor component of CRC into tumor-specific TC1-4, stem/TA-like, and goblet-like cells (Fig. 7D). Our analyses suggest that CRC cells with lower levels of MAPK activity resemble, and cluster with, normal stem, TA or immature goblet cells. Stem cell-like CRC cells are characterized by high expression of LGR5-ISC signature genes. Intestinal stem cells and the LGR5-ISC signature are known to be Wnt/β-catenin-driven[43]; however, β-catenin target genes were similarly active throughout the CRC cell clusters, suggesting additional levels of regulation. TC1 and TC4 cells have high levels of MAPK/ERK-dependent transcription, and TC1 cells are additionally associated with DNA damage. Indeed, high MAPK activity is known to cause DNA damage and replicative stress [44]. TC4 cells display high YAP transcriptional activity in addition to high MAPK levels. YAP, an effector of the Hippo pathway, is a key driver of CRC and other cancers [41]. It is of note that the CMS subtyping system developed for bulk tissue CRC transcriptomes [45] could not distinguish the CRC cell types that we identified here on the single cell level, as most epithelial cancer cells were assigned to CMS1 or CMS2 with the exception of goblet-like cells that can adopt CMS3 (Supplementary Fig. 10).

Our study identified multiple cancer traits in tumor cell clusters with relevance to therapy. For instance, TC1 cells were defined by high levels of replication stress, and tumors with high TC1 cell content were strongly positive for PARP, an important therapeutic target in many cancers [46]. MAPK activity is a key pathway for targeted therapy, as many CRC patients profit from anti-EGFR or anti-EGFR/anti-BRAF therapy blocking MAPK [47–49]. We show here that MAPK activity and CRC tumor cell differentiation states are linked. This suggests that CRC cell types identified here may also have intrinsically different thresholds to targeted therapies blocking MAPK. CRC cell types defined by oncogenic signal activities

and differentiation states may therefore provide opportunities to study mechanisms of resistance defined by CRC cell plasticity. Recent studies already point to roles of paracrine signals and the microenvironment in anti-EGFR therapy resistance of CRC [50,51]. Our CRC cell classification system could therefore aid in building a comprehensive subtyping system for CRC on the single cell level with therapeutic relevance.

Stem cells, often defined as a pool of cells sustained at the apex of cell hierarchies or developmental trajectories, are traditionally seen as unique drivers of tissue homeostasis and regeneration of normal tissue, but also of metastasis and therapy-resistance in cancer[52]. However, recent studies have shown metastasis and stemness can be uncoupled in colorectal cancer metastasis[53], and that oncogenic mutations and paracrine signals can reverse developmental trajectories so that more differentiated cells can regain stem cell characteristics[54–56]. Here, we show that a gradient of ERK target gene expression is associated with developmental latent time in CRC organoids, extending our previous finding of graded ERK activity in CRC organoids[57]. The results imply a capacity for intrinsic regulation of CRC cell lifetimes via MAPK. However, further functional studies are required to mechanistically dissect how MAPK and other signaling pathways interact to control CRC developmental trajectories. We suggest that patient-specific sets of genomic aberrations, oncogenic mutations and paracrine signals will all play roles in modulating cell development. Indeed, we see that target genes of further pathways, such as Wnt/β-catenin, Hippo/YAP or TGF-β are expressed in heterogeneous or graded fashion in CRC epithelium. These pathways could represent novel vulnerabilities of cancer cells that could be exploited by directing developmental trajectories towards endpoints of cellular lifespans.

The extension of single-cell analyses of CRC and other cancers to multi-omics, taking also in account genetic and epigenetic heterogeneity[58,59], as well as assessment of cell types and signaling network states while preserving spatial information, promises to identify cell plasticity and genetic diversity of cancer at a cellular resolution. Such approaches have the potential to improve the molecular understanding of cancer and therapy prediction for patients [60]. Our work defining the transcriptional landscape of CRC contributes to this goal.

Figure Legends

Figure 1: Generation and initial assignment of CRC single-cell RNA sequencing data. A Clinical data for the twelve patients under investigation. For histology, see Supplementary Fig. 2. For mutational data, see Supplementary table 6. Loc (Localisation): C: cecum; S: sigmoid colon; T: transverse colon; A: ascending colon; R: rectum. Prog (Predicted Progression): S: via serrated precursor; I: inflammatory/colitis-associated; C: canonical. Tissues used for single-cell RNA sequencing: N: Normal; T: Tumor; O: Organoid. **B, C** UMAPs of single cell transcriptome data, using 10 principal components. **B** UMAPs of epithelial, immune, and stromal cell transcriptomes, color-coded by patient identity. **C** UMAPs of epithelial, immune, and stromal cell transcriptomes, color-coded by tissue of origin.

Figure 2: Cell type census in normal colon and CRC. A UMAPs of single cell transcriptome data, using top ten UMAP components. UMAPs of epithelial, immune and stromal cells, separated by tissue of origin. Color code for cell type assignment. **B** Relative fractions of epithelial, immune and stromal cell types across all patient-derived libraries. For fractions per patient, see Supplementary Figure 5. **C** Single cell gene expression of key immune cell markers. Heatmap shows log2 gene activity values per cell. Color code for cell clusters as in A. **D** Assignment of immune cell clusters by single cell sequencing of tag-based antibodies against immune cell surface markers (Total-Seq), using enriched immune cells of patient P026. Heatmap shows log2 average sequencing depth per antibody tag. Yellow: high sequencing depth, blue: low sequencing depth.

Figure 3: Assignment of epithelial tumor cells by somatic copy number alterations. A Heatmap of SCNAs inferred from single-cell gene expression data. Red: copy number increases (amplifications), Blue: copy number decreases (loss-of-heterozygosity, deletions). To the left: Clustering of cells, color-coded by patient, copy-number-associated cluster, cell type. Copy-number normal cluster: grey; copy number-aberrant cluster: red. **B** Distribution of copy number aberrant cells. Color code as in A. **C** Cell type assignment for copy number normal versus aberrant cells. **D** Single cell gene expression of stem cell marker *OLFM4*, proliferative marker *MKI67*, differentiated absorptive cell marker *FABP1*, and secretory cell marker *TFF3* in UMAP space. **E** Immunofluorescence analysis for OLFM4, MKI67, FABP1, and TFF3 in normal and tumor tissue. All sections are from patient P009, except the EPCAM/OLFM4 co-staining that was done on tumor tissue of P016. Scale bars indicate 100µm

Figure 4: Patient-specific gene expression patterns. A UMAPs using top 50 UMAP components of epithelial cell transcriptomes, color-coded by tissue of origin, cell type, or patient, as indicated. **B** UMAPs of epithelial cells of the TC1-4 tumor cell clusters, using top 50 UMAP components. Cells color-coded by cell type, TC subcluster or patient, as indicated. **C** Assignment of TC subclusters to patients. Black: low cell count; Yellow: high cell count. D Heatmap of Cluster-specifically expressed genes. Genes

assessed by immunohistochemistry or immunofluorescence in this study as marked in bold. Red and blue indicate high and low expression, respectively. **E** Immunohistochemistry on patient tumor sections for selected genes found expressed in cluster-specific manner, as indicated. Scale bars 100µm.

Figure 5. Tumor-specific epithelial cell clusters are distinguished by signaling pathway activities. **A** Transcriptional activity associated with key oncogenic traits and signals, by tumor-specific cell type and patient, as indicated. Red: High activity, blue: low activity. For activities in all normal and tumor epithelial cell types, see Supplementary Figure 9. **D** Visualization of oncogenic traits and signals in tumor cell transcriptome UMAP, as in A. **C** Cell cycle distribution of TC1-4 epithelial tumor cells, as inferred from single cell transcriptomes. **D** Immunofluorescence of DNA-damage-associated nuclear protein PARP. Image shows adjacent normal and tumor crypts of tissue P009T, marked by N and T, respectively. Scale bar 100µm.

Figure 6: RNA metabolic labeling defines tumor cell trajectories in patient-derived organoids. **A** Phenotypes of patient-derived organoid lines P009T and P013T. **B** UMAPs of organoid single cell transcriptomes. Organoid lines and medium conditions as indicated. LGR5-ISC stem cell, enterocyte and Goblet cell signatures are visualized. Grey: No gene activity; purple: low gene activity; yellow, high gene activity. **C** Schematic representation of SLAM-Seq workflow to infer RNA dynamics ("RNA velocity"). In short, organoids were passaged and assigned to different medium conditions. After three days, nascent RNA was metabolically labelled for 3h using 4sU. Organoids were harvested, dissociated, and fixed. RNA in single cells was alkylated, and cells were subjected to single cell sequencing. Reads were assigned to nascent or old RNA status, depending on diagnostic T-C mutational status. **D** Developmental trajectories inferred from RNA metabolic labelling. Bold and thin arrows indicate high versus low directionality of RNA velocity. Colors below RNA velocity encode cell plasticity, used to infer trajectory starts. Small inlays show color code for mitochondrial reads, indicating apoptosis near the trajectory ends. **E** Activities of Wnt/β-catenin and MAPK target genes in organoid single cell transcriptomes, ordered along latent time. **F** Activities of *TFF3*, *FABP1*, *MKI67* and *MMP7* in organoid single cell transcriptomes, ordered along latent time. Color code for panels D-F: Red: high activity; blue: low activity.

Figure 7: MAPK activity is linked to CRC cell differentiation states. **A** Gene expression of *LGR5* and *EPHB2*, along activity gradients of LGR5-ISC or ERK/MAPK transcriptional signatures. **B** Cell-type distribution of SCN-aberrant CRC cells along gradients of LGR5-ISC or MAPK transcriptional signatures, as in A. **C** Cell type distribution of SCN-aberrant CRC tumor cells along MAPK signature activity, as in B, per tumor. **D** Model of CRC cell plasticity along a gradient of MAPK activity. Low MAPK is associated with CRC cells resembling stem/TA or immature goblet cells. Among these, stem-like CRC cells have

high expression of the LGR5-ISC signature. High MAPK is associated with TC1 and TC4 cells that are further characterized by high replication stress and YAP activity, respectively.

Funding

The work was in part funded by the Berlin Institute of Health (to PB, NB, CS, DH and MM), the German Cancer Consortium DKTK (to NB and MM) and the Helmholtz Association (RFS). TTW was supported by the DFG graduate school grant GRK2424/1.

References

1. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell 2011;144:646–674.

2. Sever R, Brugge JS. Signal transduction in cancer. Cold Spring Harb Perspect Med 2015.

3. Binnewies M, Roberts EW, Kersten K, et al. Understanding the tumor immune microenvironment (TIME) for effective therapy. Nat Med 2018.

4. Quail DF, Joyce JA. Microenvironmental regulation of tumor progression and metastasis. Nat Med 2013.

5. Cunha BR da, Domingos C, Buzzo Stefanini AC, et al. Cellular interactions in the tumor microenvironment: The role of secretome. J Cancer 2019.

6. Fearon ER. Molecular genetics of colorectal cancer. Annu Rev Pathol 2011;6:479–507.

7. Palma FDE De, D'argenio V, Pol J, et al. The molecular hallmarks of the serrated pathway in colorectal cancer. Cancers (Basel) 2019.

8. Lasry A, Zinger A, Ben-Neriah Y. Inflammatory networks underlying colorectal cancer. Nat Immunol 2016.

9. Ricci-Vitiani L, Lombardi DG, Pilozzi E, et al. Identification and expansion of human colon-cancer-initiating cells. Nature 2007;445:111–115.

10. O'Brien CA, Pollett A, Gallinger S, et al. A human colon cancer cell capable of initiating tumour growth in immunodeficient mice. Nature 2007;445:106–110.

11. Shimokawa M, Ohta Y, Nishikori S, et al. Visualization and targeting of LGR5+ human colon cancer stem cells. Nature 2017;545:187–192.

12. Merlos-Suárez A, Barriga FM, Jung P, et al. The intestinal stem cell signature identifies colorectal cancer stem cells and predicts disease relapse. Cell Stem Cell 2011;8:511–524.

13. Brabletz T, Jung A, Dag S, et al. β-catenin regulates the expression of the matrix metalloproteinase-7 in human colorectal cancer. Am J Pathol 1999.

14. Vermeulen L, Sousa E Melo F De, Heijden M van der, et al. Wnt activity defines colon cancer stem cells and is regulated by the microenvironment. Nat Cell Biol 2010;12:468–476.

15. James KR, Gomes T, Elmentaite R, et al. Distinct microbial and immune niches of the human colon. Nat Immunol 2020.

16. Zhang L, Li Z, Skrzypczynska KM, et al. Single-Cell Analyses Inform Mechanisms of Myeloid-

Targeted Therapies in Colon Cancer. Cell 2020.

17. Lee H-O, Tejpar S, Woong-Yang PArk. Lineage-dependent gene expression programs influence the immune landscape of colorectal cancer. Nat Genet 2020.

18. Adam M, Potter AS, Potter SS. Psychrophilic proteases dramatically reduce single-cell RNA-seq artifacts: A molecular atlas of kidney development. Dev 2017.

19. Mamlouk S, Childs LH, Aust D, et al. DNA copy number changes define spatial patterns of heterogeneity in colorectal cancer. Nat Commun 2017;8:14093.

20. Sato T, Stange DE, Ferrante M, et al. Long-term expansion of epithelial organoids from human colon, adenoma, adenocarcinoma, and Barrett's epithelium. Gastroenterology 2011;141:1762–1772.

21. Herzog VA, Reichholf B, Neumann T, et al. Thiol-linked alkylation of RNA to assess expression dynamics. Nat Methods 2017.

22. Manno G La, Soldatov R, Zeisel A, et al. RNA velocity of single cells. Nature 2018.

23. Hafemeister C, Satija R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. bioRxiv 2019.

24. Stuart T, Butler A, Hoffman P, et al. Comprehensive Integration of Single-Cell Data. Cell 2019.

25. Smillie CS, Biton M, Ordovas-Montanes J, et al. Intra- and Inter-cellular Rewiring of the Human Colon during Ulcerative Colitis. Cell 2019.

26. Liberzon A, Birger C, Thorvaldsdóttir H, et al. The Molecular Signatures Database Hallmark Gene Set Collection. Cell Syst 2015.

27. Köster J, Rahmann S. Snakemake-a scalable bioinformatics workflow engine. Bioinformatics 2012;28:2520–2522.

28. Döring A, Weese D, Rausch T, et al. SeqAn an efficient, generic C++ library for sequence analysis. BMC Bioinformatics 2008.

29. Wolf FA, Angerer P, Theis FJ. SCANPY: Large-scale single-cell gene expression data analysis. Genome Biol 2018.

30. Bergen V, Lange M, Peidli S, et al. Generalizing RNA velocity to transient cell states through dynamical modeling. bioRxiv 2019.

31. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and

population genetical parameter estimation from sequencing data. Bioinformatics 2011.

32. Favero F, Joshi T, Marquard AM, et al. Sequenza: Allele-specific copy number and mutation profiles from tumor sequencing data. Ann Oncol 2015.

33. Wang Y, Song W, Wang J, et al. Single-cell transcriptome analysis reveals differential nutrient absorption functions in human intestine. J Exp Med 2020.

34. McInnes L, Healy J, Saul N, et al. UMAP: Uniform Manifold Approximation and Projection. J Open Source Softw 2018.

35. Parikh K, Antanaviciute A, Fawkner-Corbett D, et al. Colonic epithelial cell diversity in health and inflammatory bowel disease. Nature 2019.

36. Lau W De, Barker N, Low TY, et al. Lgr5 homologues associate with Wnt receptors and mediate R-spondin signalling. Nature 2011.

37. Yan KS, Janda CY, Chang J, et al. Non-equivalence of Wnt and R-spondin ligands during Lgr5 + intestinal stem-cell self-renewal. Nature 2017.

38. Haramis A-PG, Begthel H, Born M van den, et al. De novo crypt formation and juvenile polyposis on BMP inhibition in mouse intestine. Sci (New York, NY) 2004;303:1684–1686.

39. Sahai E, Astsaturov I, Cukierman E, et al. A framework for advancing our understanding of cancer-associated fibroblasts. Nat Rev Cancer 2020.

40. Serra D, Mayr U, Boni A, et al. Self-organization and symmetry breaking in intestinal organoid development. Nature 2019.

41. Zanconato F, Cordenonsi M, Piccolo S. YAP/TAZ at the Roots of Cancer. Cancer Cell 2016.

42. Schubert M, Klinger B, Klünemann M, et al. Perturbation-response genes reveal signaling footprints in cancer gene expression. Nat Commun 2018.

43. Wetering M van de, Sancho E, Verweij C, et al. The beta-catenin/TCF-4 complex imposes a crypt progenitor phenotype on colorectal cancer cells. Cell 2002;111:241–250.

44. Sheu JJC, Guan B, Tsai FJ, et al. Mutant BRAF induces DNA strand breaks, activates DNA damage response pathway, and up-regulates glucose transporter-1 in nontransformed epithelial cells. Am J Pathol 2012.

45. Guinney J, Dienstmann R, Wang X, et al. The consensus molecular subtypes of colorectal cancer. Nat Med 2015;21:1350–1356.

46. Sun C, Fang Y, Labrie M, et al. Systems approach to rational combination therapy: PARP inhibitors. Biochem Soc Trans 2020.

47. Karapetis CS, Khambata-Ford S, Jonker DJ, et al. K-ras mutations and benefit from cetuximab in advanced colorectal cancer. N Engl J Med 2008;359:1757–1765.

48. Amado RG, Wolf M, Peeters M, et al. Wild-type KRAS is required for panitumumab efficacy in patients with metastatic colorectal cancer. J Clin Oncol 2008;26:1626–1634.

49. Kopetz S, Grothey A, Yaeger R, et al. Encorafenib, binimetinib, and cetuximab in BRAF V600E–mutated colorectal cancer. N Engl J Med 2019.

50. Woolston A, Khan K, Spain G, et al. Genomic and Transcriptomic Determinants of Therapy Resistance and Immune Landscape Evolution during Anti-EGFR Treatment in Colorectal Cancer. Cancer Cell 2019.

51. Misale S, Nicolantonio F Di, Sartore-Bianchi A, et al. Resistance to Anti-EGFR therapy in colorectal cancer: From heterogeneity to convergent evolution. Cancer Discov 2014;4:1269–1280.

52. Lytle NK, Barber AG, Reya T. Stem cell fate in cancer growth, progression and therapy resistance. Nat Rev Cancer 2018.

53. Fumagalli A, Oost KC, Kester L, et al. Plasticity of Lgr5-Negative Cancer Cells Drives Metastasis in Colorectal Cancer. Cell Stem Cell 2020.

54. Schwitalla S, Fingerle AA, Cammareri P, et al. Intestinal Tumorigenesis Initiated by Dedifferentiation and Acquisition of Stem-Cell-like Properties. Cell 2012.

55. Buczacki SJA, Zecchini HI, Nicholson AM, et al. Intestinal label-retaining cells are secretory precursors expressing Lgr5. Nature 2013;495:65–69.

56. Jadhav U, Saxena M, O'Neill NK, et al. Dynamic Reorganization of Chromatin Accessibility Signatures during Dedifferentiation of Secretory Precursors into Lgr5+ Intestinal Stem Cells. Cell Stem Cell 2017.

57. Brandt R, Sell T, Lüthen M, et al. Cell type-dependent differential activation of ERK by oncogenic KRAS in colon cancer and intestinal epithelium. Nat Commun 2019;10:2919.

58. Bian S, Hou Y, Zhou X, et al. Single-cell multiomics sequencing and analyses of human colorectal cancer. Science (80-) 2018.

59. Roerink SF, Sasaki N, Lee-Six H, et al. Intra-tumour diversification in colorectal cancer at the

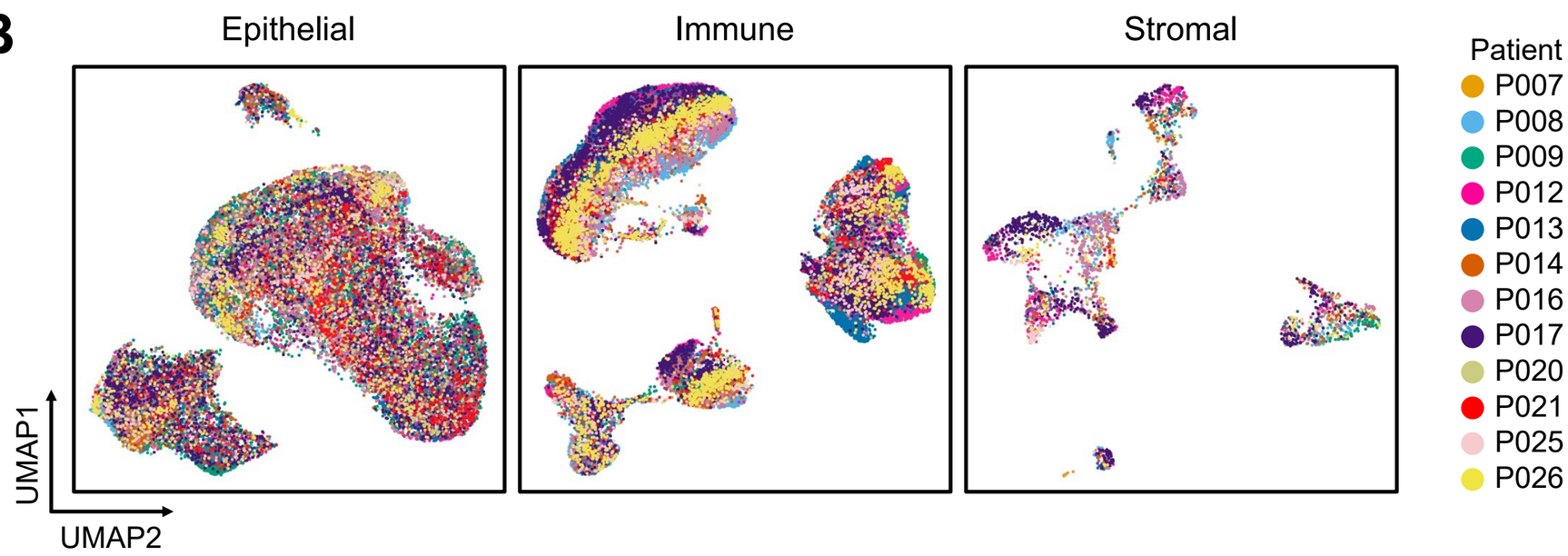
single-cell level. Nature 2018.

60. Rajewsky N, Almouzni G, Gorski SA, et al. LifeTime and improving European healthcare through cell-based interceptive medicine. Nature 2020.

A

Patient	Mutations	Stage	Grade	MS	Loc	Prog	Sex	Tissues
P007	$BRAF^{V600E}$, $TP53^{R175H}$	T2N1b	G3	MSS	C	S	m	N/T
P008	$TP53^{R175H}$	T4aN2b	G3	MSS	C	I	m	N/T
P009	APC^{LOH}, $FBXW7^{D480H}$, $TP53^{R248Q}$	T2N0	G2	MSS	S	C	m	2N/2T/O
P012	n/a	TisN0	G2	MSS	T	?	m	N/T
P013	APC^{LOH}, $NOTCH1^{H1601L}$	T4aN1bM1	G2	MSS	A	C	m	N/T/O
P014	$APC^{LOH,\ 1556insN}$, $BRAF^{V600E}$, $HRAS^{A59T}$, $PIK3CA^{E545K}$, $TGFBR2^{D547H}$	T3N2b	G3	MSS	A	?	f	N/T
P016	$APC^{LOH,\ 1396delF}$, $KRAS^{G13C}$, $PIK3CA^{E545K}$, $FBXW7^{R278*}$	T3N0	G2	n/a	R	C	m	N/T
P017	APC^{LOH}, $KRAS^{G13D}$	T3N2b	G3	MSS	A	C	f	N/T
P020	$BRAF^{V600E}$, $TP53^{R282W}$	T2N1a	G2	MSS	R	S	m	N/T
P021	APC^{LOH}, $KRAS^{A146T}$, $DNMT1^{R1384Q}$	T2N0	G2	MSS	D	C	f	N/T
P025	APC^{LOH}, $KRAS^{G13D}$, $TP53^{G245S}$	T3N0	G2	MSS	A	C	m	N/T
P026	$BRAF^{V600E}$, $TP53^{S90fs}$, $AXIN2^{LOH}$, $PI4KA^{A1198T}$	T1N0	G2	MSI	C	S	m	T

B



C

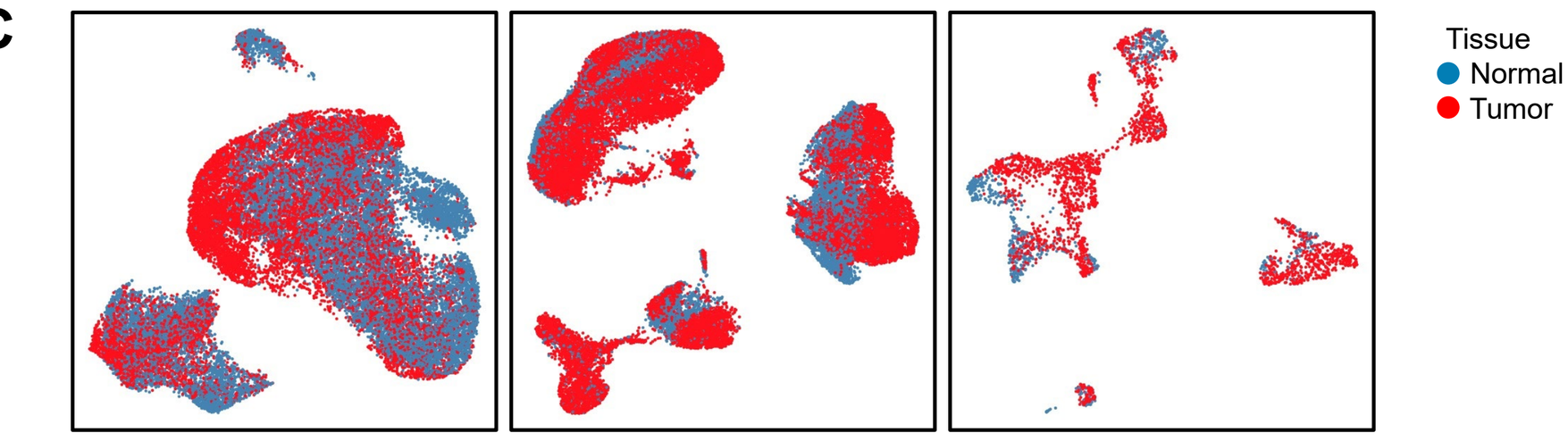


Fig. 1

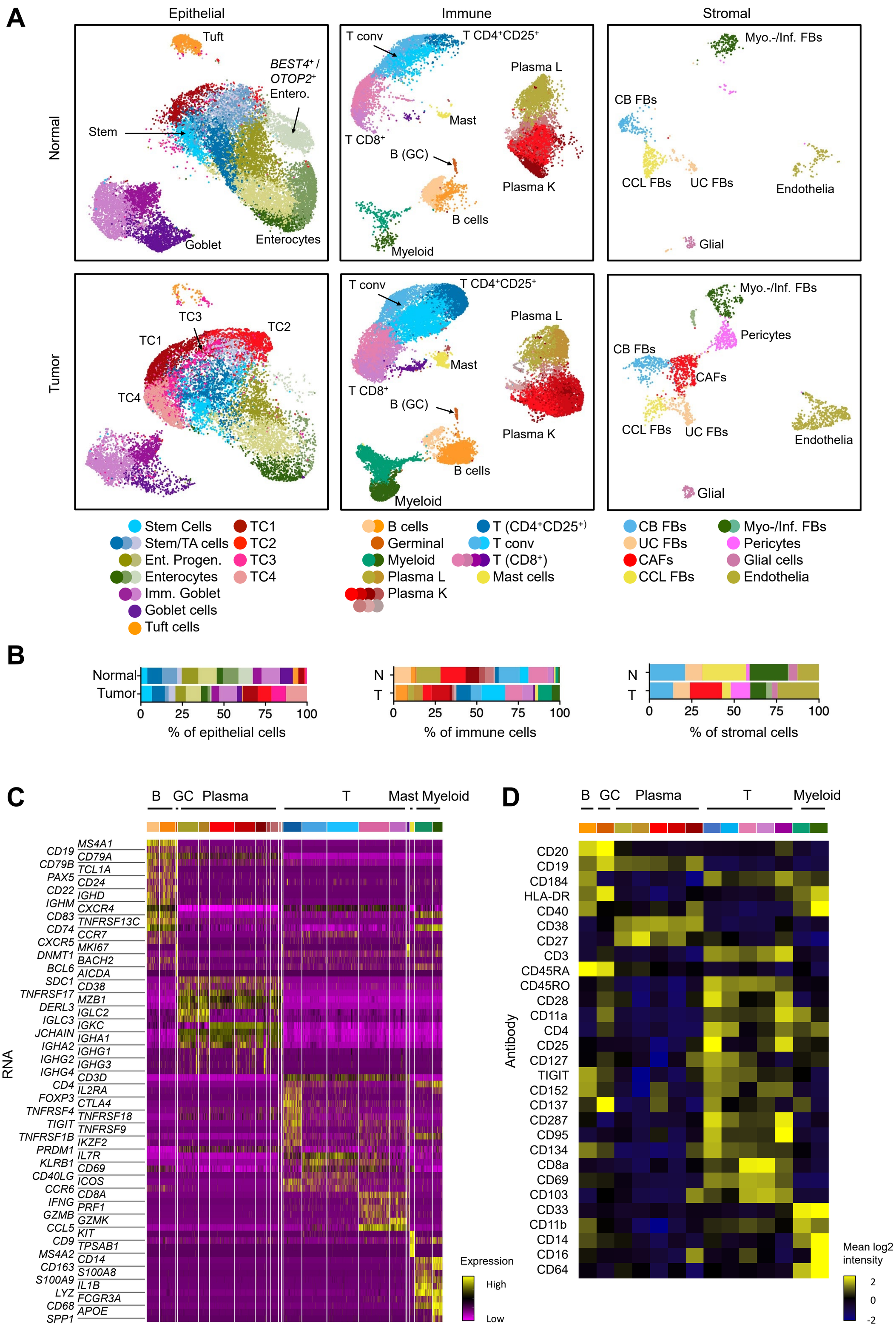

A
Epithelial
Immune
Stromal
Normal
Tumor
Tuft
BEST4+ / OTOP2+ Entero.
Stem
Goblet
Enterocytes
T conv
T CD4+CD25+
Plasma L
Mast
T CD8+
B (GC)
Plasma K
B cells
Myeloid
Myo.-/Inf. FBs
CB FBs
CCL FBs
UC FBs
Endothelia
Glial
TC3
TC2
TC1
TC4
Pericytes
CAFs
Stem Cells
Stem/TA cells
Ent. Progen.
Enterocytes
Imm. Goblet
Goblet cells
Tuft cells
TC1
TC2
TC3
TC4
B cells
Germinal
Myeloid
Plasma L
Plasma K
T (CD4+CD25+)
T conv
T (CD8+)
Mast cells
CB FBs
UC FBs
CAFs
CCL FBs
Myo-/Inf. FBs
Pericytes
Glial cells
Endothelia
B
Normal
Tumor
N
T
0
25
50
75
100
% of epithelial cells
% of immune cells
% of stromal cells
C
B
GC
Plasma
T
Mast
Myeloid
RNA
MS4A1
CD19
CD79A
CD79B
TCL1A
PAX5
CD24
CD22
IGHD
IGHM
CXCR4
CD83
TNFRSF13C
CD74
CCR7
CXCR5
MKI67
DNMT1
BACH2
BCL6
AICDA
SDC1
CD38
TNFRSF17
MZB1
DERL3
IGLC2
IGLC3
IGKC
JCHAIN
IGHA1
IGHA2
IGHG1
IGHG2
IGHG3
IGHG4
CD3D
CD4
IL2RA
FOXP3
CTLA4
TNFRSF4
TNFRSF18
TIGIT
TNFRSF9
TNFRSF1B
IKZF2
PRDM1
IL7R
KLRB1
CD69
CD40LG
ICOS
CCR6
CD8A
IFNG
PRF1
GZMB
GZMK
CCL5
KIT
CD9
TPSAB1
MS4A2
CD14
CD163
S100A8
S100A9
IL1B
LYZ
FCGR3A
CD68
APOE
SPP1
Expression
High
Low
D
B
GC
Plasma
T
Myeloid
Antibody
CD20
CD19
CD184
HLA-DR
CD40
CD38
CD27
CD3
CD45RA
CD45RO
CD28
CD11a
CD4
CD25
CD127
TIGIT
CD152
CD137
CD287
CD95
CD134
CD8a
CD69
CD103
CD33
CD11b
CD14
CD16
CD64
Mean log2 intensity
2
0
-2

Fig. 2

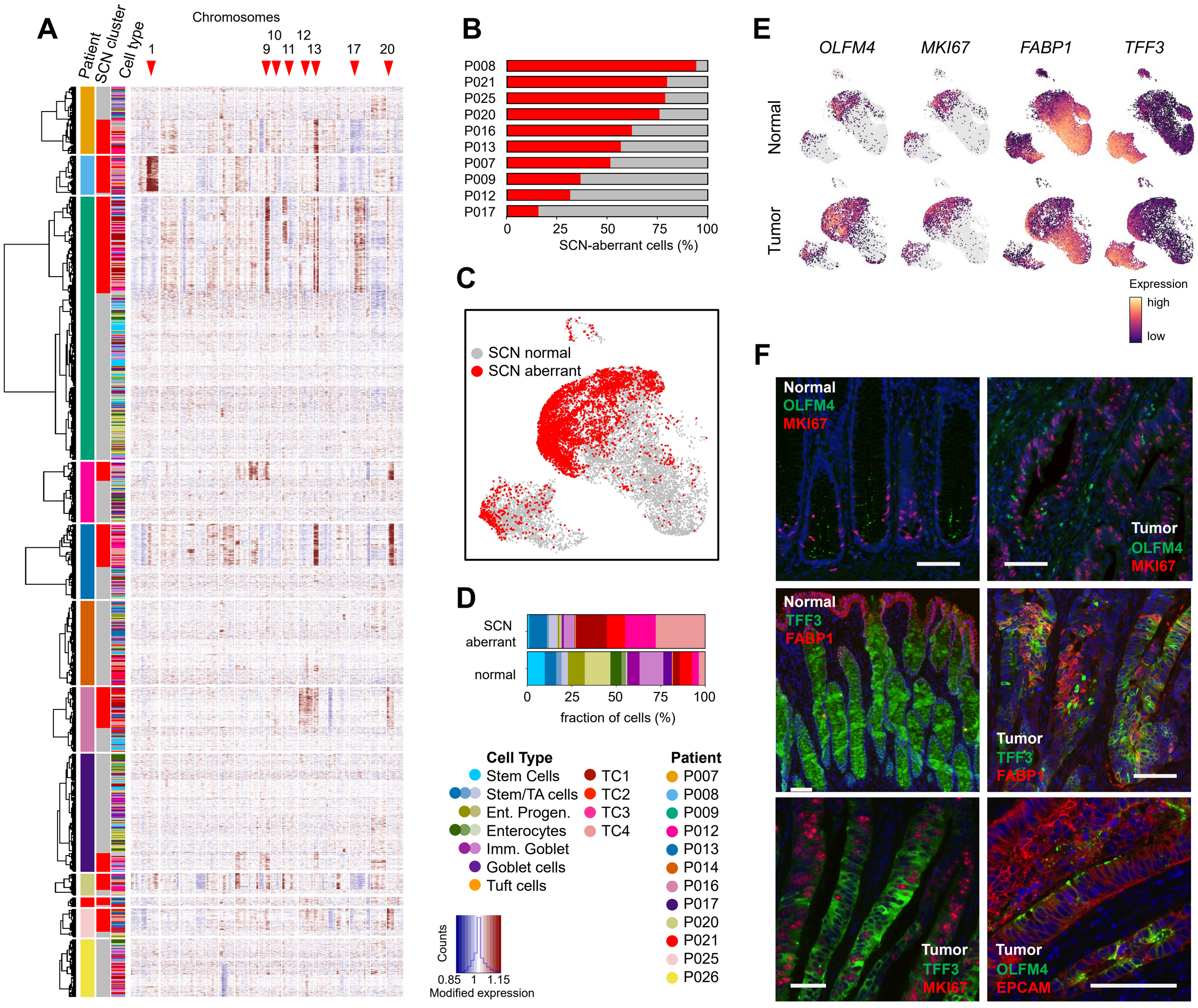

A
Patient
SCN cluster
Cell type
Chromosomes
1
9
10
11
12
13
17
20
B
P008
P021
P025
P020
P016
P013
P007
P009
P012
P017
0
25
50
75
100
SCN-aberrant cells (%)
C
SCN normal
SCN aberrant
D
SCN aberrant
normal
0
25
50
75
100
fraction of cells (%)
Cell Type
Stem Cells
Stem/TA cells
Ent. Progen.
Enterocytes
Imm. Goblet
Goblet cells
Tuft cells
TC1
TC2
TC3
TC4
Patient
P007
P008
P009
P012
P013
P014
P016
P017
P020
P021
P025
P026
Counts
0.85
1
1.15
Modified expression
E
OLFM4
MKI67
FABP1
TFF3
Normal
Tumor
Expression
high
low
F
Normal
OLFM4
MKI67
Tumor
OLFM4
MKI67
Normal
TFF3
FABP1
Tumor
TFF3
FABP1
Tumor
TFF3
MKI67
Tumor
OLFM4
EPCAM

Fig. 3

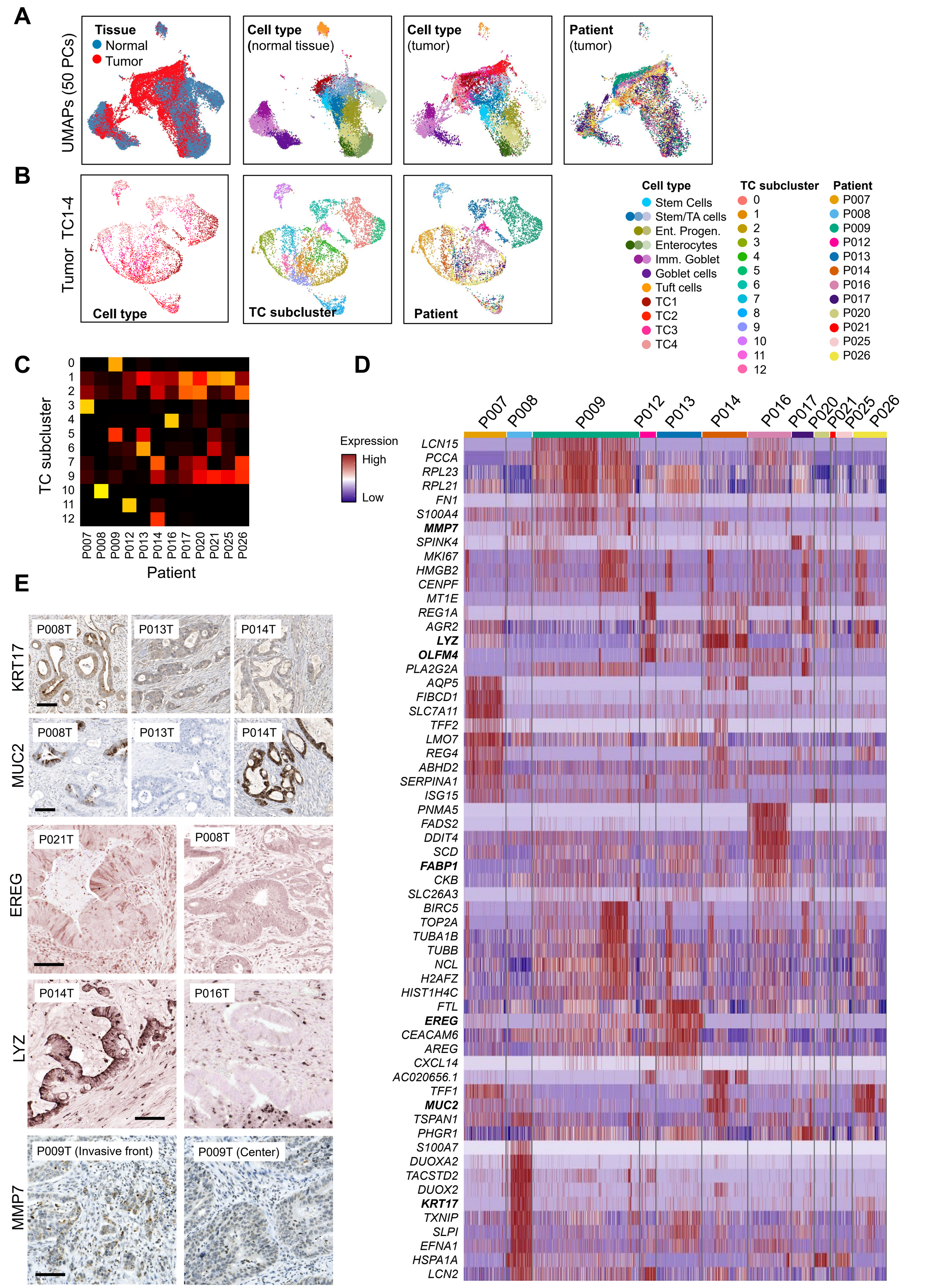

A
UMAPs (50 PCs)
Tissue
Normal
Tumor
Cell type
(normal tissue)
Cell type
(tumor)
Patient
(tumor)
B
Tumor TC1-4
Cell type
TC subcluster
Patient
Cell type
Stem Cells
Stem/TA cells
Ent. Progen.
Enterocytes
Imm. Goblet
Goblet cells
Tuft cells
TC1
TC2
TC3
TC4
TC subcluster
0
1
2
3
4
5
6
7
8
9
10
11
12
Patient
P007
P008
P009
P012
P013
P014
P016
P017
P020
P021
P025
P026
C
TC subcluster
Patient
D
Expression
High
Low
LCN15
PCCA
RPL23
RPL21
FN1
S100A4
MMP7
SPINK4
MKI67
HMGB2
CENPF
MT1E
REG1A
AGR2
LYZ
OLFM4
PLA2G2A
AQP5
FIBCD1
SLC7A11
TFF2
LMO7
REG4
ABHD2
SERPINA1
ISG15
PNMA5
FADS2
DDIT4
SCD
FABP1
CKB
SLC26A3
BIRC5
TOP2A
TUBA1B
TUBB
NCL
H2AFZ
HIST1H4C
FTL
EREG
CEACAM6
AREG
CXCL14
AC020656.1
TFF1
MUC2
TSPAN1
PHGR1
S100A7
DUOXA2
TACSTD2
DUOX2
KRT17
TXNIP
SLPI
EFNA1
HSPA1A
LCN2
E
KRT17
MUC2
EREG
LYZ
MMP7
P008T
P013T
P014T
P021T
P016T
P009T (Invasive front)
P009T (Center)

Fig. 4

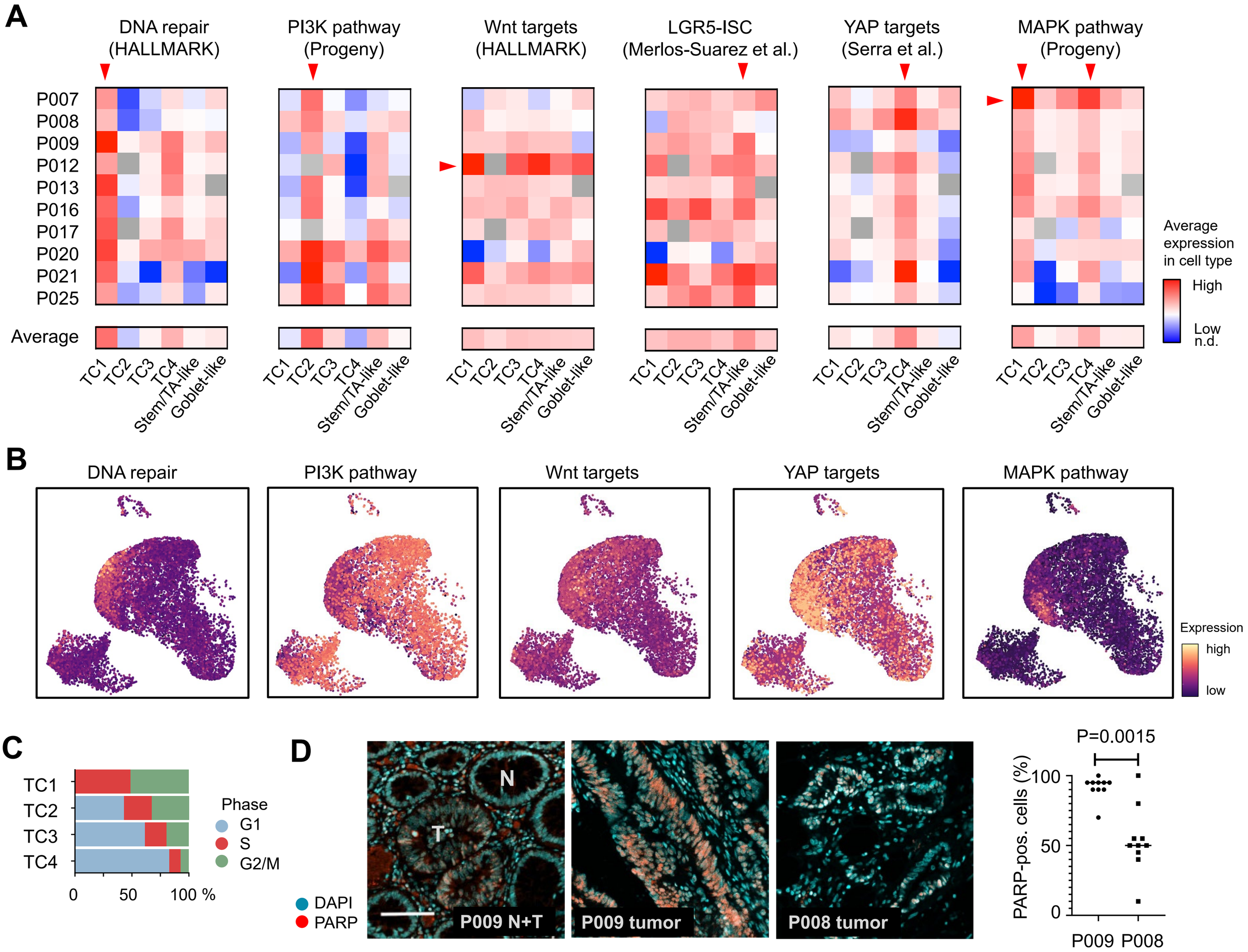

A
DNA repair (HALLMARK)
PI3K pathway (Progeny)
Wnt targets (HALLMARK)
LGR5-ISC (Merlos-Suarez et al.)
YAP targets (Serra et al.)
MAPK pathway (Progeny)
P007
P008
P009
P012
P013
P016
P017
P020
P021
P025
Average
TC1
TC2
TC3
TC4
Stem/TA-like
Goblet-like
Average expression in cell type
High
Low
n.d.
B
DNA repair
PI3K pathway
Wnt targets
YAP targets
MAPK pathway
Expression
high
low
C
TC1
TC2
TC3
TC4
0
50
100 %
Phase
G1
S
G2/M
D
DAPI
PARP
N
T
P009 N+T
P009 tumor
P008 tumor
P=0.0015
PARP-pos. cells (%)
100
50
0
P009
P008

Fig. 5

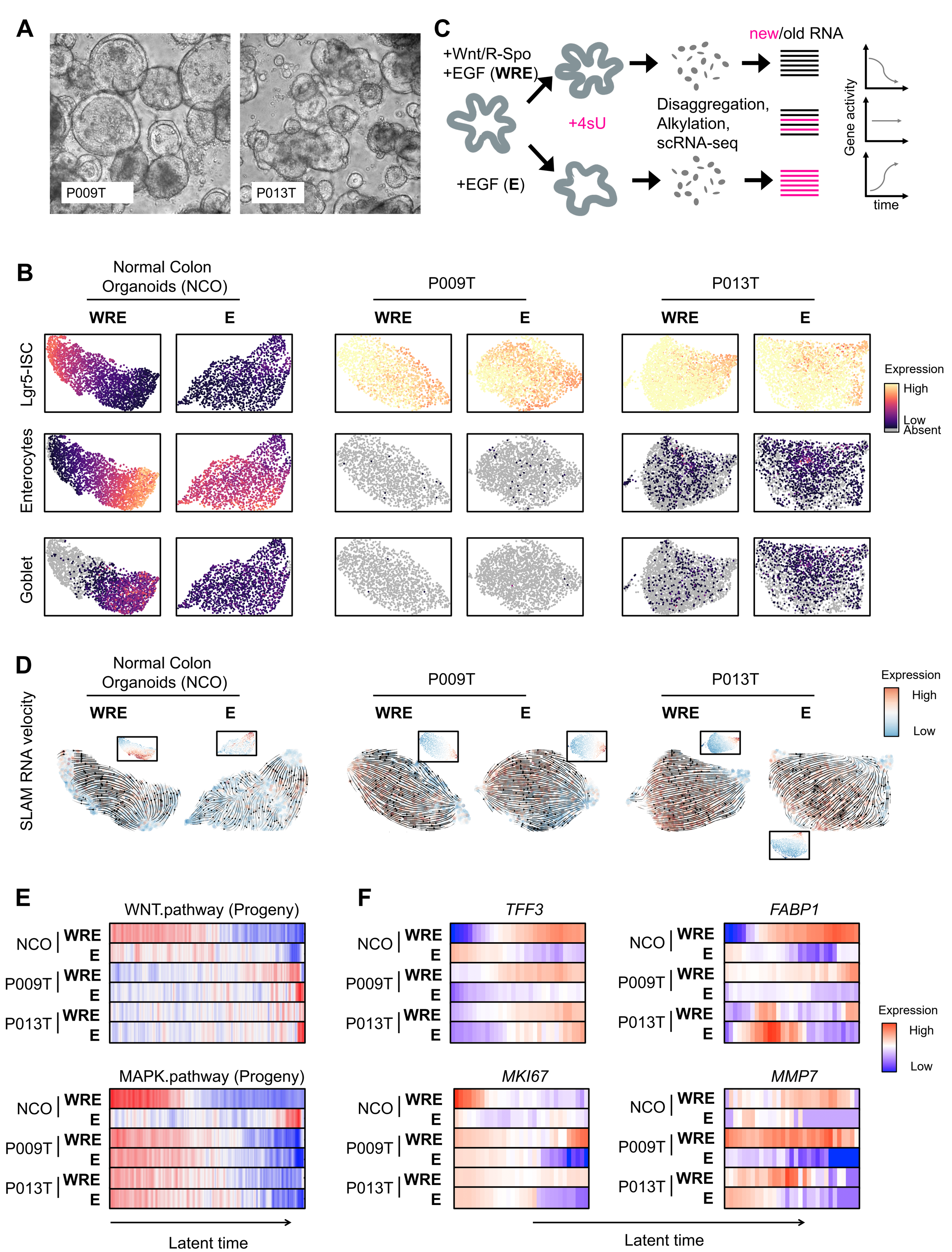

A
P009T
P013T
C
+Wnt/R-Spo
+EGF (WRE)
+4sU
+EGF (E)
Disaggregation,
Alkylation,
scRNA-seq
new/old RNA
Gene activity
time
B
Normal Colon Organoids (NCO)
P009T
P013T
WRE
E
Lgr5-ISC
Enterocytes
Goblet
Expression
High
Low
Absent
D
SLAM RNA velocity
Normal Colon Organoids (NCO)
P009T
P013T
WRE
E
Expression
High
Low
E
WNT.pathway (Progeny)
MAPK.pathway (Progeny)
NCO
P009T
P013T
WRE
E
Latent time
F
TFF3
FABP1
MKI67
MMP7
NCO
P009T
P013T
WRE
E
Expression
High
Low
Latent time

Fig. 6

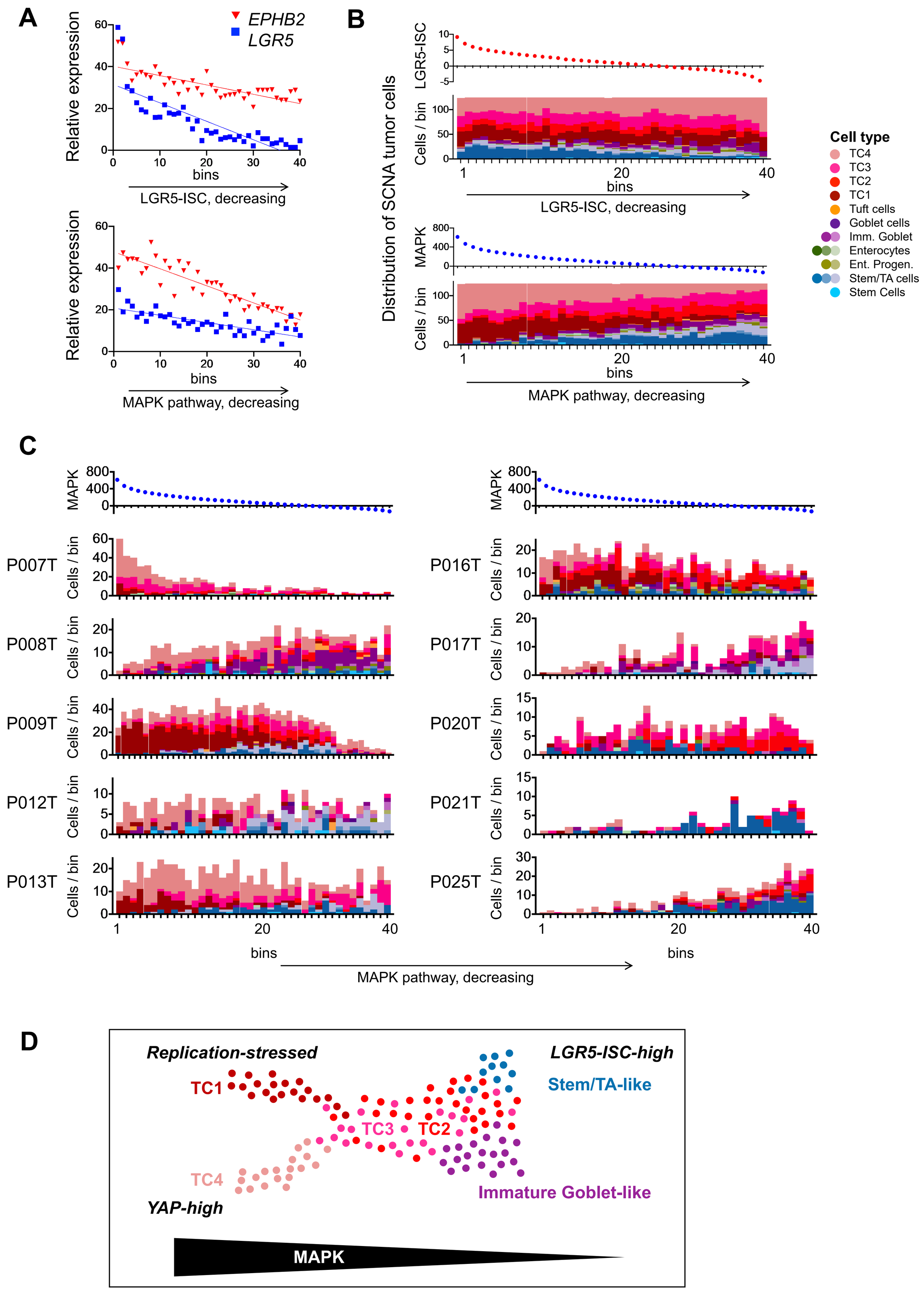

A
EPHB2
LGR5
Relative expression
bins
LGR5-ISC, decreasing
MAPK pathway, decreasing
B
Distribution of SCNA tumor cells
LGR5-ISC
MAPK
Cells / bin
bins
LGR5-ISC, decreasing
MAPK pathway, decreasing
Cell type
TC4
TC3
TC2
TC1
Tuft cells
Goblet cells
Imm. Goblet
Enterocytes
Ent. Progen.
Stem/TA cells
Stem Cells
C
P007T
P008T
P009T
P012T
P013T
P016T
P017T
P020T
P021T
P025T
MAPK pathway, decreasing
D
Replication-stressed
LGR5-ISC-high
Stem/TA-like
TC1
TC3
TC2
TC4
Immature Goblet-like
YAP-high
MAPK

Fig. 7