Keystone pathobionts associated with colorectal cancer promote oncogenic reprograming

Josh Jones[1], Qiaojuan Shi[1], Rahul R. Nath[1], Ilana L. Brito[1]*

[1] Meinig School for Biomedical Engineering, Cornell University, Ithaca, NY

*Send all correspondences to: ibrito@cornell.edu

Abstract

Fusobacterium nucleatum (Fn) and enterotoxigenic *Bacteroides fragilis* (ETBF) are two pathobionts consistently enriched in the gut microbiomes of patients with colorectal cancer (CRC) compared to healthy counterparts and frequently observed for their direct association within tumors. Although several molecular mechanisms have been identified that directly link these organisms to features of CRC in specific cell types, their specific effects on the epithelium and local immune compartment are not well-understood. To fill this gap, we leveraged single-cell RNA sequencing (scRNA-seq) on wildtype mice and mouse model of CRC. We find that Fn and ETBF exacerbate cancer-like transcriptional phenotypes in transit-amplifying and mature enterocytes in a mouse model of CRC. We also observed increased T cells in the pathobiont-exposed mice, but these pathobiont-specific differences observed in wildtype mice were abrogated in the mouse model of CRC. Although there are similarities in the responses provoked by each organism, we find pathobiont-specific effects in Myc-signaling and fatty acid metabolism. These findings support a role for Fn and ETBF in potentiating tumorigenesis via the induction of a cancer stem cell-like transit-amplifying and enterocyte population and the disruption of CTL cytotoxic function.

Introduction

Colorectal cancer (CRC) is caused by both genetic mutations and aberrant features of the gut microbiome. Specifically, two organisms, *Fusobacterium nucleatum* (Fn) and enterotoxigenic *Bacteroides fragillis* (ETBF), are commonly enriched in the gut microbiomes of CRC patients[1–7] and exacerbate intestinal tumor formation in CRC mouse models[5,8]. Although a handful of molecular mechanisms have been identified that directly link these organisms with oncogenic pathways, less is known about how they affect distinct cell types within the intestinal compartment.

Fn was originally identified as an oral pathobiont due to its role in subgingival and periodontal disease[9,10], more recent studies find that Fn is associated with a number of cancers, including esophageal cell carcinoma[11,12], breast cancer[13], and most extensively with CRC[2,7,8,14–17]. Within CRC patients, Fn is spatially enriched in both adenomas and adenocarcinomas[7,14,16–18]. Fn is often present on CRC tumor tissue and this is linked to its expression of several adhesins, including FadA[19,20], and Fap2, the latter of which binds to the sugar residue, Gal-GalNAc[21,22], overexpressed on CRC tumors[23]. In addition to these associations, Fn has been shown to play a causative role in neoplastic transformation, with several recognized mechanisms. *Fusobacterium*-specific effector protein Fap2 interacts with TIGIT (T cell immunoreceptor with immunoglobulin and ITIM domain), a potent mediator of immunosuppression, leading to reduced natural killer cell and cytotoxic T cell mediated cytotoxicity[24]. Additionally, in *in vitro* and *in vivo* models of CRC, including the commonly used $Apc^{Min/+}$ mouse model, Fn protein FadA has been shown to bind to host cells and promote host DNA damage[25]. This consequently induces beta-catenin and Wnt signaling[26] and annexin A1 expression[27], which together trigger intestinal cell proliferation[8,28].

Under homeostatic conditions, non-toxigenic *B. fragilis* strains are highly prevalent gut commensals. However, certain *B. fragilis* strains express *B. fragilis* toxin (Bft) and are a common clinicopathological feature in inflammatory bowel disease (IBD)[29–31], diarrheal disease[32], and CRC[3–6]. ETBF has been shown to play a causal role in murine models of CRC. Specifically, Bft acts as a zinc-dependent metalloprotease that degrades E-cadherin, leading to aberrant signaling by beta-catenin and c-myc, both of which support enterocyte growth and

proliferation[5,33–36] . Furthermore, ETBF exposure elicits robust pro-tumorigenic IL-17 production and Th17 and T regulatory cell responses[37–40], further establishing a pro-oncogenic role for this pathobiont.

To investigate the effects of Fn and ETBF on host intestinal cells, we exposed a mouse model of CRC, as well as wildtype (WT) mice, to these organisms and performed single-cell RNA sequencing (scRNA-seq) on harvested intestinal resections. We utilized an established CRC mouse model that carries a transversion point mutation in one copy of tumor suppressor, *adenomatous polyposis coli* (*Apc*) ($Apc^{Min/+}$). The biallelic loss of *Apc* is detected in 80-90% of CRC patient cohorts and is an initiating event in sporadic CRC[41–43]. This mutation predisposes the mice to intestinal tumors and has been previously used to study the effects of both Fn and ETBF on tumor initiation and progression[8,15,41–44]. Comparing single-cell transcriptional profiles in resections from both WT and $Apc^{Min/+}$ mice afforded the opportunity to disentangle the combined effects of genetics and pathobionts on cellular phenotypes without imposing biases upon which cells these organisms most directly affect.

Results

Fn and ETBF alter intestinal cell composition in $Apc^{Min/+}$ and wildtype mice

To determine how CRC pathobionts affect the host intestinal microenvironment, we exposed WT and $Apc^{Min/+}$ mice to Fn or ETBF. Mice received a daily oral gavage of Fn or ETBF at a concentration of 10^8 colony forming units (CFUs) to expose intestinal cells to the pathobionts[8,15,44,45] (**Figure 1A**). Although Fn and ETBF have been reported to reduce survival rates and increase tumor burdens in $Apc^{Min/+}$ mice, these effects were limited to mice pre-treated with antibiotics[8,45–47]. Although antibiotic exposure is associated with increased CRC risk in humans[48–50], we chose not to pre-treat mice with antibiotics to avoid introducing confounding effects on host tissue either directly or via altered microbiome composition. Of note, this experimental procedure does deviate from established antibiotic-aided colonization methods and may explain why our downstream findings are different from the literature.[8,15,44,45] Nonetheless, we observed greater tumor burden 10-weeks after initial pathobiont exposure in the Fn-exposed $Apc^{Min/+}$ mice (**Figure 1B**), consistent with previous reports[8,51]. We were initially surprised that

ETBF administration did not result in increased tumor burden, as it does when ETBF is administered to antibiotic-treated *Apc*$^{Min/+}$ mice[40,44,45]. ETBF administration, under antibiotic treated conditions, elicits a robust IL-17 driven inflammatory response that mediates the recruitment of myeloid cells and ultimately supports tumor cell growth and proliferation in mice[52]. However, contrary to this pro-tumor phenotype, it is also been shown that ETBF does not increase the mutations-per-megabase and copy number alterations above that observed in *Apc*$^{Min/+}$ mice that have been pre-treated with antibiotics[47]. Taken together, without antibiotic-mediated colonization and the resultant inflammation, macroscopic tumor induction post-ETBF exposure was likely tempered.

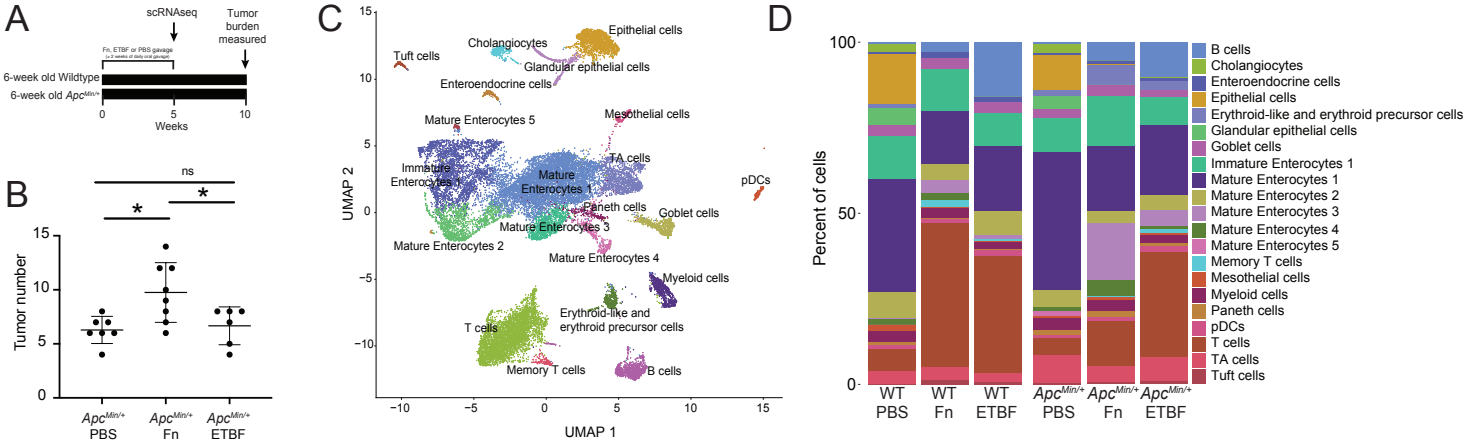


Figure 1. Exposure to CRC-associated pathobionts results in differences in cellular composition and transcriptional profiles.

(A) Depiction of the experiment. (B) Macroscopic tumor burden in *Apc*$^{Min/+}$ mice exposed to Fn or ETBF sacrificed at 16 weeks of age (n ≥ 6 *Apc*$^{Min/+}$ mice). Mice were exposed daily to CRC-associated pathobionts for at least 2 weeks starting at 6-weeks of age. (C) UMAP of transcriptomic profiles of 24,371 cells from all conditions colored according to their annotations. (D) A barplot depicting the composition of cells in each experimental condition.

We performed scRNA-seq on intestinal tissue from WT and *Apc*$^{Min/+}$ mice after oral dosing of Fn or ETBF, or phosphate buffered saline (PBS), as a control. Since Fn and ETBF are enriched in early stages of tumorigenesis (premalignant lesions and adenomas) in CRC patients[53–57], we sacrificed mice at 11-weeks of age corresponding to 5 weeks post-pathobiont exposure or PBS

treatment. We transcriptionally profiled 24,371 individual cells, which were clustered into 21 different cellular subsets, using Seurat (version 4.1.1)[58]. Cells were annotated with known cell-type specific marker genes[59,60] and cross-referenced using scMRMA, an automated single-cell annotation algorithm[61] (**Figure 1C**). Cellular compositions across treatment conditions were substantially different, including notable changes across T cells, proliferating enterocyte precursors, and mature enterocytes post-Fn and ETBF exposures (**Figure 1D**).

Fn and ETBF promote the outgrowth of cancer stem cell-like transit-amplifying cells and cancer-like enterocytes

Transit-amplifying (TA) cells are daughter cells of intestinal stem cells that further differentiate into enterocytes. Due to their high rates of proliferation, they are mutation-prone[62]. Treatment with Fn in co-culture with CRC cell lines has been found to induce the upregulation of stemness associated genes: *CD133*, *CD44, Snail1* and *ZEB1*[63,64]. Similarly, ETBF treatment leads to the increase in stemness in both CRC cell co-cultures and CRC xenograph mouse models, via the upregulation of *JMJD2B*, a histone demethylase[65]. We hypothesized that exposure to Fn and ETBF in $Apc^{Min/+}$ mice would exacerbate neoplastic transformation in these cells accordingly. TA cell transcriptomes sub-clustered into four distinct groups, including one that transcriptionally resembles cancer stem cells (CSCs), based similarities in upregulated genes and pathways between the cells we identified and the known phenotypic profile in the literature [66–69] (**Figure 2A-D)**. Using DEG analysis, we identified 91 genes delineating these CSC-like cells from the other TA cell subpopulations (**Figure 2B**). These include upregulated genes that support intestinal cell survival and proliferation, such as *Foxa1*[70–72], *Sox4*[71,73,74], *Prox1*[75–77], and *Ctnnb1*[78–80] (Fisher exact test p-values < 0.05, BH-FDR corrected p-values < 0.05, EnrichR). This subpopulation was almost exclusively found in the CRC pathobiont-exposed $Apc^{Min/+}$ mice (**Figure 2E**).

Overall, the CSC-like cells upregulated pro-oncogenic pathways, including integrin and integrin-linked kinase (ILK) signaling, MSP-RON (macrophage-stimulating protein-recepteur d'origine nantais) signaling, and Wnt/β-catenin signaling, among other pathways relating to stem cell pluripotency and the epithelial-mesenchymal transition (EMT)[81–86] (**Figure 2C)** (Fisher exact test p-values < 0.05, BH-FDR corrected p-values < 0.05, Ingenuity Platform Analysis (IPA)

canonical pathway analysis, the gene list used as the input for IPA was the result of a comparison (Wilcoxon test) between CSC-like cell clusters and the other three TA cell clusters). There were few significant differentially enriched pathways between these CSC-like TA cells specific to each pathobiont exposure, although Myc-targeting was comparably elevated in cells derived from *Apc*$^{Min/+}$ mice exposed to ETBF (**Supplemental Figure 1**). As for Fn-exposed CSC-like cell population, fatty acid metabolism was enriched compared to those exposed to ETBF, a finding which is supported by *in vitro* experiments linking this phenotype to enhanced self-renewal(**Supplemental Figure 1**)[63]. The top 50 most significant human gene-disease annotations for the DEGs in the CSC-like TA cell population are all cancers, including several related to the colon (**Figure 2D**) (Fisher exact test p-values < 0.05, BH-FDR corrected p-values < 0.05, DisGeNET). These colon-specific gene-disease annotations were unique to the CSC-like TA cells (**Supplemental Figure 2**). However, a second cluster of TA cells (proliferating TA cells 2) had similar gene-disease associations to the CSC-like TA cells, albeit different DEGs and enriched pathways. Interestingly, this cluster comprised predominantly cells from wildtype mice exposed to each of the pathobionts (**Figure 2E, Supplemental Figure 3**). These data suggest that exposure to CRC-associated pathobionts promotes the induction of cancer-stem cell-like cells within the *Apc*$^{Min/+}$ mice that possess transcriptomic hallmarks of human cancer stem cells.

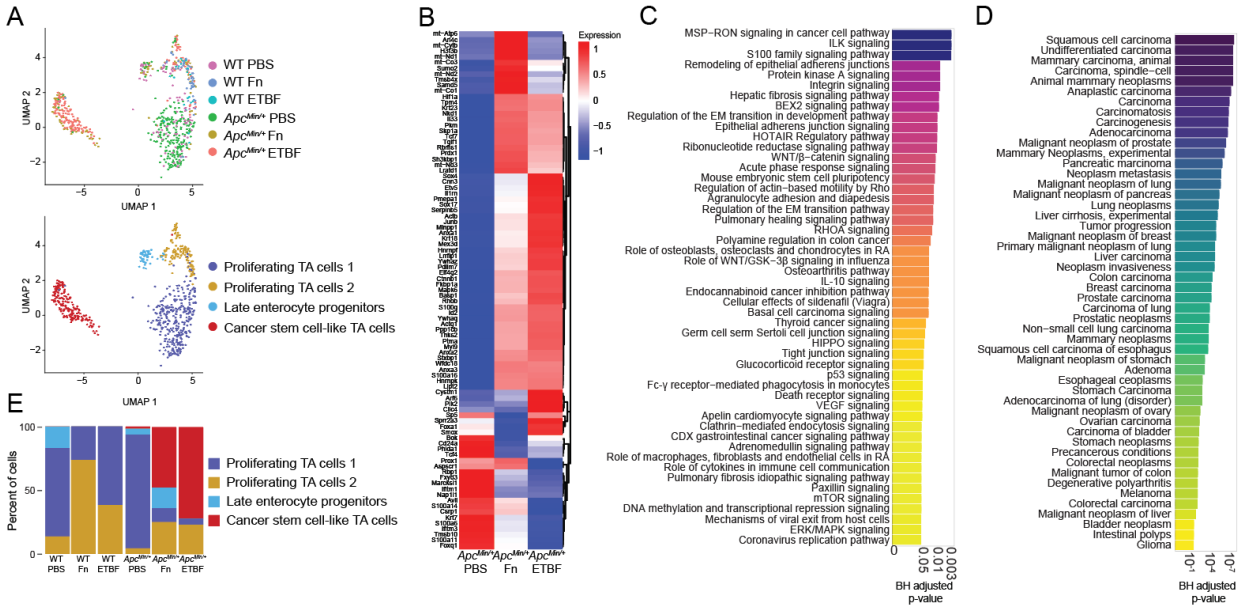


Figure 2. TA cells from *Apc*$^{Min/+}$ mice adopt cancer stem-cell like phenotypes after exposure to CRC-associated pathobionts.

(A) UMAP of transcriptomic profiles of TA cells according to experimental condition (top) and subclusters (bottom) (n=682). (B) A heatmap displaying all 91 upregulated genes for the CSC-like cell cluster (compared to the other TA populations) for each genotype-treatment, (log_2(fold-change) ≥ 0.25 (Wilcox test), Bonferroni-corrected p-value < 0.05, Seurat), plotted as average expression values. (C) A barplot depicting the top 50 IPA Canonical Pathways genesets for the cancer stem cell-like cell population, based on corrected p-values (BH-FDR-corrected p-value $<$ 0.05, IPA). (D) A barplot depicting the top 50 genesets according to DisGeNET (y-axis) for the CSC-like cell population, plotted in descending according to corrected p-values (Fisher exact test, BH-FDR corrected p-values < 0.05, EnrichR). (E) A barplot depicting the percent composition of the cell populations per genotype and treatment.

Mature enterocytes, derived from TA cells, are directly exposed to the microbiome and make up the vast majority of the cells within CRC tumors[68,87]. Both Fn and ETBF treatment increases tumor burden due to the outgrowth of transformed enterocytes in certain mouse models and drive rapid proliferation of epithelial cell lines in co-culture experiments[8,16,26,31,88,89]. Within the mature enterocyte cell population, we performed unsupervised clustering on cellular transcriptional profiles, resulting in four groups (**Figure 3A**). One group was noticeably enriched for cells derived from *Apc*$^{Min/+}$ mice exposed to Fn and ETBF and displayed a unique cancer-associated profile (**Figure 3B, Supplemental Figure 4**). Within this subset, 693 genes are differentially upregulated compared to the other three enterocyte sub-clusters, including the Wnt signaling mediator *Ctnnb1*, canonical cancer markers *STAT3* and *HIF1α*, and *Klf3*, *Klf4*, *Klf5* and *Klf6*, all of which exhibit tumor suppressive properties in many cancers, including CRC[80,90–92] (**Figure 3C, Supplemental Figure 4**). When compared to all other mature enterocyte sub-populations, the DEGs for this subset were enriched for genesets involved in PI3K/AKT/mTOR signaling, p53 signaling and apoptotic pathways (**Figure 3D**) (Fisher exact test p-values < 0.05, BH-FDR corrected p-values < 0.05, EnrichR). Analysis using the IPA platform was consistent with DisGeNET, showing a significant enrichment of disease and functional annotations associated with tumorigenesis (**Supplemental Figure 4**). We did not observe any significant differences in this sub-population that was specific to either Fn or ETBF (data not shown). Overall, these data suggest that this mature enterocyte population from pathobiont-exposed *Apc*$^{Min/+}$ mice adopts a cancer-like phenotype, like that observed in TA cells from the same mice.

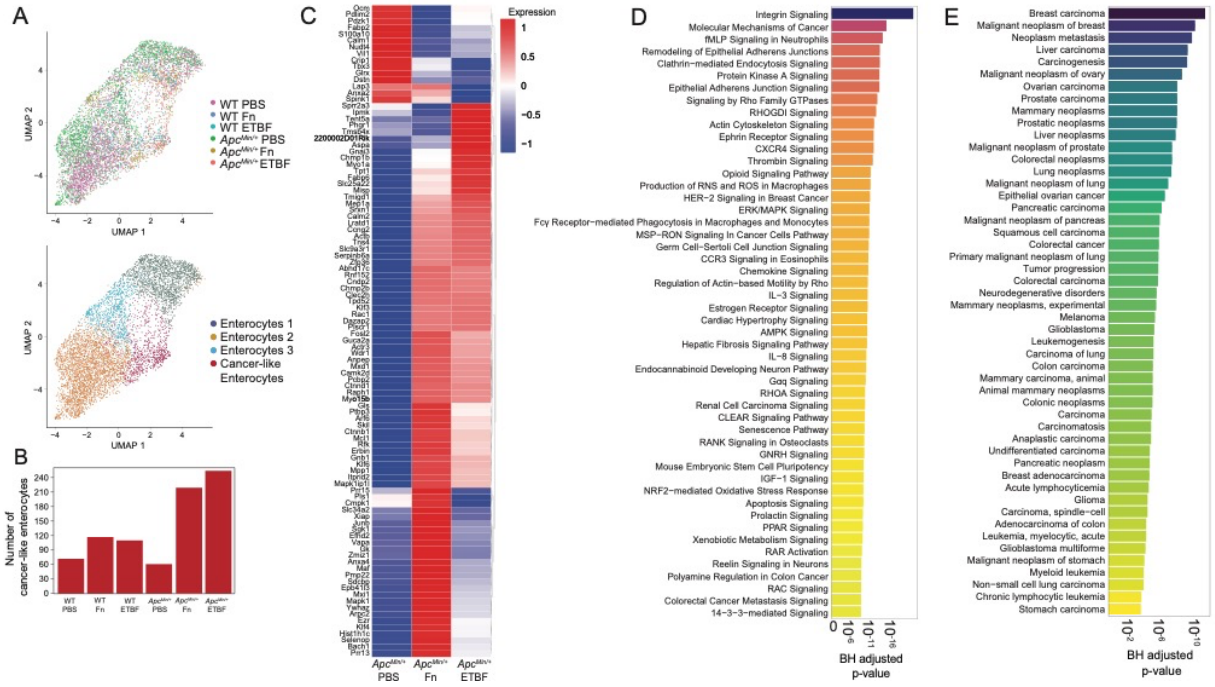


Figure 3. Mature enterocytes from pathobiont-treated *Apc*$^{Min/+}$ mice display cancer-like phenotypes.

(A) UMAP of transcriptomic profiles of 6,719 enterocyte populations colored by experimental condition(top) and by sub-clusters (bottom). (B) A barplot displaying the number of cells within each sub-cluster, according to experimental condition. (C) A clustered heatmap displaying the top 100 upregulated genes (log2(fold change) ≥ 0.25 (Wilcoxon test, BH-FDR corrected p-values < 0.05, Seurat), plotted as average expression values (Seurat) for the cancer-like enterocytes compared to all other enterocyte populations. (D)A barplot depicting the top 50 IPA Canonical Pathways genesets (y-axis) based on corrected p-values (Fisher exact test, BH-FDR corrected p-values < 0.05, IPA) for the cancer-like enterocytes. (E) A barplot depicting the top 50 genesets according to DisGeNET for the cancer-like enterocyte population, plotted in descending according to corrected p-values (Fisher exact test, BH-FDR corrected p-values < 0.05, EnrichR).

Together, these results support a model in which these pathobionts can influence cancer-associated signaling cascades, CRC initiation via CSC-like cell population induction and CRC progression by cancer-like enterocyte enrichment within the context of *Apc*$^{Min/+}$ mouse model. Supporting our work, a recent study investigating the interplay between Fn and human CRC tumors found that epithelial cell population with a high Fn burden upregulated Myc, mTORC1

and PI3K-AKT-mTOR signaling pathways. This important finding suggests that the enrichment of cell growth and proliferation signaling programs are a specific deleterious outcome elicited by Fn and in our study, ETBF as well.[7]

Pathobionts elicit similar effects in both-specific effects on cytotoxic T cells are abrogated in *Apc*$^{Min/+}$ mice

T cells are critical for tumor immunosurveillance[93,94]. However, the colorectal tumor microenvironment drives T cells, including potent anti-cancer CD8$^+$ cytotoxic T lymphocytes (CTLs), towards immunosuppressive, senescent, and exhaustive states[95–97]. In addition, CRC pathobionts Fn and ETBF exhibit profound T cell modulatory effects. In previous studies using *Apc*$^{Min/+}$ mice, ETBF exposure led to enhanced T cell differentiation skewing towards Th17 cells and away from CTLs, albeit this effect was indirect, mediated through the recruitment and activation of myeloid derived-suppressor cells (MDSCs)[40,98]. Similarly, Fn triggers the expansion of MDSCs in *Apc*$^{Min/+}$ mice, although without any effect on T cell populations[8]. However, in humans, Fn abundance within the tumor is inversely correlated with tumor-specific T cell abundances[99], and in cell culture, Fn directly binds human T cells and inhibits their function, potentially via interactions between TIGIT and Fn adhesin, Fap2[24,100]. Nevertheless, we did not observe specific changes involving TIGIT engagement because mouse TIGIT does not bind to Fap2[24]. To define the T cell subsets in our single-cell dataset, we characterized 3,101 T cells. The cells were partitioned using marker genes, yielding 4 subclusters: CTLs, γδ T cells, T regulatory cells, and mucosal-associated invariant T cells (**Figure 4A**). We focused on characterizing the CTL population, based on previous observations, and because they possess the cytotoxic function essential to the ablation of tumor growth. We also investigated whether microbe-specific transcriptional changes occurred in the myeloid cell compartments and although the myeloid cell counts were considerably low, proinflammatory macrophages derived from the Fn-treated *Apc*$^{Min/+}$ mouse were enriched for positive regulation of SMAD signaling and epithelial-to-mesenchymal transition compared with those from the WT mouse, though pathways did not pass the Bonferroni-correction threshold. (**Supplemental Figure 5**) (Fisher's exact test p-values < 0.05, Bonferroni-corrected p-values < 0.05, EnrichR). The numbers of CTLs isolated from the PBS control animals were also low and were therefore removed from downstream analyses. Of the genes that define the CTL cluster, made up of cells from

pathobiont-exposed mice, we observed that genes central to CTL function, including the cytolytic granule constituents, *Gzma* and *Gzmb*[101,102], and to a lesser extent *Cxcr6*[103–105], a chemokine receptor (not shown), are upregulated in the WT pathobiont-exposed mice, but not the $Apc^{Min/+}$ pathobiont-exposed mice (**Figure 4B and C**). These results suggest that the $Apc^{Min/+}$ background, possibly due to tumor-mediated immunosuppression, can mollify cytolytic CTL responses that are observed in wild-type post-pathobiont exposed counterparts.

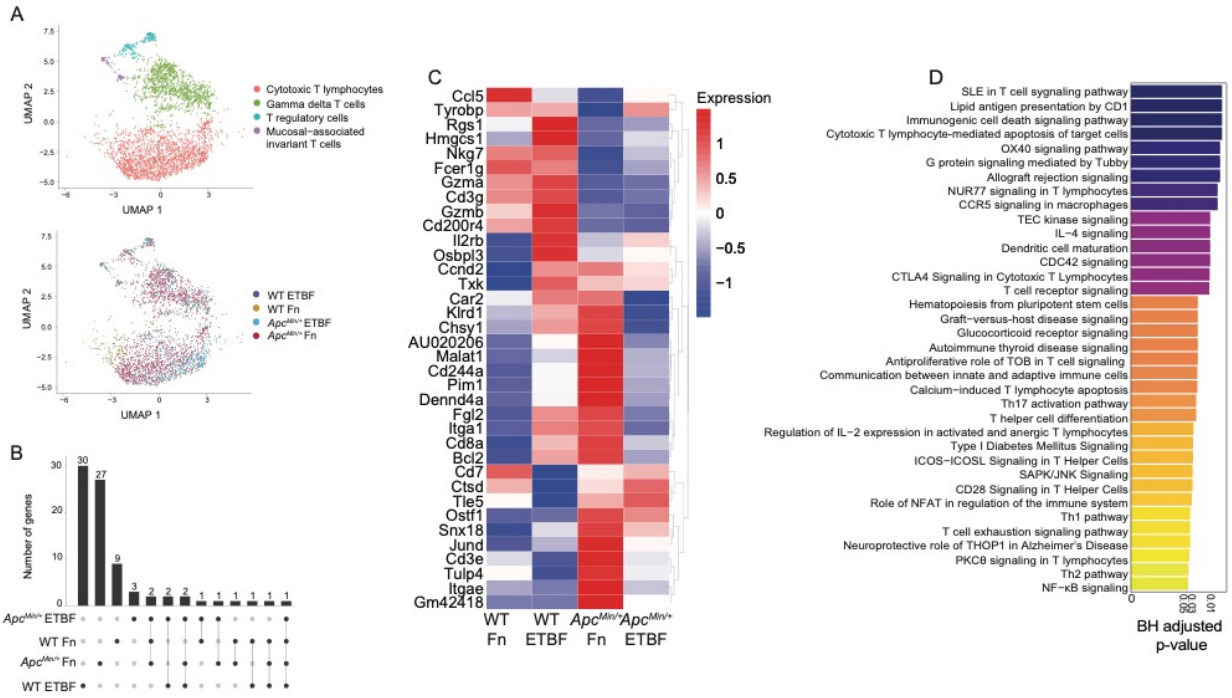


Figure 4. Pathobionts elicit similar effects in both-specific effects on cytotoxic T cells are abrogated in $Apc^{Min/+}$ mice. (A) UMAP of transcriptomic profiles of 3,101 T cell populations colored by sub-cluster (top) and by experimental condition (bottom). (B) Upset plot depicting the differentially expressed genes that each CTL population ($\log_2$(fold-change) ≥ 0.25, Wilcoxon test, BH-FDR-corrected p-value < 0.05) based on sample, the set size is the total number of genes expressed and the intersection size the number of genes that are shared by dataset, an individual sample alone indicating that the genes are only expressed by the cells in that dataset and lines representing shared genes. (C) A heatmap displaying the top 36 upregulated genes ($\log_2$(fold-change) ≥ 0.25, Wilcoxon test, BH-FDR-corrected p-value < 0.05), plotted as average expression values (Seurat) for the cytotoxic T lymphocytes across each dataset. (D) Barplot

depicting the top 36 IPA Canonical Pathways genesets (y-axis) based on corrected p-values (Fisher exact test, BH-FDR corrected p-values < 0.05, IPA). The gene list used as input for canonical pathway analysis were the genes upregulated by ETBF-exposed WT CTLs, when compared to ETBF-exposed *Apc*$^{Min/+}$ CTLs.

To better understand how the *Apc*$^{Min/+}$ model affects CTLs post-ETBF exposure, we compared the transcriptional profiles from ETBF-exposed *Apc*$^{Min/+}$ with ETBF-exposed WT mice. We found WT ETBF-exposed CTLs upregulated genesets involved in cytotoxic T lymphocyte−mediated apoptosis of target cells, T cell receptor signaling and OX40 signaling pathway[106–108], suggesting that ETBF treatment under normal conditions elicits a robust CTL response, and that this is suppressed in the *Apc*$^{Min/+}$ mice (Fisher's exact test p-values < 0.05, Bonferroni-corrected p-values < 0.05, IPA canonical pathway analysis) (**Figure 4D**). These results further support a model where CRC pathobionts induce T-cell dependent immunogenicity that is largely abrogated when tumors are present.

Discussion

Recent cancer pathophysiology studies have shown that the gut microbiota can play a significant role in tumor initiation, progression, or both[109–112]. Within CRC patients' gut microbiomes, organisms such as Fn and ETBF act as pathobionts, because of their ability to induce host inflammation, DNA damage, and cell proliferation[109–111]. These bacteria are thought to initiate the formation of carcinogenic bacterial biofilms and antagonize host immunity by tempering anti-tumor immunity[14,15,24]. Despite a growing body of evidence supporting the role of bacteria in CRC tumor burden and patient survival[109–111], much of the work uncovering the mechanisms underpinning this phenomena have been restricted to experiments using cell culture or on specific cell types isolated from mouse models.

The scRNA-seq data presented here suggests that there are cell-specific and pathobiont-specific effects evident in immune and epithelial tissue. Our analysis reveals that Fn and ETBF can provoke a CSC-like transcriptional profile in TA cells. These CSC-like TA cells bridge pathophysiological observations with specific cellular responses, including, but not limited to, known stemness genes. Moreover, mature enterocytes, which appear to be susceptible to

neoplastic transformation, are an emergent feature of *Apc*$^{Min/+}$ intestinal cell profiles post-Fn and ETBF exposure. CTLs, on the other hand, displayed transcriptiomes evident of reduced cytotoxic capacity in pathobiont-exposed *Apc*$^{Min/+}$ mice, when compared to their pathobiont-exposed wildtype counterparts. By directly comparing Fn- and ETBF-exposed mice, we observed consistent features invoked by both pathobionts in TA, enterocyte and CTL populations. These results suggest that pathobiont exposure can foster an environment conducive to the outgrowth of tumorigenic intestinal cell populations.

The effects on TA cells, enterocytes and cytotoxic T lymphocytes that we observe were each affected by the underlying genetic background of the CRC mouse model we used. The *Apc*$^{Min/+}$mouse model recapitulates a relevant mutation in human CRC (80-90% of all sporadic CRC cases) and is therefore the most widely utilized mouse model for CRC. However, there are some notable differences between this model's pathophysiology versus that which is observed in humans. For instance, the primary site of tumorigenesis in the *Apc*$^{Min/+}$ mouse is the small intestine, rather than the colon[113]. Examining the effects of CRC-associated pathobionts in additional mouse models of CRC, including those that exhibit greater colonic tumor burden (*e.g.* mice carrying inducible mutations in *Apc*, *Kras*, and *p53* specific to the colon, such as those driven by Villin or Cdx2)[114] could further enhance our understanding of colon-specific tumorigenesis mediated by Fn and ETBF. Notwithstanding these alternatives, the *Apc*$^{Min/+}$ model affords the ability to elucidate microbe-specific transcriptional responses in a system free of numerous cancer drivers and in a model within which these organisms have shown to affect tumorigenesis.

This study demonstrates the effects of repeat exposure to CRC pathobionts. There are several limitations of our experimental design. First, we did not use antibiotics nor germ-free mice, as we wanted to maintain the native murine microbiome. This came with the caveat that without antibiotics, Fn and ETBF colonization is not robust. We tracked colonization through the study using qPCR with Fn and ETBF specific probes and found that Fn and ETBF engraftment was often below the limit of detection (data not shown). Our results highlight the cellular effects of short-term repeat exposure on intestinal tissue. These results support the hit-and-run carcinogenesis model[115–118], whereby CRC pathobionts exposure is transient but the pro-tumor

effects elicited pathobionts manifest by experimental endpoints. Additionally, we were interested in providing a detailed single-cell characterization of both epithelial subtypes and immune cells from both small intestine and colon. For that reason, we pooled and sequenced cells from both anatomical sites. By doing this, we were able to capture epithelial cell heterogeneity, including the detection and characterization of cancer stem cell-like transit-amplifying cells and cancer-like enterocytes. While this method of single-cell preparation reduced our ability to capture immune cells and other lower abundance cell types such as Paneth and enteroendocrine cells in particular, we avoided examining transcriptional changes induced by cell enrichment methods[119].

Transient exposure, rather than colonization, may have tempered the pro-tumorigenic effects of ETBF (Figure 1B), and possibly Fn, via niche exclusion and/or colonization resistance[120–122]. Moreover, transient exposure and lack of antibiotic use could limit the pathobiont's access to many of the cell populations traditionally associated with their pathogenic inflammatory etiology such T cells and macrophages, which largely are in the lamina propria, and spatial distance from direct interactions with Fn and ETBF, and their pathogen-associated molecular patterns[123,124]. Nonetheless, we still find that transient exposure to Fn and ETBF in the $Apc^{Min/+}$ model triggers transcriptional programs that support the outgrowth of CSC-like cells and cancer-like enterocytes. Similar short-term exposures to ETBF induces robust cytotoxic T cell responses in wildtype mice. Taken together, this suggests that Fn and ETBF pro-tumor effects could be more robust than previous thought.

Fn and ETBF are known for their ability to trigger distinct tumor promoting mechanisms. Fn adhesin FadA modulates aberrant Wnt signaling via E-cadherin and β-catenin in enterocytes[26,27]. ETBF possesses a DNA damaging toxin, Bft, and induces Myc signaling in enterocytes and an inflammatory immune cascade largely mediated by Th17 cells and IL-17[34,35,38]. One of our study's important findings is that Fn and ETBF, despite their unique tumorigenic proclivities, mostly overlap mechanistically as evidenced by the similar cancer-associated transcriptional programs evoked in enterocyte and enterocyte pre-cursors (Table 1). This suggests that both organisms have common CRC initiating and/or supporting characteristics that affect similar cell types. These findings were enabled by the significant number of enterocytes sequenced across our murine intestinal samples. Herein lies a key shortcoming as well, which does not represent

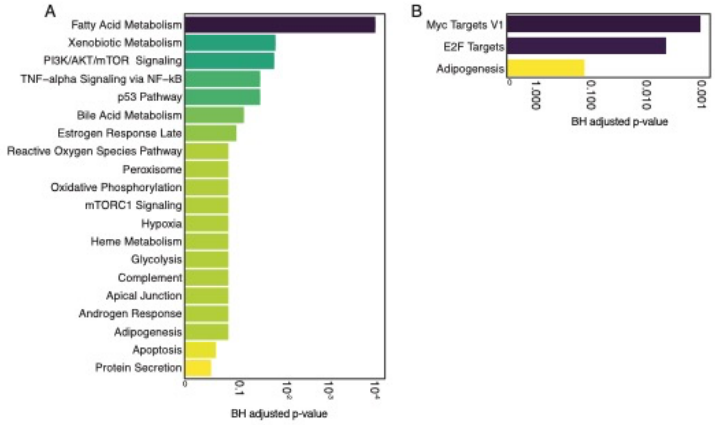
common biology. By probing thousands of enterocytes, other rarer cell types were found in smaller numbers. For this reason, comparative analyses between Fn and ETBF treatments across almost all other cell types, including across both *Apc*$^{Min/+}$ and wild-type mice, were underpowered, and we could not delineate statistically significant differences (BH corrected p-value < 0.05). Nevertheless, our findings still represent an important step in delineating enterocyte and TA cell-specific transcriptomic changes post CRC pathobiont exposure and warrants future investigations delving into larger swath of intestinal cells in depth.

Although Fn and ETBF are perhaps the most well-known CRC-associated pathobionts, a fuller picture of CRC initiation and progression likely involves other key microbial players. For example, *pks*$^{+}$ *E. coli* is an *E. coli* strain that produces colibactin, a genotoxin that cause double strand breaks in the intestinal cells' DNA also has the ability to transform cells[125–127]. The development of polymicrobial biofilms is another emergent feature of CRC. Biofilms are significantly enriched in right sided colon adenomas (precancerous lesion) versus adjacent healthy tissue and have been causally linked with CRC in mouse models[14,15,128]. Additionally, other oral pathobionts beyond Fn, such as *Parvimonas micra*, *Peptostreptococcus stomatis*, *Peptostreptococcus anaerobius* and *Gemella morbillorum*, are commonly enriched in patients with CRC[111,129,130]. Experimentally, *P. anaerobius* and *P. micra* having been shown to play a causal role in oncogenesis in azoxymethane and *Apc*$^{Min/+}$ mouse models, respectively[131,132]. Pertaining to these organisms, major questions in the field remained about how these oral microbes, in concert with gut pathobionts, seed biofilms and, if so, whether the biofilms promote tumorigenesis in the colon[126,133–136]. Performing similarly designed scRNA-seq experiments using additional organisms and eventually consortia will likely be invaluable in delineating the modulatory effects gut bacteria have on CRC tumor initiation and development.

Tumor-specific microbiomes, biofilm formation, and microbiome dysbiosis are all implicated in CRC progression. Using scRNA-seq, we were able to reconstruct cell type-specific effects that occur post-pathobiont exposure. However, recently developed approaches that enable combined host transcriptomics with microbiome species mapping[137,138] will provide additional spatial contextualization, directly associating specific gut microbiota with cell-specific transcriptional changes occurring within the tumor microenvironment. Studying the effects of Fn, ETBF and

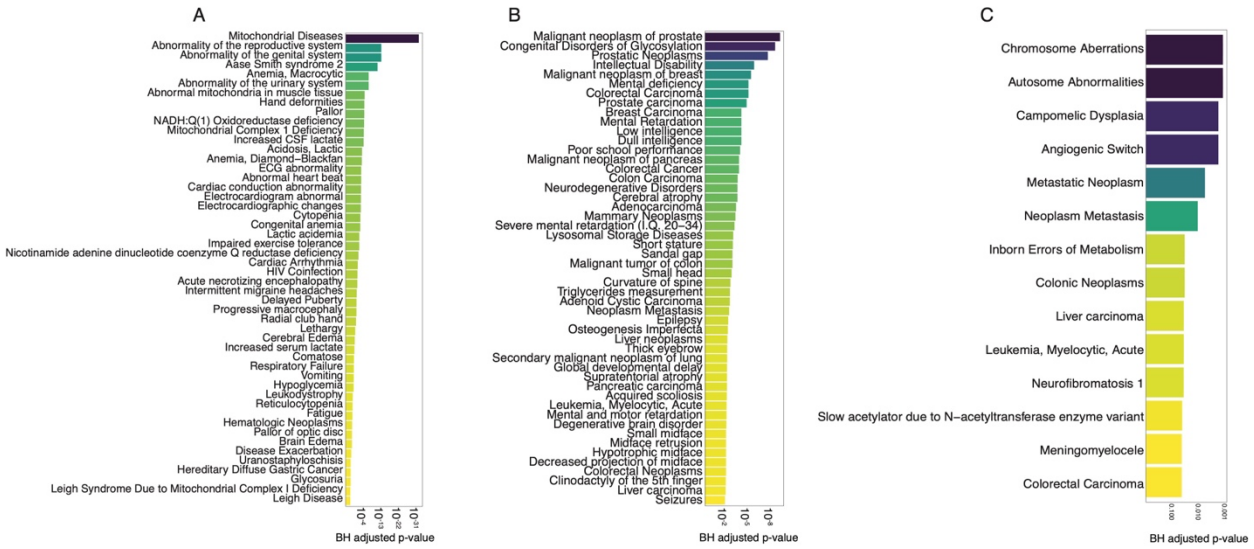
other pathobionts *in vivo*, using unbiased approaches like these offer the promise of identifying marker genes that may be used to enhance cancer diagnostics and therapeutics.

Supplemental Figures



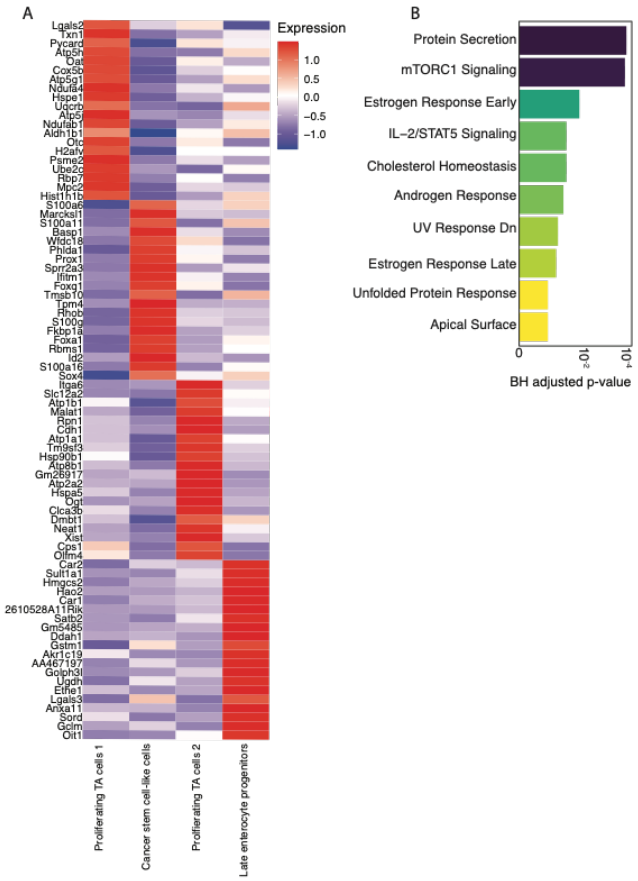
Supplemental Figure 1. CSC-like TA cells from Fn- and ETBF-exposed *Apc*$^{Min/+}$ mice differed in key pathways.

(A) Top 20 differentially enriched pathways (MSigDB Hallmarks 2020) represented in the transcriptomes of cells from CSC-like TA cells from the Fn-exposed *Apc*$^{Min/+}$ mouse as compared to the PBS-treated *Apc*$^{Min/+}$ mouse. (n= 175 cells, Fisher exact test, BH-FDR-corrected p-values < 0.05, EnrichR) (B) Top 3 differentially enriched pathways (MSigDB Hallmarks 2020) represented in the transcriptomes of cells from CSC-like TA cells from the ETBF-exposed *Apc*$^{Min/+}$ mouse as compared to the PBS-treated *Apc*$^{Min/+}$ mouse. (n= 175 cells, Fisher exact test, BH-FDR-corrected p-values < 0.05, EnrichR). Supplemental Figure 1 complements Figure 2.

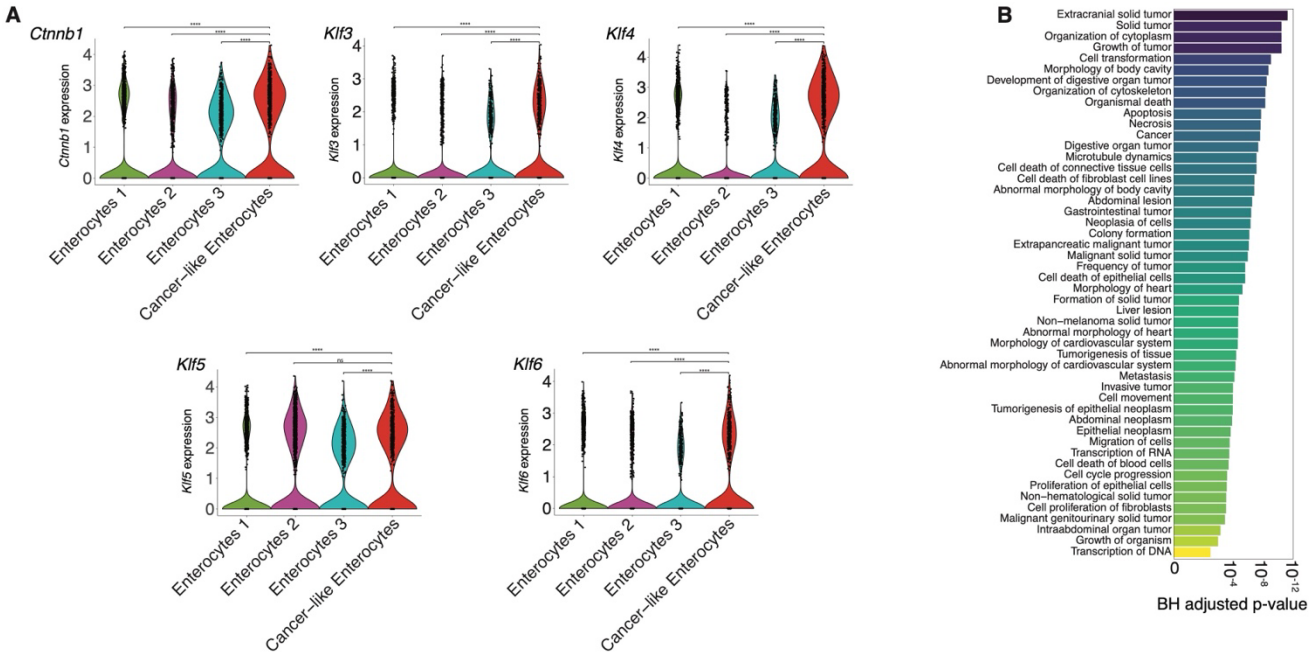


Supplemental Figure 2. Cancer-specific gene-disease associations with DEGs identified in TA cells were specific to those from pathobiont-exposed *Apc*$^{Min/+}$ mice.

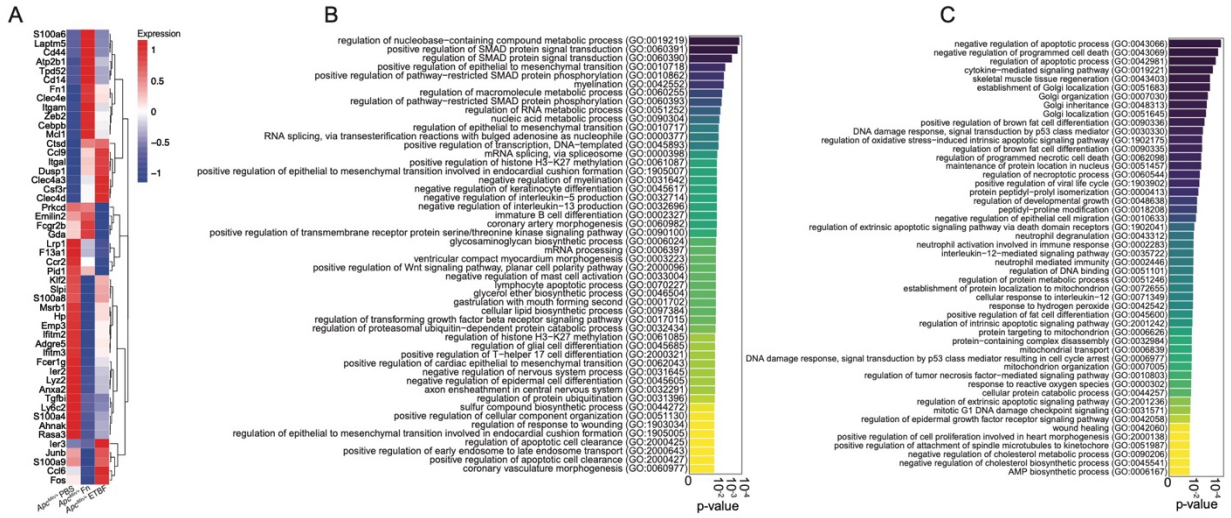
(A) A barplot depicting the top 50 genesets according to DisGeNET (y-axis) for the proliferating TA cells (1), plotted in descending according to corrected p-values (x-axis, Fisher exact test, BH-FDR corrected p-values < 0.05, EnrichR). (B) A barplot depicting the top 50 genesets according to DisGeNET (y-axis) for the proliferating TA cells (2), plotted in descending according to corrected p-values (Fisher exact test, BH-FDR-corrected p-values < 0.05, EnrichR). (C) A barplot depicting the top 14 genesets according to DisGeNET (y-axis) for the late enterocyte progenitors, plotted in descending according to corrected p-values (Fisher exact test, BH-FDR corrected p-values < 0.05, EnrichR). Supplemental Figure 2 complements Figure 2.



Supplemental Figure 3. Proliferating TA cells 2, similar to CSC-like TA cells in notable disease associations, diverge at the gene and pathway levels. (A) The TA cells depicted here are the 4 subclusters of the complete TA cell population and are an aggregate from all mouse samples (*Apc*$^{Min/+}$ mice treated with PBS, Fn or ETBF and wild type mice treated with PBS, Fn or ETBF. A heatmap displaying the top 20 upregulated genes for each TA cluster, $\log_2$(fold-change) ≥ 0.25 (Wilcox test), corrected p-value < 0.05 (Bonferroni correction), Seurat), plotted as average expression values (Seurat). (B) Differentially enriched pathways represented in the transcriptomes of proliferating TA cells 2 compared with other TA cell populations. Barplot depicting the top 10 genesets according to the Molecular Signatures Database Hallmark 2020 (MSigDB Hallmarks 2020) for the cancer-like cell population, plotted in descending according to corrected p-values (Fisher exact test, BH-FDR corrected p-values < 0.05, EnrichR). Supplemental Figure 3 complements Figure 2 and Supplemental Figure 2.



Supplemental Figure 4. Transcriptome profiles of cancer-like enterocytes were enriched in cancer-like genes and pathways. (A) Violin plots displaying selected CRC-associated genes and their expression levels across 4 enterocyte clusters ($\log_2$(fold-change) ≥ 0.25, Wilcoxon test, Bonferroni-corrected p-value < 0.05). (B) Barplot depicting the top 50 IPA Diseases and Functions annotations based on corrected p-values (Fisher exact test, BH-FDR corrected p-values < 0.05,) for the cancer-like enterocyte subpopulation. Statistical comparisons were performed using a pairwise Wilcoxon test (* = $p \leq 0.05$, ** = $p \leq 0.01$, *** = $p \leq 0.001$, **** = $p \leq 0.0001$), comparing the cancer-like enterocyte population to all other mature enterocyte clusters (see Supplemental Figure 4A). Supplemental Figure 4 complements Figure 3.



Supplemental Figure 5. Proinflammatory macrophages derived from the Fn-exposed *Apc*$^{Min/+}$ mouse upregulate pathways associated with TGF-β/SMAD signaling and epithelial-to-mesenchymal transition. (A) A heatmap displaying the top 50 upregulated genes defining the proinflammatory macrophage population compared across each dataset (log_2(fold-change) ≥ 0.25, Wilcoxon Rank Sum test, p-value < 0.05 (unadjusted), Seurat), plotted as average expression values. (B) Barplot depicting the top 50 enriched genesets according to the Gene Ontology Biological Processes 2021 (GOBP21) for proinflammatory macrophages derived from Fn-exposed *Apc*$^{Min/+}$ mice when compared to PBS control *Apc*$^{Min/+}$ mice, plotted in descending according to p-values (Fisher exact test p-values < 0.05, unadjusted, EnrichR). (C) Barplot depicting the top 50 enriched genesets according to the Gene Ontology Biological Processes 2021 (GOBP21) for proinflammatory macrophages derived from Fn-exposed *Apc*$^{Min/+}$ mice when compared to ETBF exposed *Apc*$^{Min/+}$ mice, plotted in descending according to p-values (Fisher exact test p-values < 0.05, unadjusted, EnrichR).

Materials and Methods

Ethical considerations

This study conformed to the National Institutes of Health guidelines on the care and use of laboratory animals. Mouse studies were performed following procedures approved by the Institutional Animal Care and Use Committee at Cornell University (Protocol ID #2016-0088)

468

469 **Bacterial Strains and Culturing**

470 *Fusobacterium nucleatum* subsp. nucleatum strain VPI 4355 [1612A] (ATCC 25586) was
471 purchased from American Type Culture Collection (ATCC). *Bacteroides fragilis* (Veillon and
472 Zuber) Castellani and Chalmers strain 2-078382-3 (ATCC 43858) (ETBF) was purchased from
473 American Type Culture Collection (ATCC). Fn and ETBF were grown anaerobically at 37°C on
474 Bacto™ Brain Heart Infusion Broth (BD, Sparks, MD) supplemented with 0.01% Hemin in 1M
475 NaOH, 0.1% Resazurin (25 mg/100ml distilled water), 10% NaHCO3 in distilled water, and agar
476 if bacteria were plated. Bacteria were grown overnight and diluted to 10^8 colony forming units
477 (CFU), the amount needed for oral gavage.

478

479 **Mice**

480 All mice (C57BL/6-*Apc*$^{Min/+}$/J and C57BL/6-Wild type) were maintained at the barrier mouse
481 facility at Weill Hall at Cornell University. *Apc*$^{Min/+}$ and wild-type mice were initially ordered
482 from Jackson Laboratory and then bred in the barrier facility. The *Apc*$^{Min/+}$ mice used in these
483 experiments have a chemically induced transversion point mutation (a T to an A) at nucleotide
484 2549. This results in a stop codon at codon 850, truncating the APC protein. Experimental and
485 breeding mice were provided with *ad libitum* access to autoclaved water and rodent chow
486 (autoclavable Teklad global 14% protein rodent maintenance diet #2014-S; Envigo). To avoid
487 cage effects on the microbiota, mice were housed individually at the time of initial Fn or ETBF
488 exposure. To monitor for infectious agents such as helminths, sentinel mice were used during the
489 duration of the experiment in the mouse facility to ensure that results following perturbation with
490 Fn and ETBF were a result of specific bacteria and not confounding agents. Every week, food
491 intake and animal weight were recorded, and mice were placed in clean cages with freshly
492 autoclaved chow and water weekly. Mice were handled under inside a biosafety cabinet with
493 frequent glove changes and disinfection between mice during stool collection and monitoring of
494 body weight. Stool was collected weekly throughout the course of all experiments. Bacterial oral
495 gavage experiments were performed every day for a period of at least 14 consecutive days for
496 ETBF, and up 35 days for Fn[8,25,45], beginning at 6 weeks of age. Bacteria were fed at a
497 concentration of 10^8 CFU per day. Sham treatment consisted of sterile Ca^{2+} and Mg^{2+} free
498 phosphate buffered saline gavaged daily for the entirety of the experiment. Single-cell RNA

499 experiments concluded when the mice were 11 weeks old and tumor burden experiments
500 concluded when mice were 16 weeks old.
501
502 **Tumor burden enumeration**
503 For tumor enumeration, *Apc*$^{Min/+}$ mice were euthanized at 16 weeks of age, and colons and small
504 intestines were excised. Macroscopic tumors were counted from both anatomical sites. The
505 tumor counts were plotted using Prism (version 8.2.1). For statistical analysis, Mann-Whitney
506 two-tailed tests were used to compare treatment groups using Prism. Each groups had an $n \geq 6$
507 mice.
508
509 **Single cell dissociation from fresh mouse colons and small intestines**
510 This protocol was adapted from Haber et al 2017[59]. To generated single-cell suspensions,
511 *Apc*$^{Min/+}$ and wild type mice were euthanized at 11 weeks of age, colons and small intestines
512 were excised, rinsed with ice cold sterile 1X Ca^{2+} and Mg^{2+} free PBS (Gibco, 14190144) and
513 flushed of fecal contents using a blunt 1.5-inch 22G needle filled with ice cold sterile 1X Ca^{2+}
514 and Mg^{2+} free PBS (Gibco, 14190144). The tissue was opened longitudinally and sliced into
515 small fragments roughly 1 cm in length. The tissue was incubated in RPMI supplemented with
516 L-glutamine (Corning, 45000-396), 1 mM EDTA (Neta Scientific, QB-A611-E177-10), and 10%
517 FBS (Avantor, 97068-085) for 90 minutes, shaking every 30 minutes. The tissue was then
518 incubated at 37°C for 15 minutes and continuously shaken. The supernatant was passed through
519 a 100 μm cell strainer and held on ice until loading the cells on 10X Chromium. The remaining
520 tissue was resuspended in RPMI (Corning, 45000-396) supplemented with 20% FBS (Avantor,
521 97068-085), 0.1 mg/ml DNase I (Thermo Scientific, 90083), and 0.5 mg/ml collagenase A
522 (Millipore Sigma, 10103586001) and incubated at 37°C on a shaker for 30 minutes. The tissue
523 was then gently mechanically dissociated using a rubber plunger of a syringe. The tissue and the
524 dissociated contents were passed through a 100 μm cell strainer. The single cell suspension was
525 then pelleted via centrifugation (400 x g for 10 minutes at 4°C). The cell suspension was
526 resuspended in 1X Ca^{2+} and Mg^{2+} free PBS (Gibco, 14190144) containing 0.04% weight/volume
527 BSA (VWR, 97061-420) and combined with earlier collected fraction and placed on ice. Sample
528 viability was determined before loading the cells on 10X Chromium using the Countess II

Automated Cell Counter (ThermoFisher). The desired number of transcriptomes from viable cells for each sample was 5000-6000 cells per sample.

Single-cell RNA sequencing library preparation

5000-6000 viable (≥ 70% alive) cells per sample (from colon and small intestine tissues) were targeted on the 10X Genomics Controller using one lane per mouse/sample for Gel Beads in Emulsion (GEM). Cells from the small intestine and colon were pooled together before GEM creation. Briefly, cells were separated into GEMs along with beads coated in oligos that capture mRNAs using a poly-dT sequences. This was followed by cell lysis and barcoded reverse transcription of mRNA, followed by amplification, and enzymatic fragmentation and 5′ adaptor and sample index attachment. Single-cell libraries were generated using the Chromium Next GEM Single Cell 3' Library Construction V3 Kit (10X Genomics) and were then sequenced on an Illumina NextSeq 2000 run with the 100 bp P2 kit for all samples. Sequencing data were aligned to the mouse reference, mm10 (Ensembl 84) reference genome using the Cell Ranger 5.0.1 pipeline (10X Genomics).

Single-cell RNAseq data processing and visualization

The output of Cell Ranger is a cell-by-gene unique molecular identifier (UMI) expression matrix for each sample. The expression matrices for each sample are loaded into the Seurat R package (Seurat version 4.1.1, R version 4.1.0 and 4.2.0). The standard Seurat dataset processing workflow was followed. In brief, cells with less than 200 genes, more than 2,500 genes, and more than 35% mitochondrial genes are filtered out. After filtering, the remaining cells were normalized by the total expression, multiplied by the default scale factor (10,000), and log transformed. We then used default Seurat functions to identify highly variable genes with one parameter modification. FindVariableFeatures' nfeature parameter was set to 3,000 instead of 2,000 (default). Next, we scaled the data to regress out variation from mitochondrial genes. We performed principal component analysis (PCA) on the scaled data with variable genes. The top 20 principal components were used for downstream analysis, including dimensionality reduction steps including clustering cells to identify cell populations (clusters). We implemented Uniform Manifold Approximation and Projection for dimensional reduction using the top 20 PCs and visualized.

Marker-gene identification and cell-type annotation

To define cell types for each cluster, we used Seurat's FindAllMarkers with the following parameters: a minimum percent expression value of 25%, $\log_2$fold change threshold of 0.25 and a corrected p-value < 0.05 (Bonferroni correction). We looked only at transcripts that were upregulated. We analyzed canonical markers and assigned cell annotations accordingly (see Supplemental Table 1). We cross-referenced our cell type annotations with gene lists defined in Haber et al.[59] and Moor et al.[60] We cross-reference the cell type assignments with a single cell annotation algorithm, scMRMA in R as well.[61]

Reclustering, visualization, and analysis of transit-amplifying cells, mature enterocyte (1) and T cell populations

We used the 682 TA cells, 6,719 mature enterocytes (1), and 3,101 T cells and re-clustered them using Seurat. Marker genes for each subclusters were identified using a minimum percent expression value of 25%, $\log_2$fold change threshold of 0.25 and a corrected p-value < 0.05 (Bonferroni correction) in Seurat. Cell types were assigned based on the expression of these marker genes. Cell clusters expressing marker genes from multiple unrelated cell types (doublets) were removed from analysis. All sub-clustering analysis was carried out with 20 principal components and similar resolution parameters; TA cells and T cells were analyzed with a resolution of 0.4 and mature enterocytes (1) with a resolution of 0.3 in Seurat. The marker gene list used to classify cell subtypes can be found in Supplemental Table 1. Cell populations were visualized using Uniform Manifold Approximation and Projection in Seurat. Cell were enumerated, whether as percent of sample or absolute count, using the dittoSeq's (version 1.8.1) bar plot visualization function.

Differential gene expression and geneset enrichment analysis

Differentially gene expression was carried out using Seurat's FindAllMarkers and FindMarkers functions with the following cutoffs: $\log_2$(fold change) ≥ 0.25 (Wilcox test), corrected p-value < 0.05 (Bonferroni correction) and a minimum percent expression value of either the default, 10%, or 25% for certain other analyses. For these analyses, only upregulated genes were used. We visualized DEGs using the Seurat's DoHeatmap and dittoHeatmap (dittoSeq) for heatmaps,

dittoPlot(dittoSeq) for violin plots and UpSetR (version 1.4.0) for upset plots. For statistics associated with violin plots (Supplemental Figure 4), we performed a two-sample Wilcoxon test, comparing each normal enterocyte cluster against the cancer-like enterocyte cluster using the stat_compare_means function in ggpubr (version 0.5.0). For gene set enrichment analysis, the gene list used as input were generated as detailed above using FindMarkers (Seurat). A suite of tools and databases were implement for these analysis and are as follows: Ingenuity Pathway Analysis (IPA, Qiagen) including canonical pathway and disease and function analysis, DisGeNET (version 7.0) via Enrichr[139,140], and MSigDB Hallmarks 2020 via EnrichR.[140]

Acknowledgements

We thank the de Vlaminck lab for helpful conversations. This work was supported in part by a grant from the National Institutes of Health (1R33CA235302-01A1). J.J. was funded by a fellowship from the Center for Vertebrate Genomics at Cornell University. I.L.B. was funded by the National Institutes of Health (1DP2HL141007), a Pew Foundation Research Fellowship and a Packard Foundation Fellowship.

Data Availability

Single-cell RNA-seq data are being deposited at NCBI GEO and will be made publicly available upon publication.

References

1. Kostic, A. D. *et al.* Genomic analysis identifies association of Fusobacterium with colorectal carcinoma. *Genome Res* **22**, 292–298 (2012).
2. Castellarin, M. *et al.* Fusobacterium nucleatum infection is prevalent in human colorectal carcinoma. *Genome Research* **22**, 299–306 (2012).
3. Viljoen, K. S., Dakshinamurthy, A., Goldberg, P. & Blackburn, J. M. Quantitative profiling of colorectal cancer-associated bacteria reveals associations between Fusobacterium spp.,

enterotoxigenic Bacteroides fragilis (ETBF) and clinicopathological features of colorectal cancer. *PLoS ONE* **10**, 1–21 (2015).

4. Boleij, A. *et al.* The bacteroides fragilis toxin gene is prevalent in the colon mucosa of colorectal cancer patients. *Clinical Infectious Diseases* **60**, 208–215 (2015).
5. Haghi, F., Goli, E., Mirzaei, B. & Zeighami, H. The association between fecal enterotoxigenic B. fragilis with colorectal cancer. *BMC Cancer* **19**, 879 (2019).
6. Jasemi, S. *et al.* Toxigenic and non-toxigenic patterns I, II and III and biofilm-forming ability in Bacteroides fragilis strains isolated from patients diagnosed with colorectal cancer. *Gut Pathogens* **12**, 1–7 (2020).
7. Niño, J. L. G. *et al.* Effect of the intratumoral microbiota on spatial and cellular heterogeneity in cancer. *Nature* 1–8 (2022) doi:10.1038/s41586-022-05435-0.
8. Kostic, A. D. *et al.* Fusobacterium nucleatum Potentiates Intestinal Tumorigenesis and Modulates the Tumor-Immune Microenvironment. *Cell Host & Microbe* **14**, 207–215 (2013).
9. Han, Y. W. *et al.* Interactions between periodontal bacteria and human oral epithelial cells: Fusobacterium nucleatum adheres to and invades epithelial cells. *Infect Immun* **68**, 3140–3146 (2000).
10. Liu, B. *et al.* Deep Sequencing of the Oral Microbiome Reveals Signatures of Periodontal Disease. *PLOS ONE* **7**, e37919 (2012).
11. Zhang, N. *et al.* Clinical Significance of Fusobacterium nucleatum Infection and Regulatory T Cell Enrichment in Esophageal Squamous Cell Carcinoma. *Pathol Oncol Res* **27**, 1609846 (2021).

12. Nomoto, D. *et al.* Fusobacterium nucleatum promotes esophageal squamous cell carcinoma progression via the NOD1/RIPK2/NF-κB pathway. *Cancer Lett* **530**, 59–67 (2022).
13. Parhi, L. *et al.* Breast cancer colonization by Fusobacterium nucleatum accelerates tumor growth and metastatic progression. *Nat Commun* **11**, 3259 (2020).
14. Dejea, C. M. *et al.* Microbiota organization is a distinct feature of proximal colorectal cancers. *Proc Natl Acad Sci U S A* **111**, 18321–18326 (2014).
15. Dejea, C. M. *et al.* Patients with familial adenomatous polyposis harbor colonic biofilms containing tumorigenic bacteria. *Science* **359**, 592–597 (2018).
16. Yu, T. C. *et al.* Fusobacterium nucleatum Promotes Chemoresistance to Colorectal Cancer by Modulating Autophagy. *Cell* **170**, 548-563.e16 (2017).
17. Wang, S. *et al.* Fusobacterium nucleatum Acts as a Pro-carcinogenic Bacterium in Colorectal Cancer: From Association to Causality. *Front Cell Dev Biol* **9**, 710165 (2021).
18. Gao, Y. *et al.* Fusobacterium nucleatum enhances the efficacy of PD-L1 blockade in colorectal cancer. *Sig Transduct Target Ther* **6**, 1–10 (2021).
19. Ikegami, A., Chung, P. & Han, Y. W. Complementation of the fadA mutation in Fusobacterium nucleatum demonstrates that the surface-exposed adhesin promotes cellular invasion and placental colonization. *Infect Immun* **77**, 3075–3079 (2009).
20. Meng, Q. *et al.* Fusobacterium nucleatum secretes amyloid-like FadA to enhance pathogenicity. *EMBO Rep* **22**, e52891 (2021).
21. Coppenhagen-Glazer, S. *et al.* Fap2 of Fusobacterium nucleatum is a galactose-inhibitable adhesin involved in coaggregation, cell adhesion, and preterm birth. *Infect Immun* **83**, 1104–1113 (2015).

662 22. Abed, J. *et al.* Fap2 Mediates Fusobacterium nucleatum Colorectal Adenocarcinoma
663 Enrichment by Binding to Tumor-Expressed Gal-GalNAc. *Cell Host Microbe* **20**, 215–225
664 (2016).
665 23. Yang, G. Y. & Shamsuddin, A. M. Gal-GalNAc: a biomarker of colon carcinogenesis. *Histol*
666 *Histopathol* **11**, 801–806 (1996).
667 24. Gur, C. *et al.* Binding of the Fap2 protein of fusobacterium nucleatum to human inhibitory
668 receptor TIGIT protects tumors from immune cell attack. *Immunity* **42**, 344–355 (2015).
669 25. Guo, P. *et al.* FadA promotes DNA damage and progression of Fusobacterium nucleatum-
670 induced colorectal cancer through up-regulation of chk2. *J Exp Clin Cancer Res* **39**, 202
671 (2020).
672 26. Rubinstein, M. R. *et al.* Fusobacterium nucleatum promotes colorectal carcinogenesis by
673 modulating E-cadherin/β-catenin signaling via its FadA adhesin. *Cell Host Microbe* **14**, 195–
674 206 (2013).
675 27. Rubinstein, M. R. *et al.* Fusobacterium nucleatum promotes colorectal cancer by inducing
676 Wnt/β-catenin modulator Annexin A1. *EMBO Rep* **20**, e47638 (2019).
677 28. Wang, N. & Fang, J.-Y. Fusobacterium nucleatum, a key pathogenic factor and microbial
678 biomarker for colorectal cancer. *Trends in Microbiology* **0**, (2022).
679 29. Prindiville, T. P. *et al.* Bacteroides fragilis enterotoxin gene sequences in patients with
680 inflammatory bowel disease. *Emerg Infect Dis* **6**, 171–174 (2000).
681 30. Rabizadeh, S. *et al.* Enterotoxigenic Bacteroides fragilis: A Potential Instigator of Colitis.
682 *Inflamm Bowel Dis* **13**, 1475–1483 (2007).

683 31. Cao, Y. *et al.* Enterotoxigenic Bacteroides fragilis Promotes Intestinal Inflammation and
684 Malignancy by Inhibiting Exosome-Packaged miR-149-3p. *Gastroenterology* **161**, 1552-
685 1566.e12 (2021).
686 32. Sears, C. L. *et al.* Association of Enterotoxigenic Bacteroides fragilis Infection with
687 Inflammatory Diarrhea. *Clin Infect Dis* **47**, 797–803 (2008).
688 33. Shang, S., Hua, F. & Hu, Z.-W. The regulation of β-catenin activity and function in cancer:
689 therapeutic opportunities. *Oncotarget* **8**, 33972–33989 (2017).
690 34. Wu, S., Morin, P. J., Maouyo, D. & Sears, C. L. Bacteroides fragilis enterotoxin induces c-Myc
691 expression and cellular proliferation. *Gastroenterology* **124**, 392–400 (2003).
692 35. Wu, S. *et al.* The Bacteroides fragilis toxin binds to a specific intestinal epithelial cell
693 receptor. *Infection and Immunity* **74**, 5382–5390 (2006).
694 36. Toprak, N. U. *et al.* A possible role of Bacteroides fragilis enterotoxin in the aetiology of
695 colorectal cancer. *Clin Microbiol Infect* **12**, 782–786 (2006).
696 37. Geis, A. *et al.* Enterotoxigenic Bacteroides fragilis induces oncogenic regulatory T cells
697 (TUM9P.1000). *The Journal of Immunology* **194**, 210.2-210.2 (2015).
698 38. Geis, A. L. *et al.* Regulatory T cell response to enterotoxigenic Bacteroides fragilis
699 colonization triggers IL-17-dependent colon carcinogenesis. *Cancer Discov* **5**, 1098–1109
700 (2015).
701 39. Housseau, F. *et al.* Redundant innate and adaptive sources of IL-17 production drive colon
702 tumorigenesis. *Cancer Res* **76**, 2115–2124 (2016).
703 40. Thiele Orberg, E. *et al.* The myeloid immune signature of enterotoxigenic Bacteroides
704 fragilis-induced murine colon tumorigenesis. *Mucosal Immunology* **10**, 421–433 (2017).

41. Moser, A. R. *et al.* ApcMin: a mouse model for intestinal and mammary tumorigenesis. *Eur J Cancer* **31A**, 1061–1064 (1995).

42. Kwong, L. N. & Dove, W. F. APC and its modifiers in colon cancer. *Adv Exp Med Biol* **656**, 85–106 (2009).

43. Nguyen, H. T. & Duong, H.-Q. The molecular characteristics of colorectal cancer: Implications for diagnosis and therapy (Review). *Oncology Letters* **16**, 9–18 (2018).

44. Housseau, F. & Sears, C. L. Enterotoxigenic Bacteroides fragilis (ETBF)-mediated colitis in Min (Apc+/-) mice: A human commensal-based murine model of colon carcinogenesis. *Cell Cycle* **9**, 3–5 (2010).

45. Chung, L. *et al.* Bacteroides fragilis Toxin Coordinates a Pro-carcinogenic Inflammatory Cascade via Targeting of Colonic Epithelial Cells. *Cell Host and Microbe* **23**, 203-214.e5 (2018).

46. Yang, Y. *et al.* Fusobacterium nucleatum Increases Proliferation of Colorectal Cancer Cells and Tumor Development in Mice by Activating Toll-Like Receptor 4 Signaling to Nuclear Factor−κB, and Up-regulating Expression of MicroRNA-21. *Gastroenterology* **152**, 851-866.e24 (2017).

47. Allen, J. *et al.* Colon Tumors in Enterotoxigenic Bacteroides fragilis (ETBF)-Colonized Mice Do Not Display a Unique Mutational Signature but Instead Possess Host-Dependent Alterations in the APC Gene. *Microbiology Spectrum* **10**, e01055-22 (2022).

48. Lu, S. S. M. *et al.* Antibiotics Use and Subsequent Risk of Colorectal Cancer: A Swedish Nationwide Population-Based Study. *JNCI: Journal of the National Cancer Institute* **114**, 38–46 (2022).

727 49. Zhang, J. *et al.* Oral antibiotic use and risk of colorectal cancer in the United Kingdom, 1989-
728 2012: a matched case-control study. *Gut* **68**, 1971–1978 (2019).
729 50. Armstrong, D. *et al.* The association between colorectal cancer and prior antibiotic
730 prescriptions: case control study. *Br J Cancer* **122**, 912–917 (2020).
731 51. Yang, Y. *et al.* Fusobacterium nucleatum Increases Proliferation of Colorectal Cancer Cells
732 and Tumor Development in Mice by Activating TLR4 Signaling to NFκB, Upregulating
733 Expression of microRNA-21. *Gastroenterology* **152**, 851-866.e24 (2017).
734 52. Orberg, E. T. *et al.* The Myeloid Immune Signature of Enterotoxigenic Bacteroides Fragilis-
735 Induced Murine Colon Tumorigenesis. *Mucosal Immunol* **10**, 421–433 (2017).
736 53. Wong, S. H. *et al.* Quantitation of faecal Fusobacterium improves faecal immunochemical
737 test in detecting advanced colorectal neoplasia. *Gut* **66**, 1441–1448 (2017).
738 54. Zhang, M. *et al.* Differential Mucosal Microbiome Profiles across Stages of Human
739 Colorectal Cancer. *Life (Basel)* **11**, 831 (2021).
740 55. Purcell, R. V., Visnovska, M., Biggs, P. J., Schmeier, S. & Frizelle, F. A. Distinct gut
741 microbiome patterns associate with consensus molecular subtypes of colorectal cancer.
742 *Scientific Reports* **7**, 1–12 (2017).
743 56. Zamani, S. *et al.* Enterotoxigenic Bacteroides fragilis: A Possible Etiological Candidate for
744 Bacterially-Induced Colorectal Precancerous and Cancerous Lesions. *Frontiers in Cellular*
745 *and Infection Microbiology* **9**, (2020).
746 57. Tjalsma, H., Boleij, A., Marchesi, J. R. & Dutilh, B. E. A bacterial driver-passenger model for
747 colorectal cancer: beyond the usual suspects. *Nat Rev Microbiol* **10**, 575–582 (2012).

748 58. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587.e29
749 (2021).
750 59. Haber, A. L. *et al.* A single-cell survey of the small intestinal epithelium. *Nature* **551**, 333–
751 339 (2017).
752 60. Moor, A. E. *et al.* Spatial Reconstruction of Single Enterocytes Uncovers Broad Zonation
753 along the Intestinal Villus Axis. *Cell* **175**, 1156-1167.e15 (2018).
754 61. Li, J., Sheng, Q., Shyr, Y. & Liu, Q. scMRMA: single cell multiresolution marker-based
755 annotation. *Nucleic Acids Research* **50**, e7 (2022).
756 62. Huels, D. J. & Sansom, O. J. Stem vs non-stem cell origin of colorectal cancer. *Br J Cancer*
757 **113**, 1–5 (2015).
758 63. Liu, H. *et al.* Fusobacterium nucleatum Promotes Colorectal Cancer Cell to Acquire Stem
759 Cell-Like Features by Manipulating Lipid Droplet-Mediated Numb Degradation. *Adv Sci*
760 *(Weinh)* **9**, 2105222 (2022).
761 64. Wang, Q., Yu, C., Yue, C. & Liu, X. Fusobacterium nucleatum produces cancer stem cell
762 characteristics via EMT-resembling variations. *Int J Clin Exp Pathol* **13**, 1819–1828 (2020).
763 65. Liu, Q.-Q. *et al.* Enterotoxigenic Bacteroides fragilis induces the stemness in colorectal
764 cancer via upregulating histone demethylase JMJD2B. *Gut Microbes* **12**, 1788900.
765 66. Zhou, Y. *et al.* Cancer stem cells in progression of colorectal cancer. *Oncotarget* **9**, 33403–
766 33415 (2017).
767 67. Munro, M. J., Wickremesekera, S. K., Peng, L., Tan, S. T. & Itinteang, T. Cancer stem cells in
768 colorectal cancer: a review. *Journal of Clinical Pathology* **71**, 110–116 (2018).

68. Becker, W. R. *et al.* Single-cell analyses define a continuum of cell state and composition changes in the malignant transformation of polyps to colorectal cancer. *Nat Genet* **54**, 985–995 (2022).

69. Wang, H. *et al.* Colorectal Cancer Stem Cell States Uncovered by Simultaneous Single-Cell Analysis of Transcriptome and Telomeres. *Adv Sci (Weinh)* **8**, 2004320 (2021).

70. Yue, M., Yun, Z., Li, S., Yan, G. & Kang, Z. NEDD4 triggers FOXA1 ubiquitination and promotes colon cancer progression under microRNA-340-5p suppression and ATF1 upregulation. *RNA Biology* **18**, 1981–1995 (2021).

71. Lazar, S. B. *et al.* Genome-Wide Analysis of the FOXA1 Transcriptional Network Identifies Novel Protein-Coding and Long Noncoding RNA Targets in Colorectal Cancer Cells. *Mol Cell Biol* **40**, e00224-20 (2020).

72. Park, Y.-L. *et al.* Forkhead-box A1 regulates tumor cell growth and predicts prognosis in colorectal cancer. *Int J Oncol* **54**, 2169–2178 (2019).

73. Wang, B., Li, Y., Tan, F. & Xiao, Z. Increased expression of SOX4 is associated with colorectal cancer progression. *Tumour Biol* **37**, 9131–9137 (2016).

74. Liu, J. *et al.* SOX4 maintains the stemness of cancer cells via transcriptionally enhancing HDAC1 revealed by comparative proteomics study. *Cell & Bioscience* **11**, 23 (2021).

75. Petrova, T. V. *et al.* Transcription factor PROX1 induces colon cancer progression by promoting the transition from benign to highly dysplastic phenotype. *Cancer Cell* **13**, 407–419 (2008).

76. Ragusa, S. *et al.* PROX1 promotes metabolic adaptation and fuels outgrowth of Wnt(high) metastatic colon cancer cells. *Cell Rep* **8**, 1957–1973 (2014).

77. Wiener, Z. *et al.* Prox1 promotes expansion of the colorectal cancer stem cell population to fuel tumor growth and ischemia resistance. *Cell Reports* **8**, 1943–1956 (2014).

78. Fevr, T., Robine, S., Louvard, D. & Huelsken, J. Wnt/β-Catenin Is Essential for Intestinal Homeostasis and Maintenance of Intestinal Stem Cells. *Molecular and Cellular Biology* **27**, 7551–7559 (2007).

79. Li, B. *et al.* Impaired Wnt/β-catenin pathway leads to dysfunction of intestinal regeneration during necrotizing enterocolitis. *Cell Death Dis* **10**, 1–11 (2019).

80. Gregorieff, A. & Clevers, H. Wnt signaling in the intestinal epithelium: from endoderm to cancer. *Genes Dev.* **19**, 877–890 (2005).

81. Zheng, C.-C. *et al.* Significance of integrin-linked kinase (ILK) in tumorigenesis and its potential implication as a biomarker and therapeutic target for human cancer. *Am J Cancer Res* **9**, 186–197 (2019).

82. TSOUMAS, D. *et al.* ILK Expression in Colorectal Cancer Is Associated with EMT, Cancer Stem Cell Markers and Chemoresistance. *Cancer Genomics Proteomics* **15**, 127–141 (2018).

83. Almasabi, S., Ahmed, A. U., Boyd, R. & Williams, B. R. G. A Potential Role for Integrin-Linked Kinase in Colorectal Cancer Growth and Progression via Regulating Senescence and Immunity. *Front Genet* **12**, 638558 (2021).

84. Huang, L. *et al.* MSP-RON Pathway: Potential Regulator of Inflammation and Innate Immunity. *Front Immunol* **11**, 569082 (2020).

85. Li, C. *et al.* MSP-RON Signaling Is Activated in the Transition From Pancreatic Intraepithelial Neoplasia (PanIN) to Pancreatic Ductal Adenocarcinoma (PDAC). *Frontiers in Physiology* **10**, (2019).

86. Yao, H.-P., Zhou, Y.-Q., Zhang, R. & Wang, M.-H. MSP-RON signalling in cancer: pathogenesis and therapeutic potential. *Nature Reviews Cancer* **13**, 466–482 (2013).

87. Li, J., Ma, X., Chakravarti, D., Shalapour, S. & DePinho, R. A. Genetic and biological hallmarks of colorectal cancer. *Genes Dev.* **35**, 787–820 (2021).

88. Bao, Y. *et al.* Long noncoding RNA BFAL1 mediates enterotoxigenic Bacteroides fragilis-related carcinogenesis in colorectal cancer via the RHEB/mTOR pathway. *Cell Death and Disease* **10**, (2019).

89. Purcell, R. V., Permain, J. & Keenan, J. I. Enterotoxigenic Bacteroides fragilis activates IL-8 expression through Stat3 in colorectal cancer cells. *Gut Pathog* **14**, 16 (2022).

90. Ghaleb, A. M. & Yang, V. W. The Pathobiology of Krüppel-like Factors in Colorectal Cancer. *Curr Colorectal Cancer Rep* **4**, 59–64 (2008).

91. Cao, D. *et al.* Expression of HIF-1alpha and VEGF in colorectal cancer: association with clinical outcomes and prognostic implications. *BMC Cancer* **9**, 432 (2009).

92. Corvinus, F. M. *et al.* Persistent STAT3 Activation in Colon Cancer Is Associated with Enhanced Cell Proliferation and Tumor Growth. *Neoplasia* **7**, 545–555 (2005).

93. Borst, J., Ahrends, T., Bąbała, N., Melief, C. J. M. & Kastenmüller, W. CD4+ T cell help in cancer immunology and immunotherapy. *Nat Rev Immunol* **18**, 635–647 (2018).

94. Philip, M. & Schietinger, A. CD8+ T cell differentiation and dysfunction in cancer. *Nat Rev Immunol* **22**, 209–223 (2022).

95. Di, J. *et al.* Phenotype molding of T cells in colorectal cancer by single-cell analysis. *International Journal of Cancer* **146**, 2281–2295 (2020).

834 96. Zhang, L. *et al.* Lineage tracking reveals dynamic relationships of T cells in colorectal cancer.
835 *Nature* **564**, 268–272 (2018).
836 97. Kim, C. G. *et al.* VEGF-A drives TOX-dependent T cell exhaustion in anti–PD-1–resistant
837 microsatellite stable colorectal cancers. *Science Immunology* **4**, eaay0555 (2019).
838 98. Geis, A. L. & Housseau, F. Procarcinogenic regulatory T cells in microbial-induced colon
839 cancer. *OncoImmunology* **5**, e1118601 (2016).
840 99. Mima, K. *et al.* Fusobacterium nucleatum and T Cells in Colorectal Carcinoma. *JAMA Oncol*
841 **1**, 653–661 (2015).
842 100. Galaski, J. *et al.* Fusobacterium nucleatum CbpF Mediates Inhibition of T Cell Function
843 Through CEACAM1 Activation. *Frontiers in Cellular and Infection Microbiology* **11**, (2021).
844 101. Cullen, S. P., Brunet, M. & Martin, S. J. Granzymes in cancer and immunity. *Cell Death*
845 *Differ* **17**, 616–623 (2010).
846 102. Trapani, J. A. Granzymes: a family of lymphocyte granule serine proteases. *Genome Biol*
847 **2**, reviews3014.1-reviews3014.7 (2001).
848 103. Wang, B. *et al.* CXCR6 is required for antitumor efficacy of intratumoral CD8+ T cell. *J*
849 *Immunother Cancer* **9**, e003100 (2021).
850 104. Muthuswamy, R. *et al.* CXCR6 by increasing retention of memory CD8+ T cells in the
851 ovarian tumor microenvironment promotes immunosurveillance and control of ovarian
852 cancer. *J Immunother Cancer* **9**, e003329 (2021).
853 105. Di Pilato, M. *et al.* CXCR6 positions cytotoxic T cells to receive critical survival signals in
854 the tumor microenvironment. *Cell* **184**, 4512-4530.e22 (2021).

855 106. Pan, P.-Y., Zang, Y., Weber, K., Meseck, M. L. & Chen, S.-H. OX40 ligation enhances
856 primary and memory cytotoxic T lymphocyte responses in an immunotherapy for hepatic
857 colon metastases. *Mol Ther* **6**, 528–536 (2002).
858 107. Pham Minh, N. *et al.* In vivo antitumor function of tumor antigen-specific CTLs
859 generated in the presence of OX40 co-stimulation in vitro. *Int J Cancer* **142**, 2335–2343
860 (2018).
861 108. Bansal-Pakala, P., Halteman, B. S., Cheng, M. H.-Y. & Croft, M. Costimulation of CD8 T
862 Cell Responses by OX40. *The Journal of Immunology* **172**, 4821–4825 (2004).
863 109. Wirbel, J. *et al.* Meta-analysis of fecal metagenomes reveals global microbial signatures
864 that are specific for colorectal cancer. *Nat Med* **25**, 679–689 (2019).
865 110. Thomas, A. M. *et al.* Metagenomic analysis of colorectal cancer datasets identifies cross-
866 cohort microbial diagnostic signatures and a link with choline degradation. *Nat Med* **25**,
867 667–678 (2019).
868 111. Yachida, S. *et al.* Metagenomic and metabolomic analyses reveal distinct stage-specific
869 phenotypes of the gut microbiota in colorectal cancer. *Nature Medicine* **25**, 968–976
870 (2019).
871 112. Sepich-poore, G. D. *et al.* The microbiome and human cancer. **4552**, (2021).
872 113. Muzny, D. M. *et al.* Comprehensive molecular characterization of human colon and
873 rectal cancer. *Nature* **487**, 330–337 (2012).
874 114. Byun, A. *et al.* Colon-specific tumorigenesis in mice driven by Cre-mediated inactivation
875 of Apc and activation of mutant Kras. *Cancer Lett* **347**, 191–195 (2014).

876 115. Knippel, R. J., Drewes, J. L. & Sears, C. L. The Cancer Microbiome: Recent Highlights and
877 Knowledge Gaps. *Cancer Discov* **11**, 2378–2395 (2021).
878 116. Sears, C. L. & Garrett, W. S. Microbes, Microbiota, and Colon Cancer. *Cell Host &*
879 *Microbe* **15**, 317–328 (2014).
880 117. Hatakeyama, M. Helicobacter pylori CagA and Gastric Cancer: A Paradigm for Hit-and-
881 Run Carcinogenesis. *Cell Host & Microbe* **15**, 306–316 (2014).
882 118. Ternes, D. *et al.* Microbiome in Colorectal Cancer: How to Get from Meta-omics to
883 Mechanism? *Trends in Microbiology* **28**, 401–423 (2020).
884 119. van den Brink, S. C. *et al.* Single-cell sequencing reveals dissociation-induced gene
885 expression in tissue subpopulations. *Nat Methods* **14**, 935–936 (2017).
886 120. Khan, I. *et al.* Mechanism of the Gut Microbiota Colonization Resistance and Enteric
887 Pathogen Infection. *Frontiers in Cellular and Infection Microbiology* **11**, (2021).
888 121. Lawley, T. D. & Walker, A. W. Intestinal colonization resistance. *Immunology* **138**, 1–11
889 (2013).
890 122. Ley, R. E., Peterson, D. A. & Gordon, J. I. Ecological and Evolutionary Forces Shaping
891 Microbial Diversity in the Human Intestine. *Cell* **124**, 837–848 (2006).
892 123. Ma, H., Tao, W. & Zhu, S. T lymphocytes in the intestinal mucosa: defense and tolerance.
893 *Cell Mol Immunol* **16**, 216–224 (2019).
894 124. Yip, J. L. K., Balasuriya, G. K., Spencer, S. J. & Hill-Yardin, E. L. The Role of Intestinal
895 Macrophages in Gastrointestinal Homeostasis: Heterogeneity and Implications in Disease.
896 *Cell Mol Gastroenterol Hepatol* **12**, 1701–1718 (2021).

897 125. Pleguezuelos-Manzano, C. *et al.* Mutational signature in colorectal cancer caused by
898 genotoxic pks + E. coli. *Nature* **580**, 269–273 (2020).
899 126. Cuevas-Ramos, G. *et al.* Escherichia coli induces DNA damage in vivo and triggers
900 genomic instability in mammalian cells. *Proceedings of the National Academy of Sciences of*
901 *the United States of America* **107**, 11537–11542 (2010).
902 127. Iftekhar, A. *et al.* Genomic aberrations after short-term exposure to colibactin-producing
903 E. coli transform primary colon epithelial cells. *Nat Commun* **12**, 1003 (2021).
904 128. Tomkovich, S. *et al.* Human colon mucosal biofilms from healthy or colon cancer hosts
905 are carcinogenic. *J Clin Invest* **129**, 1699–1712.
906 129. Yang, Y. *et al.* Integrated microbiome and metabolome analysis reveals a novel interplay
907 between commensal bacteria and metabolites in colorectal cancer. *Theranostics* **9**, 4101–
908 4114 (2019).
909 130. Dai, Z. *et al.* Multi-cohort analysis of colorectal cancer metagenome identified altered
910 bacteria across populations and universal bacterial markers. *Microbiome* **6**, 70–70 (2018).
911 131. Zhao, L. *et al.* Parvimonas micra promotes colorectal tumorigenesis and is associated
912 with prognosis of colorectal cancer patients. *Oncogene* **41**, 4200–4210 (2022).
913 132. Tsoi, H. *et al.* Peptostreptococcus anaerobius Induces Intracellular Cholesterol
914 Biosynthesis in Colon Cells to Induce Proliferation and Causes Dysplasia in Mice.
915 *Gastroenterology* **152**, 1419-1433.e5 (2017).
916 133. Drewes, J. L. *et al.* High-resolution bacterial 16S rRNA gene profile meta-analysis and
917 biofilm status reveal common colorectal cancer consortia. *NPJ biofilms and microbiomes* **3**,
918 34–34 (2017).

919 134. Zijnge, V. *et al.* Oral Biofilm Architecture on Natural Teeth. *PLOS ONE* **5**, e9321 (2010).

920 135. Flemer, B. *et al.* The oral microbiota in colorectal cancer is distinctive and predictive. *Gut*

921 **67**, 1454–1463 (2018).

922 136. Cieplik, F. *et al.* Microcosm biofilms cultured from different oral niches in periodontitis

923 patients. *J Oral Microbiol* **11**, 1551596 (2018).

924 137. Shi, H. *et al.* Highly multiplexed spatial mapping of microbial communities. *Nature* **588**,

925 676–681 (2020).

926 138. Lötstedt, B., Stražar, M., Xavier, R., Regev, A. & Vickovic, S. Spatial host-microbiome

927 sequencing. 2022.07.18.500470 Preprint at https://doi.org/10.1101/2022.07.18.500470

928 (2022).

929 139. Piñero, J. *et al.* DisGeNET: a discovery platform for the dynamical exploration of human

930 diseases and their genes. *Database (Oxford)* **2015**, bav028 (2015).

931 140. Chen, E. Y. *et al.* Enrichr: interactive and collaborative HTML5 gene list enrichment

932 analysis tool. *BMC Bioinformatics* **14**, 128 (2013).

933