bioRxiv preprint doi: https://doi.org/10.1101/2020.01.10.901579; this version posted January 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

A census of cell types and paracrine interactions in colorectal cancer

Florian Uhlitz[1,2,3,*], Philip Bischoff[1,*], Anja Sieber[1,2], Benedikt Obermayer[5], Eric Blanc[5], Mareen Lüthen[1,3], Birgit Sawitzki[6], Carsten Kamphues[3,4], Dieter Beule[5], Christine Sers[1,3,7], David Horst[1,3,7], Nils Blüthgen[1,2,3,7,#] and Markus Morkel[1,3,7,#]

[1] Institute of Pathology, Charité Universitätsmedizin Berlin, Charitéplatz 1, 10117 Berlin, Germany

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

[3] German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany

[4] Department of Surgery, Charité Universitätsmedizin Berlin, Hindenburgdamm 30, 12203 Berlin

[5] Core Unit Bioinformatics (CUBI), Charité Universitätsmedizin Berlin, MDC Berlin, and Berlin Institute of Health, Charitéplatz 1, 10117 Berlin

[6] Institute of Medical Immunology, Charité Universitätsmedizin Berlin, Augustenburger Platz 1, 13353 Berlin

[7] Berlin Institute of Health, Anna-Louisa-Karsch-Straße 2, 10178 Berlin, Germany

[*] 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

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.10.901579; this version posted January 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

In colorectal cancer, oncogenic mutations transform a hierarchically organized and homeostatic epithelium into invasive cancer tissue. To define differences in cellular composition between the normal colon and colorectal cancer, and to map potential cellular interactions between tumor cells and their microenvironment, we profiled transcriptomes of >50,000 single cells from tumors and matched normal tissues of eight colorectal cancer patients. We find that tumor formation is accompanied by changes in epithelial, immune and stromal cell compartments in all patients. In the epithelium, we identify a continuum of five tumor-specific stem cell and progenitor-like populations, and persistent multilineage differentiation. We find multiple stromal and immune cell types to be consistently expanded in tumor compared to the normal colon, including cancer-associated fibroblasts, pericytes, monocytes, macrophages and a subset of T cells. We identify epithelial tumor cells and cancer-associated fibroblasts as relevant for assigning colorectal cancer consensus molecular subtypes. Our survey of growth factors in the tumor microenvironment identifies cell types responsible for increased paracrine EGFR, MET and TGF-β signaling in tumor tissue compared to the normal colon. We show that matched colorectal cancer organoids retain cell type heterogeneity, allowing to define a distinct differentiation trajectory encompassing stem and progenitor-like tumor cells. In summary, our single-cell analyses provide insights into cell types and signals shaping colorectal cancer cell plasticity.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.10.901579; this version posted January 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Introduction

All cells in the human body exist in contact with other cells in finely tuned microenvironments. Paracrine communication between cells ensures tissue homeostasis. Cancer cells are compromised in their ability to maintain homeostasis, as the 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 environment[2]. Thirdly, cancer cells are often 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. Interactions between immune, stromal and cancer cells are critical for tumor progression and therapy response[4,5].

Colorectal cancer (CRC) most often 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 signaling cascade[6,7]. Less frequently, CRC can arise by initiating mutations in BRAF, or from chronic inflammation increasing the mutation rate in the tissue. Regardless of the order of oncogenic mutations, genetic CRC drivers have direct and indirect effects on the cellular composition of CRC and its microenvironment. There is substantial evidence for the existence of tumor cell subpopulations in CRC. For instance, cancer stem cells with high clonogenic potential can be sorted from CRC based on surface markers like PROM1 (also known as CD133) or LGR5[8–10]. Furthermore, CRC cells at the invasive front express matrix metalloproteinases such as MMP7 and display epithelial-to-mesenchymal transition, in contrast to cells residing in more central locations in the tumor [11–13]. However, it has not been investigated systematically how cell types and cell plasticity differ between the normal colon and CRC.

Here, we use droplet-based single-cell RNA sequencing to profile cell types and their differentiation states in normal colon and tumor tissues of eight CRC patients, and in matching CRC organoids. We identify consistent changes occurring in epithelial differentiation programs between normal and tumor epithelium in the colon, resulting in the emergence of tumor-specific epithelial cell clusters expressing genes relevant for cancer traits such as stemness, invasion, and epithelial-to-mesenchymal transition. We also catalog cell types expanded in the tumor microenvironment, for instance, stromal cancer-associated fibroblasts and several types of immune cells. We provide evidence for cancer-cell-specific re-wiring of morphogenetic signaling informed by oncogenic mutations and differences in paracrine signaling networks.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.10.901579; this version posted January 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Results

Single-cell RNA sequencing of CRC

To capture the cellular diversity in CRC and track changes from normalcy to disease, we performed single-cell transcriptome analysis of eight previously untreated CRC patients (Fig. 1A). We utilized tissue samples that included the invasive tumor front and matched normal tissues (Supplementary Fig. 1). Tumors under investigation 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. Panel sequencing of genomic tumor DNA uncovered mutations in APC, KRAS and/or TP53 in tumors P008, P009, P013, P016, and P017; these mutations are characteristic for the canonical CRC progression pathway initiated by loss of APC. P007 harbored $BRAF^{V600E}$ and TP53 mutations; this mutational pattern is in line with a tumor initiated by BRAF activation. P008 carried a TP53 mutation and was colitis-associated. Tumor P014 contained putative driver mutations in APC, BRAF, HRAS, and PIK3CA, albeit at a lower frequency, suggesting the possibility of distinct subclones contributing to this tumor.

We enzymatically dissociated the fresh normal and cancer tissues to single cells, produced single-cell transcriptome libraries from each tissue using a commercial droplet-based system, and sequenced the libraries to obtain transcriptomes covering 500 to 5,000 genes per cell. Singe-cell profiles were partitioned into epithelial, immune or stromal transcriptome subsets for each library, using known marker genes[14], and then merged. We observed varying fractions of cell types per library, but stromal cells were generally less abundant (in total across all libraries: >25,000 epithelial cells; >25,000 immune cells and 2,691 stromal cells). Fluctuations in epithelial, immune or stromal cell abundance between libraries could reflect differences in tissue cell content or technical variances related to ischemic time during operation or the dissociation process.

We observed that single-cell transcriptomes derived from all patients intermingled within the epithelial, immune, and stromal compartments (Fig. 1B). When distinguishing normal versus tumor samples, distributions of single-cell profiles largely overlapped, although several regions within each plot were preferentially inhabited by transcriptomes derived from either normal or tumor tissues (Fig. 1C). This indicates that our data are generally free from patient- or sample-specific batch effects confounding the cell type distributions, but that general differences occur between normal and tumor transcriptome distributions.

Cell type census in CRC versus normal colon

We next clustered the single-cell profiles of the epithelial, immune and stromal compartments, and used cell-type-specific signatures and marker genes to annotate cell types[14] (Fig. 2A; Supplementary

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.10.901579; this version posted January 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Fig. 2 for genes over- or underrepresented between tumor and normal samples). In the normal epithelium, we identified a zone populated by profiles of undifferentiated cells with high activity of stem cell markers such as *OLFM4* and transiently-amplifying proliferative markers. This region bordered on transcriptome clusters annotated as enterocyte progenitors, and, ultimately, mature absorptive enterocytes with high expression of markers such as *KRT20* and *FABP1*. *BEST4*- and *OTOP2*-expressing enterocytes formed a discrete cluster (Fig. 2A, identified by the lightest shade of green) [15]. Further separate epithelial clusters were populated by profiles annotated as immature and mature secretory goblet cells defined by expression of *MUC2*, *TFF1*, and *TFF3,* and *TRMP5*-expressing tuft cells. In the tumor samples, the zone of undifferentiated epithelial cells was expanded by five largely tumor-specific clusters (TC1-5, Fig. 2A, B; Supplementary Fig. 3 for data per patient). In contrast, clusters of differentiated absorptive and secretory cell transcriptomes were reduced in size. Profiles representing tuft cells and *BEST4*/*OTOP2*-positive enterocytes were vastly underrepresented in the tumor cell libraries.

We used immunofluorescence to verify the spatial distributions of epithelial cell types, using marker genes identified in the single-cell data (Fig. 2C, D). We detected the stem cell marker OLFM4 exclusively at the base of normal crypts. However, in tumor sections, OLFM4, as well as the proliferation marker MKI67, stained cells scattered throughout the epithelium, as validated by co-staining with the epithelial marker EPCAM. 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. Clusters of TFF3- and FABP1-positive cells that were largely negative for MKI67 suggest the presence of differentiated cells in CRC, in agreement with the single-cell sequencing data.

The five tumor-specific epithelial cell clusters TC1 to TC5 were represented in different proportions in all eight CRCs under investigation (Fig. 2E). TC1 and TC2 were assigned as highly stem cell-like using prior classifiers[14], while TC3-4 showed the strongest similarity to transiently-amplifying cell types and TC5 shared similarity with both, transient-amplifying and stem cells. Transcriptome-based cell cycle analysis revealed that TC1 is highly proliferative (Fig. 2F). Furthermore, TC clusters shared a couple of defining genes that have previously been linked to oncogenic processes in CRC (Supplementary Fig. 4; Supplementary Table 1). The CRC stem cell marker *CD44*[16] was expressed prominently in TC1 and TC4. *MMP7,* encoding a matrix metalloproteinase responsible for CRC invasion[11], was expressed highest in TC4, but also in TC1 and TC3. *VIM*[17] and *S100A4*[18], key markers of the epithelial-to-mesenchymal transition, were among the genes defining TC2. *PLA2G2A,* encoding a phospholipase that controls inflammation and homeostasis in the intestinal stem cell niche[19], and the intestinal stem cell marker *OLFM4* was overrepresented in TC5.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.10.901579; this version posted January 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

We annotated immune transcriptome clusters by lineage-specific marker genes (Fig. 2A, Supplementary Fig. 5). Two main clusters of myeloid cells were assigned as monocytes and macrophages by expression of *CD14* and *CD68*. *ITGAX* (encoding CD11c) was also expressed in this domain of the UMAP, indicating that these clusters, in addition, also encompass dendritic cells. A large cluster of T cell profiles, as identified by *CD3*, could be subgrouped into *CD8*-positive cytotoxic T cells and *CD4*-positive T helper cells. Subclusters of conventional and regulatory T helper cells were assigned by higher relative levels of *IL7R* (encoding CD127), *FOXP3* and *IL2RA* (encoding CD25). B cell and plasma cell clusters were defined by *CD19*, *MS4A1* (encoding CD20) and *SDC1* (encoding CD138), respectively. Among T, B, and plasma cells we could distinguish several subclusters that were represented in tissue samples across the patients. Six of the 26 immune cell clusters in our analysis were expanded in all eight CRC. These comprise the macrophage/monocyte clusters, regulatory T cells, two clusters of plasma cells (termed Plasma 5 and 8, see Supplementary Table 2) and one cluster of CD8-positive T cells expressing high levels of *IL17A* (CD8+ cluster 4, Supplementary Table 2). This interleukin has been implicated in CRC progression[20], and a similar type of T cell was recently found expanded in single-cell analyses of colitis patients[14].

Among stromal transcriptomes, we annotated an interconnected supercluster of fibroblasts. 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 matrix metalloproteinase-encoding genes such as *MMP1*, *MMP11*, *MMP3* and *MMP2*, suggesting roles of these cells in the degradation 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 ligand *WNT2B* and the Wnt amplifier *RSPO3*[21,22], upper crypt fibroblasts expressing genes such as *BMP2* and *BMP4* encoding differentiation-associated growth factors[23] (Fig. 2A, Supplementary Fig. 6, Supplementary Table 3) 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 *ACTA2* and *DES* and pericytes marked by *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.

Paracrine signaling in CRC ecosystems

As we discovered multiple clusters of epithelial tumor cells (TC1-5) and expanded clusters of stromal and immune cells in their microenvironment, we investigated possible paracrine interactions. For this, we mapped cognate ligand-receptor pairs in our single-cell data, taking into account expression levels

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.10.901579; this version posted January 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

of ligands, the prevalence of the ligand-expressing cell (that is, cluster size for that cell type), and fractions of receptor-expressing cells. We focused on ligand-encoding genes active in immune or stromal cells and genes encoding matching receptors in proliferative epithelial cells, that is, in stem/TA cells and the five tumor-specific clusters TC1-5 (Fig. 3A, Supplementary Fig. 7A, and Supplementary Table 4). Possible ligand-receptor connections appeared relatively sparse in the normal tissue. However, we found many more potential paracrine signaling connections in the tumor. This was mainly due to three features of the tumor ecosystem: Firstly, CRC contained novel epithelial cell types TC1-5 expressing multiple receptors, including high levels of the receptor tyrosine kinase MET[24] and others (Supplementary Fig. 7B). Secondly, several immune and stromal cell populations expressing cognate ligands for the receptors in the epithelium were expanded (Supplementary Figs. 3 and 7B; Supplementary Table 5). In particular, CAFs dominated signaling in the microenvironment of tumors P008 and P016, and, to a lesser extent, P007 and P017, and express genes encoding ligands of cancer-relevant signaling pathways. Among these are *GREM1*, *WNT2B*, *WNT5A*, *HGF*, multiple EGFR ligands including *AREG* and many others that have been linked to CRC initiation and progression[25–28]. Finally, some cell types in the cancer microenvironment expressed additional ligands compared to their normal tissue counterparts, for instance, crypt-base-like fibroblasts in the tumor express *FGF7* and *IGF* (Supplementary Fig. 7B).

We took a detailed look at Wnt signaling, as this pathway drives stem cell maintenance and tumor initiation in the gut. A signature of Wnt/β-catenin target genes was most active in the TC1-5 and stem cell compartment of tumor epithelium (Fig. 3B), while activity was lower among the differentiated CRC cells, similar to the normal colon epithelium. Indeed, it has been shown that Wnt/β-catenin is dynamic in CRC and can be activated by Wnt ligands[29,30], although the pathway is frequently activated by loss of APC. We detect the highest connectivity for Wnts and the R-Spondin family of Wnt amplifiers between CAFs, expressing *WNT2* and *WNT5A*, and stem cells and the tumor-specific cell clusters (Fig. 3A).

We next investigated EGFR-RAS-RAF-ERK signaling that plays a central role in CRC development and is a key target of therapy. ERK-regulated genes[31] were most active in TC4 (Fig. 3B). We identified CAFs, endothelial cells, but also immune cells as potential sources of EGFR family ligands in the CRC microenvironment (Fig 3A). Monocytes and macrophages that we found consistently enriched in tumor versus normal tissues, express the ligand-encoding genes *AREG*, *EREG*, and *HBEGF* (Fig 3C and Supplementary Table 3). These ligands could play roles in the activation of the EGFR-RAS-RAF-ERK cascade in CRC, particularly in tumors lacking mutations in RAS, RAF or other activating components of the pathway. Macrophages and CAFs also express *HGF*, encoding the MET receptor tyrosine kinase ligand driving cancer progression at the invasive tumor front[13]. Indeed, we found macrophages to be

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.10.901579; this version posted January 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

enriched in tumors specifically at the invasive front (Fig. 3C, D), suggesting that these could play roles in the induction of epithelial-mesenchymal transition via HGF-MET.

We additionally detected TGF-β target gene activity, along with expression of a gene signature of epithelial-to-mesenchymal transition in the tumor cell cluster TC4, and to a lesser degree in TC2 (Fig. 3B, E). Genes encoding TGF-β ligands were expressed in multiple stromal cell types, including CAFs, and cognate receptors were present in the TC clusters (Fig. 3E, Supplementary Fig. 7A, B). While the connectivity of the TC2 cluster was generally low, maybe also due to lower sequencing depth per cell, the cluster was defined by expression of *VIM* and *S100A4* (see above, Supplementary Fig. 4 and Supplementary Table 1), supporting the association of TC2 cells with the epithelial-to-mesenchymal transition. Interestingly, the size of cluster TC2 was highly correlated with the size of the CAF population in the cancer microenvironment in the eight tumor profiles (Fig. 3F), suggesting a potential role for CAFs in the support of the tumor cells clustering in TC2.

Cell type composition informs Consensus Molecular Subtypes

Consensus molecular subtypes (CMS) represent a transcriptome-based classification system for CRC with clinical utility[32]. We applied the CMS classifier to our single-cell data and found strong inter- and intratumor heterogeneity. Epithelial cells were exclusively assigned to CMS1-3 (Fig. 4A). The continuous cluster comprising intestinal stem cells, TA cells, and enterocytes of the normal tissue, as well as the TC1-5 tumor cell subtypes consists mainly of intermingled CMS1 and CMS2 cells. In our limited set of CRCs, we find that epithelial cells of the two cancers probably arising via serrated precursors (P007) or inflammation-induced progression (P008) are scoring predominantly CMS1, while epithelial cells of the other cancers score mainly as CMS2 (Supplementary Fig. 7). CMS3 appears to be confined to cells differentiating into the secretory lineage.

CMS subtypes were also unevenly distributed in the tumor microenvironment. Almost all immune cells were assigned CMS2, and only a minority scored as the “immune” subtype, CMS1 (Fig. 4B). Stromal cells were mostly CMS2, but a minority population of normal fibroblasts that we assigned previously as potential crypt base fibroblasts were CMS4 (Fig. 4C). Most interestingly, CAFs in the tumor tissue provided a strong CMS4 component that in our tumors was most prominent in P008 and P016 (Supplementary Fig. 7). As CMS4 was assigned to fibroblasts exclusively, we suggest that this cell type also drives the “mesenchymal” CMS4 assignment from bulk CRC tissue. It is of note that CMS4 cancers have a worse clinical prognosis which may, therefore, also be linked to the presence of stromal tumor-specific fibroblasts. In summary, we found that different cell types of the tumor ecosystem are preferentially assigned to different CMS, and thus, that bulk tissue CMS assignments report combinations of cell state and cell type prevalence in the tumor.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.10.901579; this version posted January 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Cell type composition of patient-derived CRC organoids and matched tumor samples

We established organoid lines of the tumor samples P009 and P013, using standard culture conditions with media containing EGF, FGF, and p38- and TGF-β inhibitors (Fig. 5A)[33,34]. P009 tumor tissue initially grew out unevenly, however, formed a uniformly spheroidal organoid culture within three passages. P013 tumor tissue grew out swiftly and uniformly, forming complexly folded organoids. We used panel sequencing to confirm the identity of the organoids with the matched tumor tissue on a mutational level (Supplementary Table 6).

We used cell suspensions of the organoid lines for single-cell RNA sequencing to generate profiles before the first passage (designated as p0), essentially sequencing the primary tissue after one to two weeks of expansion in culture. We also sequenced transcriptomes of the established lines P009 and P013 after two and three passages, respectively. Single-cell profiles of organoids were of a higher quality than those from epithelial cells of the tumors, as judged by the uniformly low fraction of mitochondrial reads, despite having used similar conditions for disaggregation at 37°C. To exclude confounding effects of technical differences in the different single-cell profiles, we anchored the organoid profiles in the space of the matched epithelial tumor transcriptomes[35].

In the resulting integrated data set, organoid cell transcriptomes of the two patients and the different passage numbers intermingled (Fig. 5B). Next, we re-clustered the transcriptomes and assigned cell identities per cluster by matching cell types with the previous annotation (Fig. 5C, D; for the previous annotation, see Fig. 2A, B). We found that profiles corresponding to TA cells, differentiated cells, and the tumor-specific TC1-TC5 cell types were present in the organoids. Stem cells, tuft cells and BEST4/OTOP2-positive enterocytes, which are all cell types that were present only in small numbers in the original primary tumor samples, were not called in any organoid single-cell library. Surprisingly, despite the different phenotypic appearance of the organoid cultures, cell type distributions were very comparable. In particular, TC cell types were present at similar fractions regardless of organoid line and passage number, unlike the dissimilar TC cell type ratios in the matched tumors. In all four organoid libraries, the highly proliferative TC1 cell type was overrepresented compared to the matched tumors.

Differentiation trajectories of CRC organoids

As all major epithelial cell types were present in the organoids, we used the transcriptomes to establish how the tumor cells are related to each other within differentiation trajectories. We used diffusion maps to take into account a pseudotemporal order of the transcriptomes that is related to cell differentiation. Profiles of both organoid lines ordered into a structure containing cells expressing stem cell markers such as *LGR5*, *PROM1* or *CD44* merging into an extended projection containing cells

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.10.901579; this version posted January 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

positive for the absorptive and secretory differentiation markers *FABP1* and *TFF3* (Supplementary Fig. 9). Markers specific for the TC cell clusters, that is, *MMP7*, *S100A4*, and *VIM* were found overlapping and adjacent to the stem cell zone defined by *LGR5, PROM1* and *CD44*.

The organoid line P009 showed an enlarged zone inhabited by transcriptomes positive for the TC cell markers and was, therefore, suitable to infer a more granular picture of cell plasticity by taking into account RNA velocity, that is, direction of cell differentiation defined by ratios of immature unspliced versus mature spliced mRNAs (Fig. 5E). We could define a single trajectory origin in the vicinity of cells expressing the normal intestinal stem cell marker *LGR5* and *CD44*, which is a CRC stem cell marker highly expressed in TC1. The main differentiation trajectory extended towards the *FABP1*- and *TFF3*-positive differentiated cells. A shorter trajectory stretched towards the area with the highest expression of the TC4 cell marker *MMP7*. We conclude that the TC cells can inhabit a zone of cell plasticity encompassing CRC stem cells and progenitor-like descendants that are, however, distinct from absorptive or secretory progenitors.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.10.901579; this version posted January 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Discussion

Here, we use droplet-based single-cell sequencing of transcriptomes to characterize patient-derived matched CRC and normal colon tissue, as well as CRC organoids. We find that CRC, regardless of specific genetic mutations or clinical parameters, contains a large proportion of tumor-specific undifferentiated epithelial cells in addition to cells that resemble differentiated cell types of the normal colon epithelium. We assign the tumor-specific epithelial cells to five clusters, TC1-5, that form a continuum of cell plasticity. The TC cell types are distinguished by proliferative activity, oncogenic signaling, and gene expression patterns related to stemness, tissue invasion, and epithelial-to-mesenchymal transition. We conclude that these cancer traits are unevenly distributed between tumor cell subpopulations. Furthermore, we identify stromal and immune cell types, including CAFs, macrophages, monocytes, and subsets of CD8+ T cells as cell types enriched in the tumor microenvironment that are sources of multiple growth factors initiating or amplifying Wnt-, TGF-β-, EGFR- and HGF-MET-signaling. This data suggest that paracrine signaling is a defining factor shaping the CRC ecosystem.

Our single-cell analysis illuminates a couple of clinically relevant features of CRC. Classification of CRC by bulk cancer transcriptome analysis can be achieved by the consensus molecular subtype system[32]. CMS4, also termed as the “mesenchymal” subtype, is notable for its worse relapse-free and overall survival. We show here on a single-cell level that CMS4 transcriptomes stem specifically from fibroblasts, in particular, CAFs, while epithelial tumor cells can only assume CMS1 - CMS3. We conclude, therefore, that CMS4 is assigned to CRCs with a high content of CAFs that could possibly also contain a large fraction of the correlated TC2 tumor cell subtype. As we find CAFs to produce multiple pro-oncogenic growth factors including HGF and TGF-β ligands, it appears as a plausible strategy to target paracrine signaling as a future therapeutic option for the CMS4 CRC subtype. Furthermore, we could assign epithelial differentiation states to CMS subtypes. Our findings could be used to incorporate further informative and cell-type-specific genes into the CMS classifier, in order to assign CRCs with greater specificity, including those CRCs that currently cannot be assigned to a CMS subtype.

Anti-EGFR antibodies serve as first-line targeted therapy for patients with metastatic disease and no mutations in the EGFR-RAS-RAF-ERK signaling axis. Treatment success in this cohort has been linked to the production of the EGFR ligands AREG and EREG, and to immune infiltration in separate publications[28,36]. Connecting these phenomena, we have identified here *AREG* and *EREG* to be expressed by immune cells in the CRC ecosystems that we investigated here, in addition to epithelial cells. *AREG* and *EREG* were most strongly active in monocytes, and *AREG* was additionally expressed in other immune cell types. We hypothesize that *AREG*-/*EREG*-expressing immune cells contribute to

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.10.901579; this version posted January 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the paracrine signaling loop activating ERK in KRAS-, NRAS- and BRAF-wildtype CRC cells, and possibly influence anti-EGFR antibody therapy outcome.

Our results imply that CRC cells display considerable cell plasticity and have multilineage differentiation capacity, in agreement with pioneering single-cell sequencing studies[37,38]. Stem cells have traditionally been seen as unique cells driving tissue homeostasis and regeneration of normal tissue, but also therapy-resistance in cancer. However, recent studies have shown that combinations of oncogenic mutations and paracrine signals can steer the cell-intrinsic signaling network so that differentiation trajectories are reversed and more differentiated cells can regain stem cell characteristics[39–43]. In CRC, stem cells can be maintained and induced by factors such as HGF[13] and IL17A[20]. We found Macrophages and CAFs to express HGF and EGFR ligands, extending previous observations[44,45]. We also observed an expansion of IL17A-expressing CD8+ T cells. As these cells also showed transcription of *KLRB1*, these cells might be mucosal-associated invariant T cells (also known as MAITs)[46,47]. To verify this, an analysis of the invariant T cell receptor Valpha7.2 could be performed in future studies.

Some tumor samples in our analysis contained a small proportion of normal epithelial tissue (see Supplementary Fig. 1). To ascertain the origin of transcriptomes assigned to differentiated epithelial cell clusters, we calculated probabilities for single-cell transcriptomes to be derived from the tumor, taking into account RNA reads covering somatic mutations. While this approach successfully assigned a small fraction of transcriptomes as deriving from normal or tumor cells, respectively, a large majority of single-cell profiles remained unassigned (Supplementary Fig. 10). We conclude that single-cell transcriptomes acquired by our droplet-based sequencing platform contain insufficient information for mutation-based tumor cell assignment. However, several lines of evidence support that CRC contains many cells similar to normal differentiated cell types: firstly, we find enterocyte- and goblet cell-like transcriptomes in tumor samples, but tuft cells and BEST4/OTOP2-positive enterocytes are selectively depleted, arguing against normal tissue contamination. Secondly, we could stain a substantial proportion of tumor cells using antibodies against the goblet cell and enterocyte differentiation markers TFF3 and FABP1, respectively. Thirdly, organoids cultured in a medium supporting the specific outgrowth of tumor cells maintained differentiated cell populations, with the exception of tuft cells and BEST4/OTOP2-positive enterocytes.

In organoid cultures, the complex *in vivo* microenvironment is substituted by a uniform extracellular matrix and only a few growth factors. It has not been examined in detail how organoid culture conditions affect tumor cell heterogeneity compared to the tissue of origin. We show here that CRC organoids maintain all main cell types of CRC, including the tumor-specific cell types TC1-TC5, in the absence of stromal or immune cells. However, organoids were enriched for profiles of the strongly proliferative TC1 cells in both patient-derived cultures and multiple passages, probably reflecting the

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.10.901579; this version posted January 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

growth conditions with high concentrations of EGF, but lacking a complex cellular microenvironment (see Fig. 5D). It is of note that improved experimental procedures are now available incorporating fibroblasts into organoid cultures to better mimic paracrine interactions present *in vivo*[48]. Single-cell analysis of such models could inform the dissection of cellular interdependences in CRC, and would also provide a more realistic scenario for preclinical drug tests.

Our workflow for dissociating clinical samples resulted in the acquisition of many types of major epithelial, immune and stromal cell transcriptomes. In addition, we capture subtle transcriptome differences within T cells, plasma cells, fibroblasts, and the tumor-specific TC cells that were assigned to multiple similar clusters. Analysis of single-cell transcriptomes is subject to multiple confounding factors, including numbers of genes detected, fractions of spliced versus unspliced mRNAs, and the fraction of mitochondrial reads. Indeed, particularly the epithelial cell transcriptomes that we captured for the present study varied on these quality parameters within, but also between the clusters. Reassuringly, the distinction of phenotypic features between clusters was largely independent of the confounders. However, we note that the discrimination of five tumor-specific epithelial clusters is purely heuristic and, presently, we consider tumor-specific epithelial cells to have continuous phenotypic plasticity rather than inhabiting fixed states that can be clustered with confidence. Improvements to methods for tissue dissociation have recently been published[49] and should be incorporated into future clinical workflows to diminish tissue processing artifacts. However, with clinical samples, cell-type-specific degradation of transcriptome quality during the ischemic time window between the restriction of blood flow and completion of the operation probably cannot be avoided completely.

The extension of single-cell analyses of CRC to multi-omics, taking also in account genetic and epigenetic heterogeneity[50,51], promises to identify cell plasticity and genetic diversity in cancer at a cellular resolution. We believe that such approaches will aid the future identification and eradication of CRC cell populations responsible for therapy resistance.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.10.901579; this version posted January 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Methods

Acquisition and processing of clinical specimens

Fresh normal colon and colorectal cancer tissues were acquired during the intraoperational pathologist´s examination. Tissues (approx. 0.1-0.4g) were minced using scalpels and stored short-term on ice in Tissue Storage Solution (Miltenyi # 130-100-008) for transport. Next, tissues were processed using the Miltenyi human Tumor Dissociation Kit (Miltenyi, #130-095-929) and a Miltenyi gentleMACS Octo Tissue Dissociator with heaters (Miltenyi, #130-096-427), using program 37C_h_TDK_1 for 30-45min. Cell suspensions were filtered using 100µm filters, and all subsequent steps were performed at 4°C or on ice. Cells were pelleted by centrifugation in BSA-coated low-binding tubes, and cells were treated with 1ml ACK erythrocyte lysis buffer for 60 seconds and washed with DMEM. Cells were pelleted, resuspended in PBS, cell suspensions were filtered using 20µm filters, debris was removed using the Debris Removal Solution (Miltenyi #130-109-398), and cells were counted using a Neubauer chamber. At least 10^4 cells of all suspensions were analyzed for cell viability >75% using LIVE/DEAD Fixable Dead Cell Stain Kit (488nm; Thermo Fisher) and a BD Accuri cytometer.

Hematoxylin-and-eosin staining and immunostaining

3-5 µm tissue sections of fresh frozen or formalin-fixed and paraffin-embedded (FFPE) tissue were used for immunofluorescence, immunohistochemistry and hematoxylin and eosin staining. Immunohistochemical and immunofluorescence stainings of FFPE tissue sections were performed on the BenchMark XT immunostainer (Ventana Medical Systems). For antigen retrieval, tissue sections were incubated in CC1 mild buffer (Ventana Medical Systems) for 30 min at 100°C. Sections were incubated with primary antibodies for 60 min and with secondary antibodies for 30 minutes at room temperature diluted in Dako Real Antibody Diluent (Dako, S2022). 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, ThermoScientific, MS-144-P1), rabbit anti-Ki67 (1:400, Abcam, ab16667), mouse anti-Ki67 (1:50, Dako, M7240), mouse anti-CD68 (1:100, Dako, M0876). Images were taken using an Axio Vert.A1 fluorescence microscope (Zeiss) equipped with an Axiocam 506 color camera (Zeiss).

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/library to a mean library saturation of 50%. This resulted in 35.000 to 120.000 reads per cell.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.10.901579; this version posted January 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

DNA Sequencing

For panel sequencing for frequent oncogenic driver mutations, tumor-enriched areas (> 40 % tumor cells) were macrodissected from FFPE tissue sections, DNA was extracted using the Maxwell RSC DNA FFPE Kit (Promega) and sequenced using the Ion AmpliSeq Cancer Hotspot Panel v2, and an IonTorrent sequencer (ThermoFisher) according to the manufacturer's instructions. For variant calling the Sequence Pilot Software (Version 4.4.0, JSI Medical Systems) was used. For the sequencing of exomes (patients P007, P008, P009), DNA was isolated from fresh frozen tumor tissue after the pathologist´s examination using the DNeasy Blood and Tissue Kit (Qiagen). Exomes were sequenced using the AllExon Human SureSelect v7 Kit (Agilent).

Organoid Culture

Cell filtrates from patient-derived tumor tissues that were retained in the 20µm filters after dissociation were washed in Advanced DMEM/F12 medium (Gibco), embedded in Matrigel, and cultured in 24-well plates, according to published procedures. Rho-kinase inhibitor Y27632 (10µM, Sigma) was used for the first passage to avoid anoikis. Cells originally embedded in Matrigel (Corning) were termed passage 0 (p0), and outgrowing organoids were passaged by removal from Matrigel, washing in PBS, and partial digestion using TrypLE cell dissociation solution (Gibco) at 37°C, washing in medium and re-embedding in Matrigel. For single-cell sequencing, organoids were dissociated completely using TrypLE and DNAseI, and filtering via a 20µm filter.

Single-cell RNA-seq data analysis

For each sample, UMIs were quantified using cellranger 3.0.2 with reference transcriptome GRCh38. Spliced, unspliced and ambiguous UMIs were quantified with velocyto[52] (mode: run10x, default parameters). Quality control filters were set to only 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[53], while regressing out fractions of mitochondrial reads and differences in S-Phase and G2M-Phase scored with Seurat v3[35]. Next, main cell types (epithelium, stromal, and immune cells) were identified by scoring cell type markers across Louvain clusters for each sample (resolution = 1). In samples where more than 2000 cells were assigned to a certain cell type, a random sample of 2000 cells was used of this cell type for the given sample. Normalized subsets were merged for each main cell type of normal and tumor samples without further batch correction. Louvain cluster-specific marker genes of merged normal and tumor samples were used to identify sub cell types among epithelial, stromal and immune subsets. Gene expression sets were taken from the hallmark signature

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.10.901579; this version posted January 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

collection of the Broad institute[54], unless otherwise referenced in the main text, and were scored with Seurat v3. Epithelial subsets of tumor and matched organoid samples were integrated using Seurat v3 with tumor samples as reference[35]. Consensus molecular subtypes were scored using random forrest approach included in R package CMSclassifier.

Diffusion map analysis and RNA velocity were performed using scanpy[55] and scvelo[56]. 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 tutorial, 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.

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.

Tumor cell calling

For tumor-specific single-nucleotide variant (SNV) calling on single-cell data, we employed exome sequence data of patients P007, P008 and P009. We used Mutect2[57] to detect SNVs, retaining only events classified as somatic. We additionally filtered these results by removing variants on non-canonical chromosomes or within repeat regions (UCSC genome browser RepeatMasker track). We then used cellSNP [58] to quantify the total number dij of (UMI-collapsed) reads covering variant i in cell j (reference and variant allele), and the number aij of reads supporting the alternative allele, using all variants detected with at least one read in any cell, and all cells containing at least 3 variant-covering reads. Adapting the EM-type approach proposed in McCarthy et al.[59] to our situation, we evaluate the likelihood pT,j that cell j is a tumor cell using a binomial model:

$$p_{T,j} \propto \prod_i \binom{d_{ij}}{a_{ij}} \theta_T^{a_{ij}} (1-\theta_T)^{d_{ij}-a_{ij}}$$

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.10.901579; this version posted January 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Here, θ_T is the "success probability" for the somatic variants, measuring how likely it is to get a read supporting the variant allele. Similarly, we compute pN,j as the likelihood that cell j is normal, with a fixed parameter θ_N=0.01 allowing for sequencing errors and uncertainties in the variant calls. We calculate pT,j and pN,j in the E-step and estimate the parameter θ_T in the M-step as weighted sum over the counts dij and aij:

$$\theta_T = \frac{\sum_j \frac{p_{T,j}}{p_{T,j}+p_{N,j}} \sum_i a_{ij}}{\sum_j \frac{p_{T,j}}{p_{T,j}+p_{N,j}} \sum_i d_{ij}}$$

E- and M-steps are iterated until convergence of the likelihood

$$\ln \mathcal{L} = \sum_j \ln(p_{T,j} + p_{N,j})$$

Finally, the criterion pT,j > pN,j is used to define likely tumor cells.

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.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.10.901579; this version posted January 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Funding

The work was in part funded by the Berlin Institute of Health (to NB, CS, DH and MM), and the German Cancer Consortium DKTK (to NB and MM).

Author contributions

PB, FU, MM, AS, ML conducted and analysed experiments; FU, NB, BO, EB performed bioinformatic analyses; MM, NB, CS, BS, DH conceived, designed, interpreted experiments and/or supervised parts of the study; PB, MM, DH, CK 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.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.10.901579; this version posted January 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

References

1. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646–674 (2011).

2. Sever, R. & Brugge, J. S. Signal transduction in cancer. *Cold Spring Harb. Perspect. Med.* (2015) doi:10.1101/cshperspect.a006098.

3. Binnewies, M. *et al.* Understanding the tumor immune microenvironment (TIME) for effective therapy. *Nat. Med.* (2018) doi:10.1038/s41591-018-0014-x.

4. Becht, E. *et al.* Immune and stromal classification of Colorectal cancer is associated with molecular subtypes and relevant for precision immunotherapy. *Clin. Cancer Res.* (2016) doi:10.1158/1078-0432.CCR-15-2879.

5. Kather, J. N. & Halama, N. Harnessing the innate immune system and local immunological microenvironment to treat colorectal cancer. *British Journal of Cancer* (2019) doi:10.1038/s41416-019-0441-6.

6. Fearon, E. R. Molecular genetics of colorectal cancer. *Annu. Rev. Pathol.* **6**, 479–507 (2011).

7. The Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487**, 330–337 (2012).

8. Ricci-Vitiani, L. *et al.* Identification and expansion of human colon-cancer-initiating cells. *Nature* **445**, 111–115 (2007).

9. O'Brien, C. A. *et al.* A human colon cancer cell capable of initiating tumour growth in immunodeficient mice. *Nature* **445**, 106–110 (2007).

10. Shimokawa, M. *et al.* Visualization and targeting of LGR5+ human colon cancer stem cells. *Nature* **545**, 187–192 (2017).

11. Brabletz, T., Jung, A., Dag, S., Hlubek, F. & Kirchner, T. β-catenin regulates the expression of the matrix metalloproteinase-7 in human colorectal cancer. *Am. J. Pathol.* (1999) doi:10.1016/S0002-9440(10)65204-2.

12. Spaderna, S. *et al.* A Transient, EMT-Linked Loss of Basement Membranes Indicates Metastasis and Poor Survival in Colorectal Cancer. *Gastroenterology* (2006) doi:10.1053/j.gastro.2006.06.016.

13. Vermeulen, L. *et al.* Wnt activity defines colon cancer stem cells and is regulated by the microenvironment. *Nat. Cell Biol.* **12**, 468–476 (2010).

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.10.901579; this version posted January 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

14. Smillie, C. S. *et al.* Intra- and Inter-cellular Rewiring of the Human Colon during Ulcerative Colitis. *Cell* (2019) doi:10.1016/j.cell.2019.06.029.

15. Parikh, K. *et al.* Colonic epithelial cell diversity in health and inflammatory bowel disease. *Nature* (2019) doi:10.1038/s41586-019-0992-y.

16. Du, L. *et al.* CD44 is of functional importance for colorectal cancer stem cells. *Clin. Cancer Res.* (2008) doi:10.1158/1078-0432.CCR-08-1034.

17. Toiyama, Y. *et al.* Increased expression of slug and vimentin as novel predictive biomarkers for lymph node metastasis and poor prognosis in colorectal cancer. *Carcinogenesis* (2013) doi:10.1093/carcin/bgt282.

18. Gongoll, S. *et al.* Prognostic significance of calcium-binding protein S100A4 in colorectal cancer. *Gastroenterology* (2002) doi:10.1053/gast.2002.36606.

19. Schewe, M. *et al.* Secreted Phospholipases A2 Are Intestinal Stem Cell Niche Factors with Distinct Roles in Homeostasis, Inflammation, and Cancer. *Cell Stem Cell* **19**, 38–51 (2016).

20. Lotti, F. *et al.* Chemotherapy activates cancer-associated fibroblasts to maintain colorectal cancer-initiating cells by IL-17A. *J. Exp. Med.* (2013) doi:10.1084/jem.20131195.

21. De Lau, W. *et al.* Lgr5 homologues associate with Wnt receptors and mediate R-spondin signalling. *Nature* (2011) doi:10.1038/nature10337.

22. Yan, K. S. *et al.* Non-equivalence of Wnt and R-spondin ligands during Lgr5 + intestinal stem-cell self-renewal. *Nature* (2017) doi:10.1038/nature22313.

23. Haramis, A.-P. G. *et al.* De novo crypt formation and juvenile polyposis on BMP inhibition in mouse intestine. *Sci. (New York, NY)* **303**, 1684–1686 (2004).

24. Birchmeier, C., Birchmeier, W., Gherardi, E. & Vande Woude, G. F. Met, metastasis, motility and more. *Nat. Rev. Mol. Cell Biol.* **4**, 915–925 (2003).

25. Rohlin, A. *et al.* GREM1 and POLE variants in hereditary colorectal cancer syndromes. *Genes Chromosom. Cancer* (2016) doi:10.1002/gcc.22314.

26. Miyoshi, H., Ajima, R., Luo, C. T., Yamaguchi, T. P. & Stappenbeck, T. S. Wnt5a potentiates TGF-β signaling to promote colonic crypt regeneration after tissue injury. *Sci. (New York, NY)* **338**, 108–113 (2012).

27. Bakker, E. R. M. *et al.* Wnt5a promotes human colon cancer cell migration and invasion but does not augment intestinal tumorigenesis in apc1638N mice. *Carcinogenesis* (2013)

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.10.901579; this version posted January 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

doi:10.1093/carcin/bgt215.

28. Jacobs, B. *et al.* Amphiregulin and epiregulin mRNA expression in primary tumors predicts outcome in metastatic colorectal cancer treated with cetuximab. *J. Clin. Oncol.* (2009) doi:10.1200/JCO.2008.21.3744.

29. Brabletz, T. *et al.* Nuclear overexpression of the oncoprotein β-Catenin in colorectal cancer is localized predominantly at the invasion front. *Pathol. Res. Pract.* (1998) doi:10.1016/S0344-0338(98)80129-5.

30. Voloshanenko, O. *et al.* Wnt secretion is required to maintain high levels of Wnt activity in colon cancer cells. *Nat. Commun.* (2013) doi:10.1038/ncomms3610.

31. Uhlitz, F. *et al.* An immediate–late gene expression module decodes ERK signal duration. *Mol. Syst. Biol.* **13**, 928 (2017).

32. Guinney, J. *et al.* The consensus molecular subtypes of colorectal cancer. *Nat. Med.* **21**, 1350–1356 (2015).

33. Sato, T. *et al.* Long-term expansion of epithelial organoids from human colon, adenoma, adenocarcinoma, and Barrett's epithelium. *Gastroenterology* **141**, 1762–1772 (2011).

34. Schütte, M. *et al.* Molecular dissection of colorectal cancer in pre-clinical models identifies biomarkers predicting sensitivity to EGFR inhibitors. *Nat. Commun.* **8**, 14262 (2017).

35. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* (2019) doi:10.1016/j.cell.2019.05.031.

36. Pozzi, C. *et al.* The EGFR-specific antibody cetuximab combined with chemotherapy triggers immunogenic cell death. *Nat. Med.* (2016) doi:10.1038/nm.4078.

37. Dalerba, P. *et al.* Single-cell dissection of transcriptional heterogeneity in human colon tumors. *Nat. Biotechnol.* (2011) doi:10.1038/nbt.2038.

38. Li, H. *et al.* Reference component analysis of single-cell transcriptomes elucidates cellular heterogeneity in human colorectal tumors. *Nat. Genet.* **49**, 708–718 (2017).

39. van Es, J. H. *et al.* Dll1(+) secretory progenitor cells revert to stem cells upon crypt damage. *Nat. Cell Biol.* **14**, 1099–1104 (2012).

40. Schwitalla, S. *et al.* Intestinal Tumorigenesis Initiated by Dedifferentiation and Acquisition of Stem-Cell-like Properties. *Cell* (2012).

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.10.901579; this version posted January 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

41. Buczacki, S. J. A. *et al.* Intestinal label-retaining cells are secretory precursors expressing Lgr5. *Nature* **495**, 65–69 (2013).

42. Jadhav, U. *et al.* Dynamic Reorganization of Chromatin Accessibility Signatures during Dedifferentiation of Secretory Precursors into Lgr5+ Intestinal Stem Cells. *Cell Stem Cell* (2017) doi:10.1016/j.stem.2017.05.001.

43. Tomic, G. *et al.* Phospho-regulation of ATOH1 Is Required for Plasticity of Secretory Progenitors and Tissue Regeneration. *Cell Stem Cell* (2018) doi:10.1016/j.stem.2018.07.002.

44. Nabeshima, A. *et al.* Tumour-associated macrophages correlate with poor prognosis in myxoid liposarcoma and promote cell motility and invasion via the HB-EGF-EGFR-PI3K/Akt pathways. *Br. J. Cancer* (2015) doi:10.1038/bjc.2014.637.

45. Vlaicu, P. *et al.* Monocytes/macrophages support mammary tumor invasivity by co-secreting lineage-specific EGFR ligands and a STAT3 activator. *BMC Cancer* (2013) doi:10.1186/1471-2407-13-197.

46. Walker, L. J. *et al.* Human MAIT and CD8αα cells develop from a pool of type-17 precommitted CD8 + T cells. *Blood* (2012) doi:10.1182/blood-2011-05-353789.

47. Dusseaux, M. *et al.* Human MAIT cells are xenobiotic-resistant, tissue-targeted, CD161 hi IL-17-secreting T cells. *Blood* (2011) doi:10.1182/blood-2010-08-303339.

48. Karpus, O. N. *et al.* Colonic CD90+ Crypt Fibroblasts Secrete Semaphorins to Support Epithelial Growth. *Cell Rep.* (2019) doi:10.1016/j.celrep.2019.02.101.

49. Adam, M., Potter, A. S. & Potter, S. S. Psychrophilic proteases dramatically reduce single-cell RNA-seq artifacts: A molecular atlas of kidney development. *Dev.* (2017) doi:10.1242/dev.151142.

50. Bian, S. *et al.* Single-cell multiomics sequencing and analyses of human colorectal cancer. *Science (80-.).* (2018) doi:10.1126/science.aao3791.

51. Roerink, S. F. *et al.* Intra-tumour diversification in colorectal cancer at the single-cell level. *Nature* (2018) doi:10.1038/s41586-018-0024-3.

52. La Manno, G. *et al.* RNA velocity of single cells. *Nature* (2018) doi:10.1038/s41586-018-0414-6.

53. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *bioRxiv* (2019) doi:10.1101/576827.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.10.901579; this version posted January 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

54. Liberzon, A. *et al.* The Molecular Signatures Database Hallmark Gene Set Collection. *Cell Syst.* (2015) doi:10.1016/j.cels.2015.12.004.

55. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: Large-scale single-cell gene expression data analysis. *Genome Biol.* (2018) doi:10.1186/s13059-017-1382-0.

56. Bergen, V., Lange, M., Peidli, S., Wolf, F. A. & Theis, F. J. Generalizing RNA velocity to transient cell states through dynamical modeling. *bioRxiv* (2019) doi:10.1101/820936.

57. Cibulskis, K. *et al.* Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.* (2013) doi:10.1038/nbt.2514.

58. Huang, Y., McCarthy, D. J. & Stegle, O. Vireo: Bayesian demultiplexing of pooled single-cell RNA-seq data without genotype reference. *bioRxiv* (2019) doi:10.1101/598748.

59. McCarthy, D. J. *et al.* Cardelino: Integrating whole exomes and single-cell transcriptomes to reveal phenotypic impact of somatic variants (under review at Nature Methods). *bioRxiv* (2018) doi:10.1101/413047.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.10.901579; this version posted January 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

A

Name	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/2O
P012	n/a	TisN0	G2	MSS	T	n/a	m	N/T
P013	APC^{LOH}	T4aN1bM1	G2	MSS	A	C	m	N/T/2O
P014	APC^{LOH}, $BRAF^{V600E}$, $HRAS^{A59T}$, $PIK3CA^{E545K}$	T3N2b	G3	MSS	A	n/a	f	N/T/O
P016	$KRAS^{G13C}$, $PIK3CA^{E545K}$, $FBXW7^{R278STOP}$	T3N0	G2	n/a	R	C	m	N/T
P017	$KRAS^{G13D}$ (5% allele frequency)	T3N2b	G3	MSS	A	n/a	f	N/T

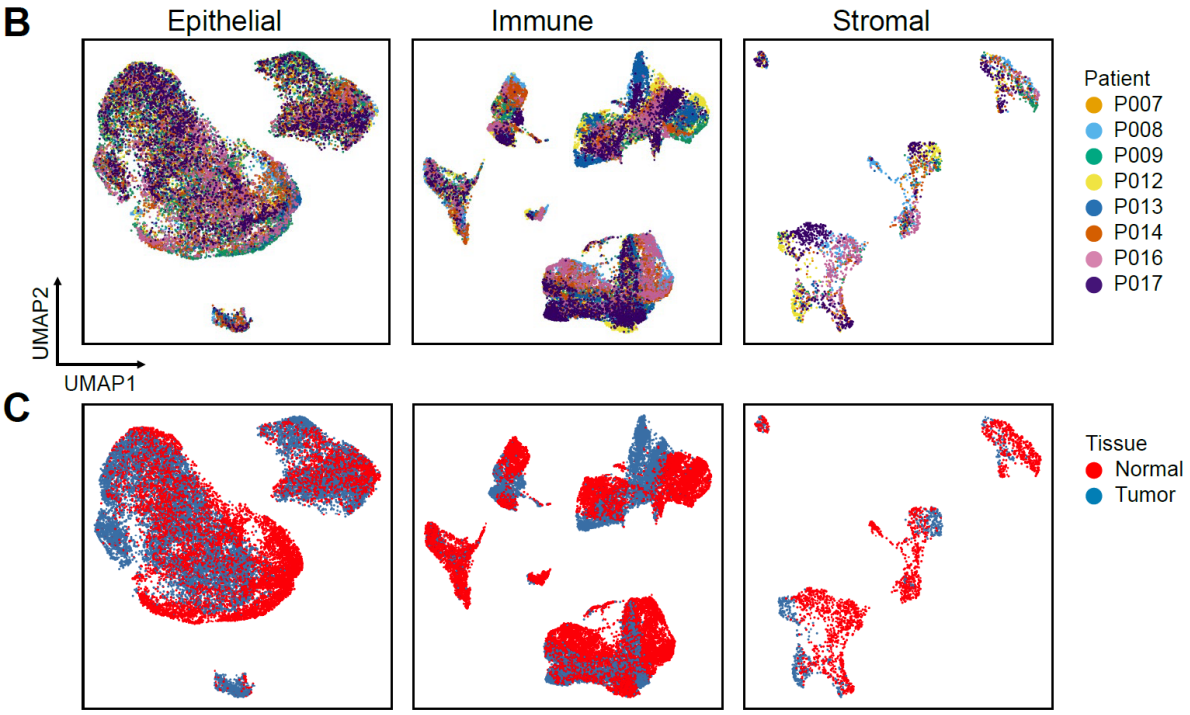

Figure 1: Generation and initial assignment of CRC single-cell RNA sequencing data. A Clinical data for the eight patients under investigation in this study. For mutational data, see also Supplementary table 6. Loc (Localisation): C: cecum; S: sigmoid colon; T: transverse colon; A: ascending colon; R: 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** UMAPs of epithelial, immune and stromal cells, color-coded for patients. **C** UMAPs of epithelial, immune and stromal cells, color-coded by tissue of origin.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.10.901579; this version posted January 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

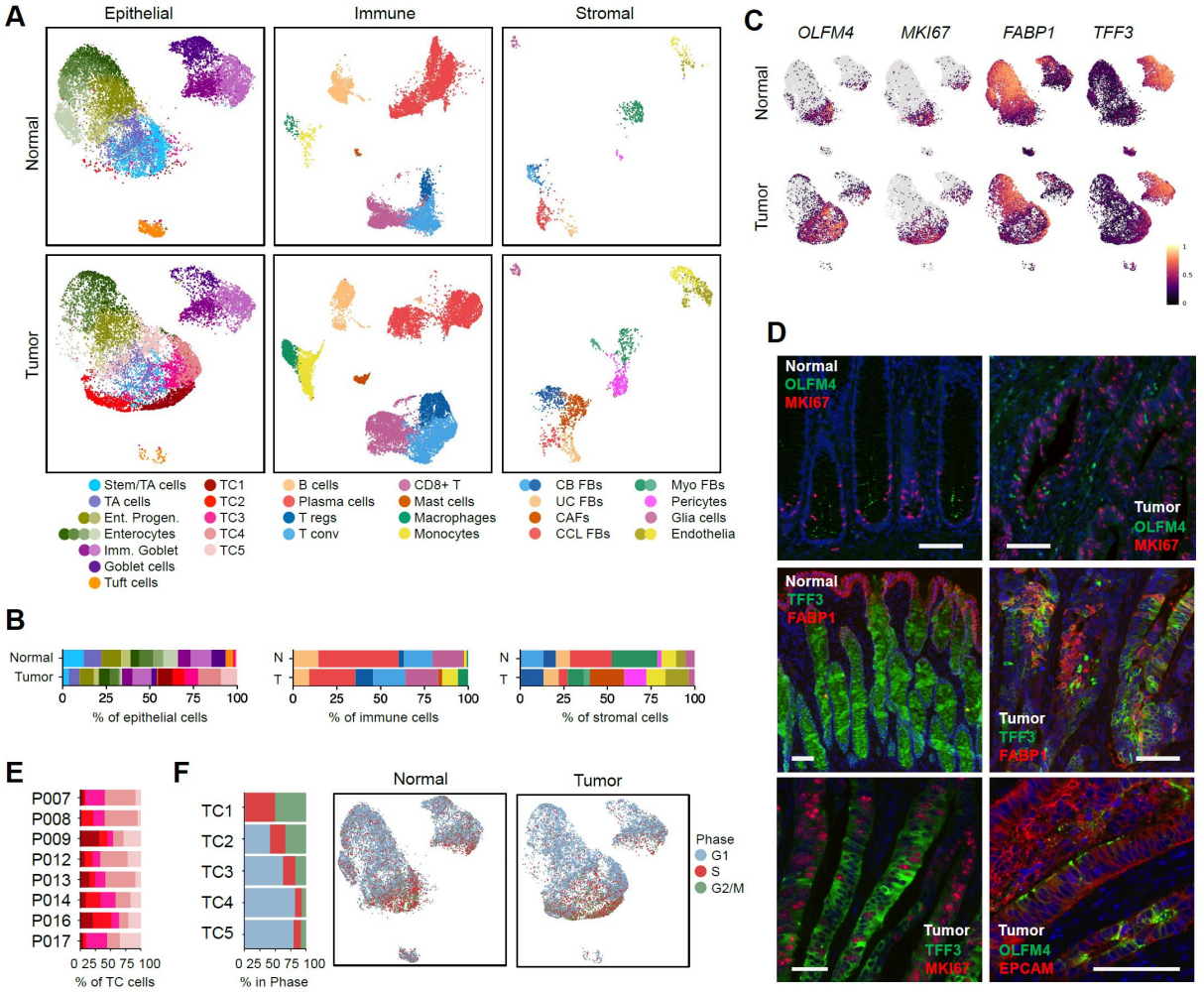

Figure 2: Cell type census in normal colon and CRC. A UMAPs of epithelial, immune and stromal cells, separated by tissue of origin. Color code for cell type assignment. **B** Relative fractions of epithelial, immune and stromal cell types across all patient-derived libraries. For fractions per patient, see Supplementary Figure 3. **C** Single cell gene activities of stem cell marker *OLFM4*, proliferative marker *MKI67*, differentiated absorptive cell marker *FABP1*, and secretory cell marker *TFF3*. **D** 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. **E** Relative fraction of TC1-5 in the tumor tissues of the patients. **F** Transcriptome-inferred cell-cycle distribution in the tumor cell fractions TC1-TC5 (bar graph), and in the UMAPs of normal and tumor epithelium.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.10.901579; this version posted January 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

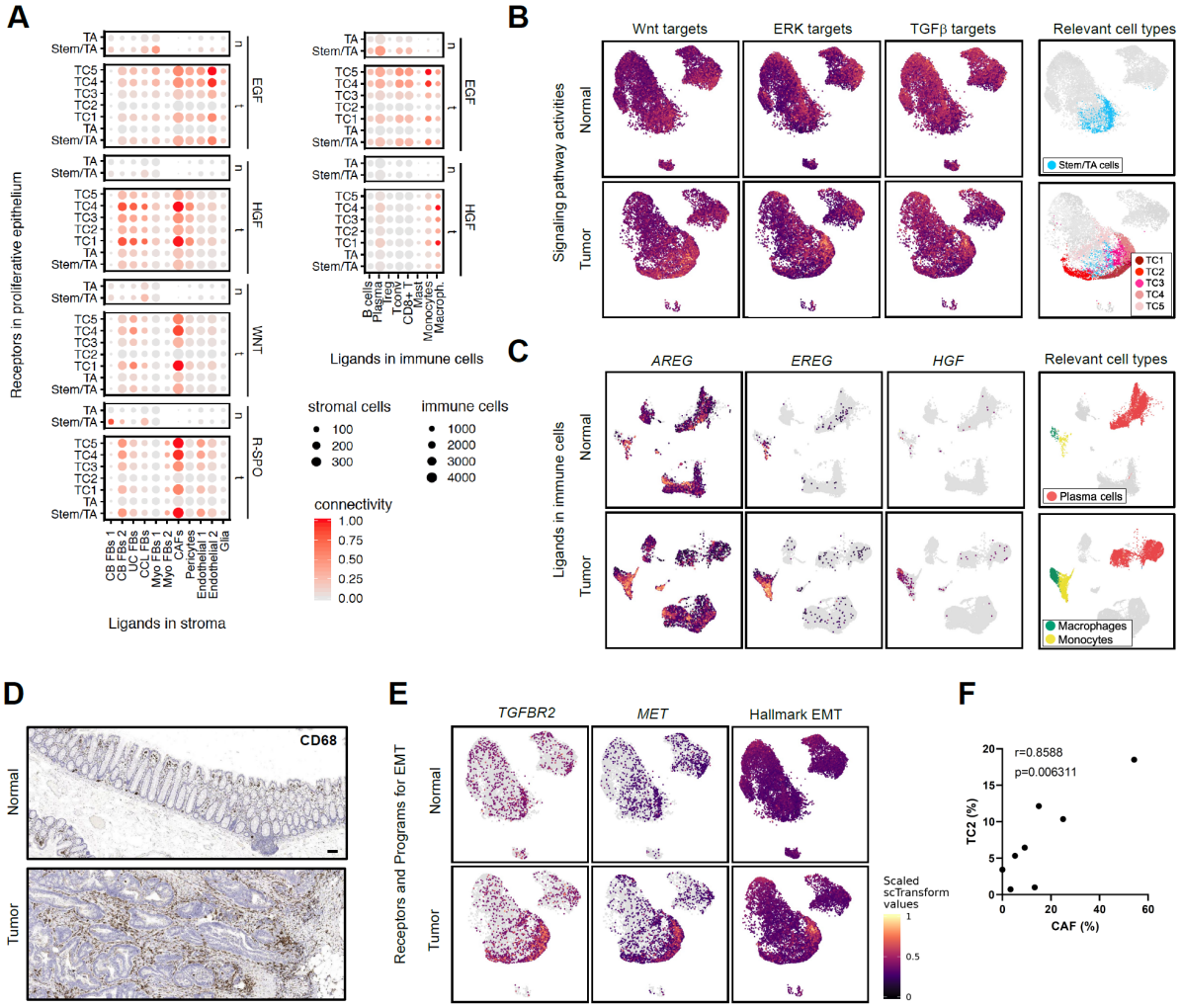

Figure 3: A connectivity map of potential paracrine interactions in CRC. A Connectivity analysis for ligand expression in stromal and immune cells, and receptor expression in proliferative epithelial cell clusters. Connectivity takes into account expression levels of ligands, the prevalence of the ligand-expressing cell, and fractions of receptor-expressing cells. See methods for details. Circle sizes: Cell numbers for ligand-expressing cells, as in figure legend. Red: High connectivity; grey: low connectivity. **B** Activities of Wnt, ERK and TGFβ target genes in the UMAPs of normal and tumor epithelium. **C** Expression of key ligands for CRC progression and therapy response in immune cells. **D** Immunohistochemistry for macrophages in normal and tumor tissue of P009, using an antibody against CD68. In the tumor, macrophages are more prevalent, in line with the single-cell data. Macrophages are enriched at the invasive front (tumor, to the right). Scale bar indicates 100µm. **E** Activity of receptors implicated in the epithelial-to-mesenchymal transition, and activity of genes comprising the EMT hallmark signature. **F** Correlation between fractions of CAFs in the stromal cell compartment and TC2 tumor cells in the epithelial tumor compartment.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.10.901579; this version posted January 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

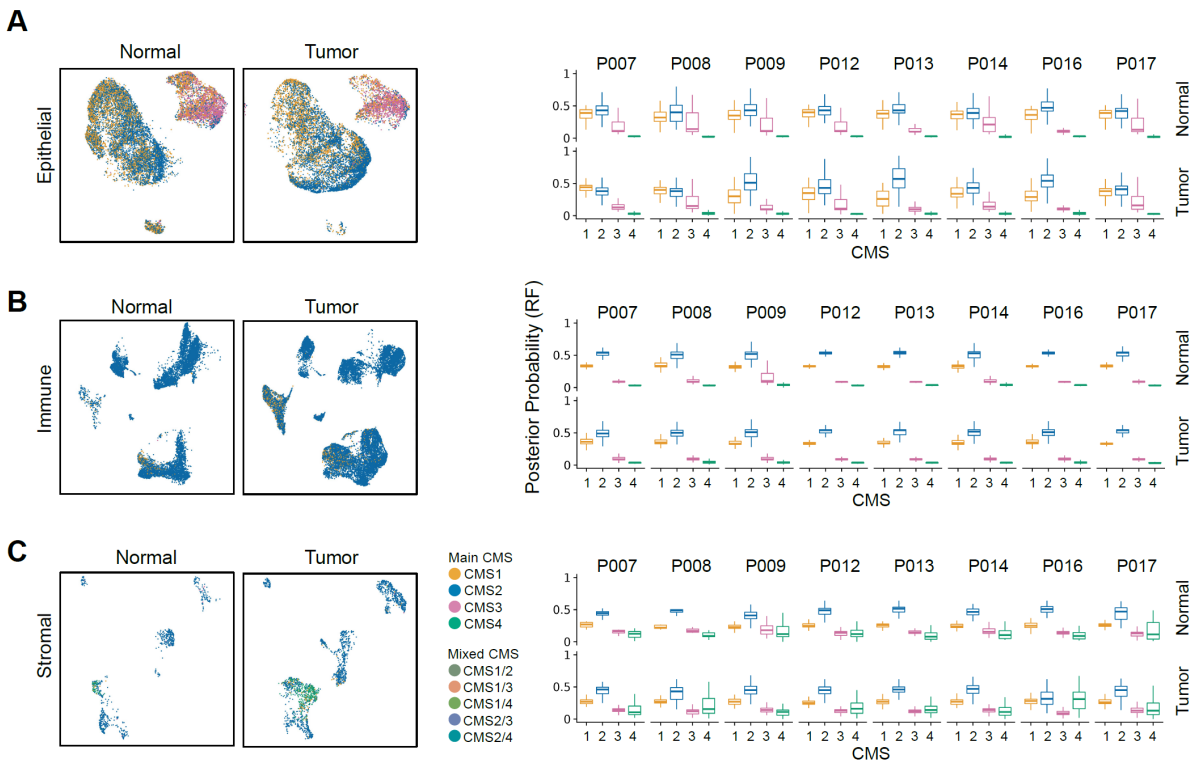

Figure 4: Consensus molecular subtype analysis on the single-cell level. A CMS calling on cells of the epithelial compartment. **B** CMS calling on cells of the immune compartment. **C** CMS calling on cells of the stromal compartment. In all subfigures, CMS subtypes are mapped to the normal and tumor tissue UMAPs to the left, and posterior probability scores for the complete tumor and normal cells in the respective compartments per patient are given to the right.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.10.901579; this version posted January 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

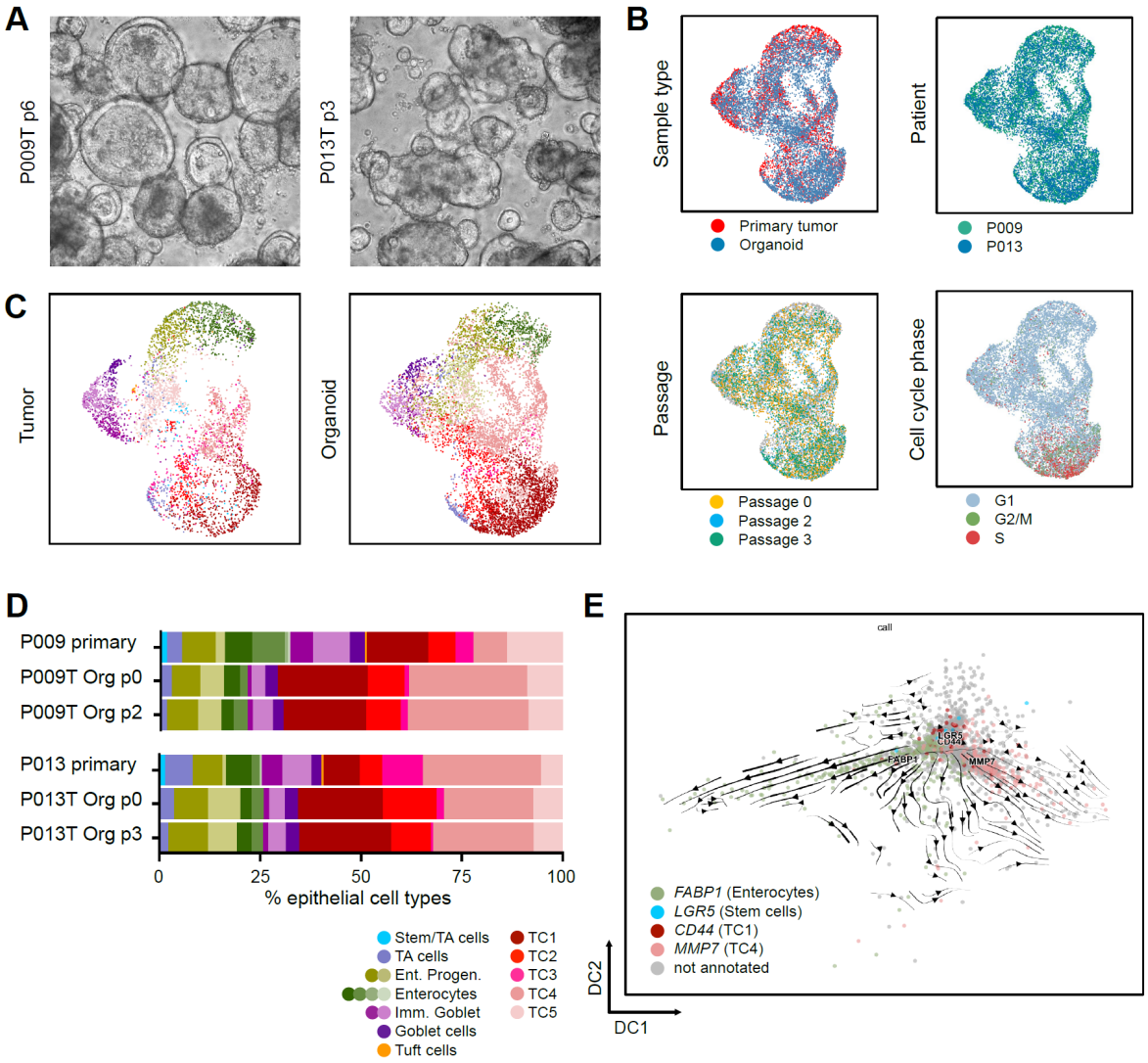

Figure 5: Epithelial cell types in matched tumors and organoids. A Phenotypes of organoid lines established from study patients. Patient and passage number are given to the left. **B** UMAPs of tumor-derived epithelial tissue profiles and anchored organoid profiles, color coded for sample type, patient, organoid passage number and cell cycle phase, as indicated. **C** UMAPs of tumor epithelial and organoid cells, color-coded by cell-type assignment. **D** Cell type distributions in tumor tissue and organoids. **E** RNA velocity of organoid line P009. Cells highly expressing the enterocyte marker FABP1, the normal stem cell markers *LGR5* the TC1 stem cell marker *CD44*, and the TC4 tumor cell marker *MMP7* are highlighted by green, blue and red and rose color, respectively. Arrows indicate two trajectories with a common root near the *LGR5*- and *CD44*-expressing cells, and extending towards the FABP1 and *MMP7*-positive cells.