



OPEN

The gut microbial metabolite formate exacerbates colorectal cancer progression

Dominik Ternes^{1,13}, Mina Tsenkova^{1,13}, Vitaly Igorevich Pozdeev^{1,13}, Marianne Meyers¹, Eric Koncina¹, Sura Atatri¹, Martine Schmitz¹, Jessica Karta¹, Maryse Schmoetten¹, Almut Heinken^{2,3}, Fabien Rodriguez¹, Catherine Delbrouck⁴, Anhoula Gaigneaux¹, Aurelien Ginolhac¹, Tam Thuy Dan Nguyen⁵, Lea Grandmougin⁶, Audrey Frachet-Bour⁶, Camille Martin-Gallaussiaux⁶, Maria Pacheco¹, Lorie Neuberger-Castillo⁷, Paulo Miranda⁸, Nikolaus Zuegel⁹, Jean-Yves Ferrand¹⁰, Manon Gantenbein¹⁰, Thomas Sauter¹⁰, Daniel Joseph Slade¹⁰, Ines Thiele^{2,3,11,12}, Johannes Meiser⁴, Serge Haan¹, Paul Wilmes^{1,6} and Elisabeth Letellier¹✉

The gut microbiome is a key player in the immunomodulatory and protumorigenic microenvironment during colorectal cancer (CRC), as different gut-derived bacteria can induce tumour growth. However, the crosstalk between the gut microbiome and the host in relation to tumour cell metabolism remains largely unexplored. Here we show that formate, a metabolite produced by the CRC-associated bacterium *Fusobacterium nucleatum*, promotes CRC development. We describe molecular signatures linking CRC phenotypes with *Fusobacterium* abundance. Cocultures of *F. nucleatum* with patient-derived CRC cells display protumorigenic effects, along with a metabolic shift towards increased formate secretion and cancer glutamine metabolism. We further show that microbiome-derived formate drives CRC tumour invasion by triggering AhR signalling, while increasing cancer stemness. Finally, *F. nucleatum* or formate treatment in mice leads to increased tumour incidence or size, and Th17 cell expansion, which can favour proinflammatory profiles. Moving beyond observational studies, we identify formate as a gut-derived oncometabolite that is relevant for CRC progression.

Colorectal cancer (CRC) is one of the most frequently diagnosed types of cancer worldwide, ranking third in terms of tumour incidence and second in cancer mortality¹. The gut microbiome has been associated with molecular mechanisms underlying CRC initiation, progression and treatment. Isolated gut microbes have been shown to induce genetic and microenvironmental changes, ultimately supporting cancerous cell growth. In the past decades, extensive taxonomic and functional profiling of the CRC-associated microbiome has emerged. Along with ^{pks+}*Escherichia coli*, enterotoxigenic *Bacteroides fragilis* and *Streptococcus gallolyticus*, *F. nucleatum* (*Fn*) isolates were identified as a major protumorigenic bacterium. However, in CRC microbiome research, there is an urgent need to move beyond descriptive studies.

Sophisticated study models, such as organoids and gut-on-chip models, can help advance microbiome research in a mechanistic direction. The gut organoid model was recently used to reveal a causative role of a genotoxic ^{pks+}*E. coli* mutational signature in CRC². In vitro modular perfusion bioreactor systems, such as HuMiX, have been used to study host metabolic, transcriptional and immune responses to microbes in CRC^{3,4}. Moreover, genome-scale

metabolic models based on human gut metagenomics data show how microbes can modulate the human metabolism in health and disease, ultimately predicting drug- or dietary-based therapies for CRC^{5,6}. Finally, mouse models remain the golden standard in CRC microbiome research, allowing for the study of tumours, their microenvironment and their immune landscape⁷.

The integral results of published in vitro, in vivo and in silico models strongly support a scenario whereby *Fn* colonizes tumours and promotes CRC growth. Several virulence factors are involved in this process, such as Fap2, FadA, RadD and FomA. Fap2 contributes to tumour colonization and bacterial invasion into Gal-GalNAc expressing epithelial cells⁸. Furthermore, Fap2 binding to host cells increases IL8 and CXCL1 production, driving cancer cell migration⁹ and tumour immune evasion¹⁰. FadA mediates host-cell binding and invasion and promotes cancer cell proliferation¹¹. RadD can cause host-cell death and bridge tumour colonization for other pathogens in biofilms¹². FomA stimulates TLR2-dependent NF-κB signalling in gut epithelial cells¹³. Elaborated secretome analyses on *Fn* are providing further insights on putative and known virulence proteins¹⁴, yet virulence expression is tightly controlled and only subordinate to pathogen metabolism¹⁵. In particular, microbiome-derived

¹Molecular Disease Mechanisms Group, Department of Life Sciences and Medicine, Faculty of Science, Technology and Medicine, University of Luxembourg, Belvaux, Luxembourg. ²School of Medicine, National University of Ireland, Galway, Ireland. ³Ryan Institute, National University of Galway, Galway, Ireland. ⁴Cancer Metabolism Group, Department of Cancer Research, Luxembourg Institute of Health, Luxembourg, Luxembourg. ⁵Department of Biochemistry, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA. ⁶Systems Ecology Group, Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg. ⁷Integrated BioBank of Luxembourg, Dudelange, Luxembourg. ⁸National Center of Pathology, Laboratoire National de Santé, Dudelange, Luxembourg. ⁹Department of Surgery, Centre Hospitalier Emile Mayrisch, Esch-sur-Alzette, Luxembourg. ¹⁰Clinical and Epidemiological Investigation Center, Department of Population Health, Luxembourg Institute of Health, Luxembourg, Luxembourg. ¹¹Discipline of Microbiology, School of Natural Sciences, National University of Ireland, Galway, Ireland. ¹²APC Microbiome, Cork, Ireland. ¹³These authors contributed equally: D. Ternes, M. Tsenkova, V. I. Pozdeev. ✉e-mail: Elisabeth.Letellier@uni.lu

oncometabolites, such as L-2-hydroxyglutarate, succinate, fumarate, D-2-hydroxyglutarate or lactate can accumulate in cancers, thereby fuelling malignancy⁷. Moreover, oncometabolites can hijack meta-static signalling via gene expression regulation¹⁶. In this context, very little is known about the fusobacterial metabolic contribution to a malignant state.

Fusobacteria have previously been shown to migrate to the metastatic site¹⁷, potentially shuttled by disseminated cells via the hematogenous route⁹, demonstrating *Fusobacterium*'s persistence throughout the metastatic process. It has been suggested that metastatic colonization (that is, outgrowth at a secondary site) is restricted to cancer cells with tumour-initiating properties (cancer stem cells, CSCs)¹⁸. CSCs are defined by key functional features including their ability to self-renew and replenish tumour heterogeneity, thereby driving tumour initiation and progression¹⁹. *Fn* was shown to induce CSC characteristics in CRC cells in vitro²⁰, which can be associated with late-stage *Fn* positive cases²¹. Nevertheless, the mechanisms behind *Fn*-associated cancer stemness and invasion remain unclear.

In this study, we aimed to integrate in vitro, in vivo and in silico approaches to gain an ecosystem-level mechanistic understanding of *Fn*'s metabolic role in CRC pathogenesis. We propose that *Fn* not only directly interacts with and contributes to CRC cell malignancy, but that its microbe–host metabolic crosstalk is involved in the progression of the disease. More specifically, we identified the small-molecule formate as an oncometabolite that contributes to CRC stemness, invasion and metastasis.

Results

***Fn* is elevated in patients with CRC and associated with consensus molecular subtype 3 (CMS3).** To investigate the CRC microbiome in an in-house cohort (Supplementary Table 1), we analysed the bacterial compositions of stool samples from patients with CRC ($n=63$) and compared them with the stool microbiomes of a healthy cohort ($n=52$) using 16S ribosomal RNA gene sequencing. The differential abundance analysis (Fig. 1a) identified bacterial genera that were significantly depleted in CRC donor stools, such as the potential probiotics Ruminococcaceae and Lachnospiraceae, and genera that were significantly enriched, such as the well-described *Fusobacterium*, *Porphyromonas*, *Pseudomonas*, *Streptococcus* and *Gemella* (marked in red).

We used PathSeq, a valuable tool for the analysis of metagenomic data²², to identify CRC-tissue associated bacterial genera and species in datasets from The Cancer Genome Atlas (TCGA) ($n=595$) and the European Genome-Phenome Archive (EGA) ($n=69$). The analysis of the tissue microbiome identified bacteria that were associated with a cancerous tissue state (Fig. 1b). Among these bacteria were *Fusobacterium*, *Parvimonas*, *Gemella* and *Streptococcus* (marked in red). The *Bacteroides* genus was depleted in tumour

tissue across all stages of the disease and elevated abundances of *Fusobacterium* were observed in stage II CRC, *Campylobacter* in stages II and IV, and *Gemella* in stages I, II and IV (Extended Data Fig. 1a). Of these bacteria, *Fn* was also enriched in tumour tissues in the EGA cohort, when compared with matching adjacent tissue (Extended Data Fig. 1b, $P<0.02$), which has also been previously observed in independent datasets^{23,24}.

TCGA and the EGA databases contain valuable gene expression data from patient tissue, allowing us to link patient transcriptomic and microbial profiles (metadata for these datasets can be found in Supplementary Tables 2 and 3). We were particularly interested in associating higher levels of *Fusobacteria* in patients to the CMS classification of CRC. This classification uses the transcriptomic profile of tumours to group them into CMS1 (hypermethylated and immunogenic), CMS2 (characterized by the activation of the WNT and MYC pathways), CMS3 (metabolically deregulated) and CMS4 (strong stromal infiltration associated to lower relapse-free and overall survival expectancies)²⁵. To this end, we classified patients into no (score, 0), low (score, $\leq t$, where t is the median value of positive scores) and high (score, $> t$) fusobacterial abundance (Fig. 1c). The results show that CMS1 and CMS3 patients have a similarly high proportion of *Fusobacterium* ($\chi^2(2, n=595)=1.3$, $P>0.05$), which was significantly different from the proportions observed in CMS2 and CMS4 patients ($P<0.05$) (Fig. 1d). This observation is also in agreement with very recent work from Salvucci and collaborators²⁶. In addition, CMS3 tended to harbour high proportions of *Fusobacterium*^{high} cases in the EGA dataset (Extended Data Fig. 1c,d). We therefore suggest that *Fusobacterium* not only plays a role in CMS1 (ref. 24), but also in CMS3, which is enriched in metabolic pathways including glutamine, fatty acid and lysophospholipid metabolism²⁷. Next, we performed differential gene expression analyses comparing *Fusobacterium*^{high} to *Fusobacterium*^{no} cases. Our pathway enrichment analyses showed that inflammatory-related signalling pathways, such as IL-8 signalling or Th17 activation, were upregulated (Extended Data Fig. 1e). In addition, pathways related to cellular organization, movement and invasion, as well as metabolic pathways involving cholesterol and proteoglycan metabolism, were affected (Extended Data Fig. 1e,f). Notably, aryl hydrocarbon receptor (AhR) signalling was found activated in both the TCGA and the EGA datasets when *Fusobacteria* were abundant (Extended Data Fig. 1e,g).

***Fn* induces protumorigenic signalling and formate secretion.**

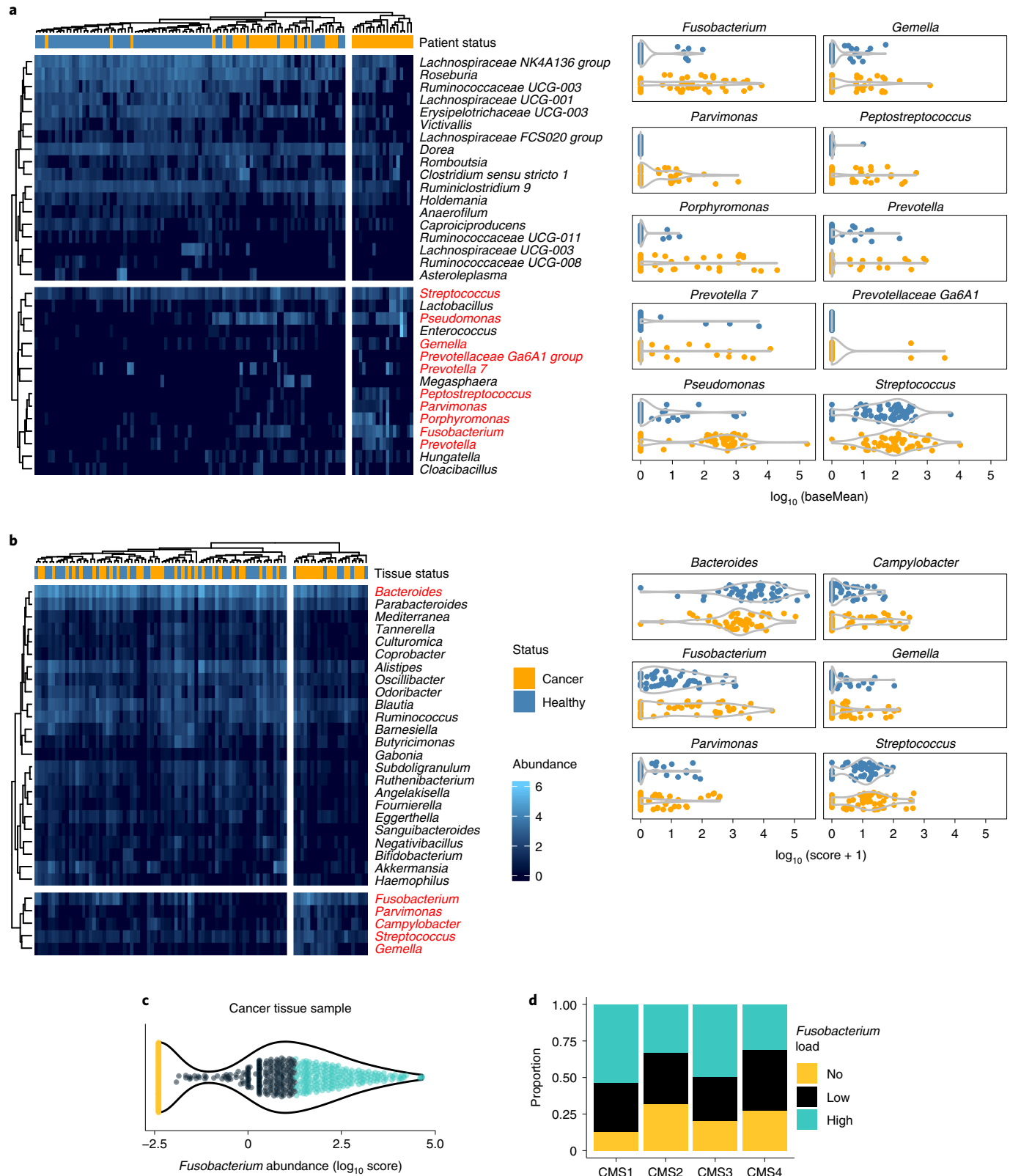
To specifically address how *Fn* influences tumour cell metabolism, we used the modular, perfusion bioreactor system HuMiX. The model previously helped us to study metabolic, transcriptional, and immune responses of CRC cells in coculture with the probiotic *Lactobacilli*⁴. HuMiX represents a unique system to study host-microbiome metabolic crosstalk where the exchanged

Fig. 1 | *Fusobacterium* levels are elevated in stool and tissue from patients with CRC and are associated with CMS1 and metabolic-driven CMS3.

a, Differential abundance of bacterial genera in CRC stool samples. Clustered heatmap (Pearson correlation) shows the significant differentially abundant bacterial genera in an in-house cohort of samples from patients with CRC ($n=52$ independent biological samples) in comparison with healthy donor ($n=63$ independent biological samples) ($P<0.05$; negative binomial Wald Benjamini–Hochberg testing) as identified by 16S rRNA gene sequencing. Top CRC-enriched bacteria⁷ are marked in red and separately plotted as abundance (baseMean) of a bacterium per patient (right). Heatmap intensities (blue colour scale) represent \log_{10} base mean values. **b**, Differential abundance of bacterial genera in CRC-tissue samples, following a PathSeq analysis on whole-exome sequencing (WXS) data (TCGA). Clustered heatmap (Pearson correlation) showing the significant differentially abundant bacterial genera in tumour tissue samples from patients with CRC ($n=50$) in comparison with adjacent healthy mucosal samples ($n=50$, $P<0.05$; negative binomial Wald Benjamini–Hochberg testing). Top CRC-enriched bacteria are marked in red and separately plotted as abundance (score) of a bacterium per sample (right). Heatmap intensities (blue colour scale) represent \log_{10} scores. **c**, *Fusobacterium* abundance in tissue samples from patients with CRC ($n=595$ independent biological samples, TCGA). WXS data were analysed for the logarithmic score abundance of *Fusobacterium* across all samples via quantile-based classification (colour code). **d**, *Fusobacterium* abundance across CMS subtypes. The cohort (**c**) was subjected to gene expression analysis and CMS classification of matching RNA-seq⁶⁷. Coloured, segregated bars show the proportion of patients with fusobacterial loads per CMS. Significant differences were observed for CMS1 versus CMS2 ($P=0.003103733$) or CMS4 ($P=0.006062911$) and CMS3 versus CMS2 ($P=0.038151361$) or CMS4 ($P=0.035902419$) Chi-squared tests. $n^{\text{CMS1}}=62$, $n^{\text{CMS2}}=207$, $n^{\text{CMS3}}=69$, $n^{\text{CMS4}}=139$ biologically independent samples.

metabolites can be independently analysed in each chamber. To test the effects of host–*Fusobacterium* crosstalk, we cocultured the patient-derived, CRC cell line HT-29 with *Fn* ATCC 25586 in HuMiX (Fig. 2a and Extended Data Fig. 2a,b). HT-29 cells were obtained from ATCC and were derived from a 44 year old female patient with grade II adenocarcinoma and selected on the basis of

the stage II enrichment of *Fn* in patients with CRC (Extended Data Fig. 1a). The HuMiX system allowed us to study the effects of *Fn* on cancer cells at near-physiological oxygen levels (Extended Data Fig. 2c) and under a constant supply of fresh medium to the basal cell surface. The 24-hour coculture had no effect on cell viability, nor on *Fn* and HT-29 proliferation (Extended Data Fig. 2d,e).



The transcriptional response of HT-29 cells to *Fn* exposure showed effects on Wnt, IL8 and mitogen-activated protein kinase (MAPK) downstream pathway activation (Fig. 2b), traits that have been linked with *Fusobacterium* in CRC in the past^{11,28}. Further activated pathways included Th17 activation, AhR receptor signalling and CRC metastasis signalling. Moreover, when investigating potential functional effects, the ingenuity pathway analysis (IPA) predicted high activation scores for cancer cell morphology and invasion (Fig. 2c). To confirm this functional regulation, we looked at the activation of the MAPK signalling pathway, which is important for multiple biological processes (for example, cell proliferation, differentiation, migration and invasion). As typical responders to extracellular stimulation, the phosphorylation of extracellular signal-regulated kinase (ERK) and slight activation of MEK and p38 were observed on the molecular exchange between *Fn* and HT-29 (Extended Data Fig. 2f,g). In addition, phosphorylation of the p65 (NF- κ B subunit) tended to be increased due to *Fusobacterium* (Extended Data Fig. 2f,g).

The global transcriptional changes observed in the HuMiX experiments, including the activation of Th17 and AhR receptor signalling, but also the induction of cancer stemness genes (for example SOX2), correlated with those observed in tissues from *Fusobacterium*^{high} patients (Extended Data Fig. 2h,i, see also Fig. 2b in comparison with Extended Data Fig. 1e,g). To address the soluble factors underlying the observed effects, we performed untargeted and targeted metabolite analyses. The results of the untargeted analysis showed that HT-29 cell culture chambers contained lower glutamic acid and glutamine levels, suggesting a boosted consumption (Fig. 2d and Extended Data Fig. 3a,b). In addition, lower levels of lactic acid in the HT-29 chambers compared to the monoculture suggest a metabolic switch from glucose to glutamine as the main carbon source. Regarding fusobacterial metabolism, large metabolite level increases were observed for succinic acid, lactic acid and alanine (Extended Data Fig. 3c). A range of amino acids, such as isoleucine, lysine, methionine, phenylalanine, tyrosine and valine were reduced in the coculture bacterial chambers compared to the respective monoculture chambers and the inflow metabolite profile, suggesting fusobacterial use (Extended Data Fig. 3b,c). The results from the targeted gas chromatography–mass spectrometry (GC–MS) analyses on exometabolites from the HuMiX chambers indicated that upon coculture, *Fn* secreted almost four times more acetate, propionate, butyrate and formate; the latter being the predominant secretion product (Fig. 2e). To verify that the increase in formate and other metabolites is common to different *Fn* strains when interacting with different CRC cells, we performed further HuMiX-based experiments focused on the metabolic crosstalk using different *Fn* strain isolates (*Fn* 23726 and one clinical isolate *Fn* ssp. *animalis* 7_1). Analysis of the metabolic profile of the different strains in coculture with HT-29 or Caco-2 (human colon epithelial cell line) clearly showed that formate is a common *Fn*-derived metabolite that is secreted upon coculture between different human CRC-derived cells and *Fn* strains (Fig. 2f and Extended Data Fig. 3d). The clinical isolate showed the highest secretion of formate compared to ATCC 23726 and 25586 *Fn* strains, further supporting the

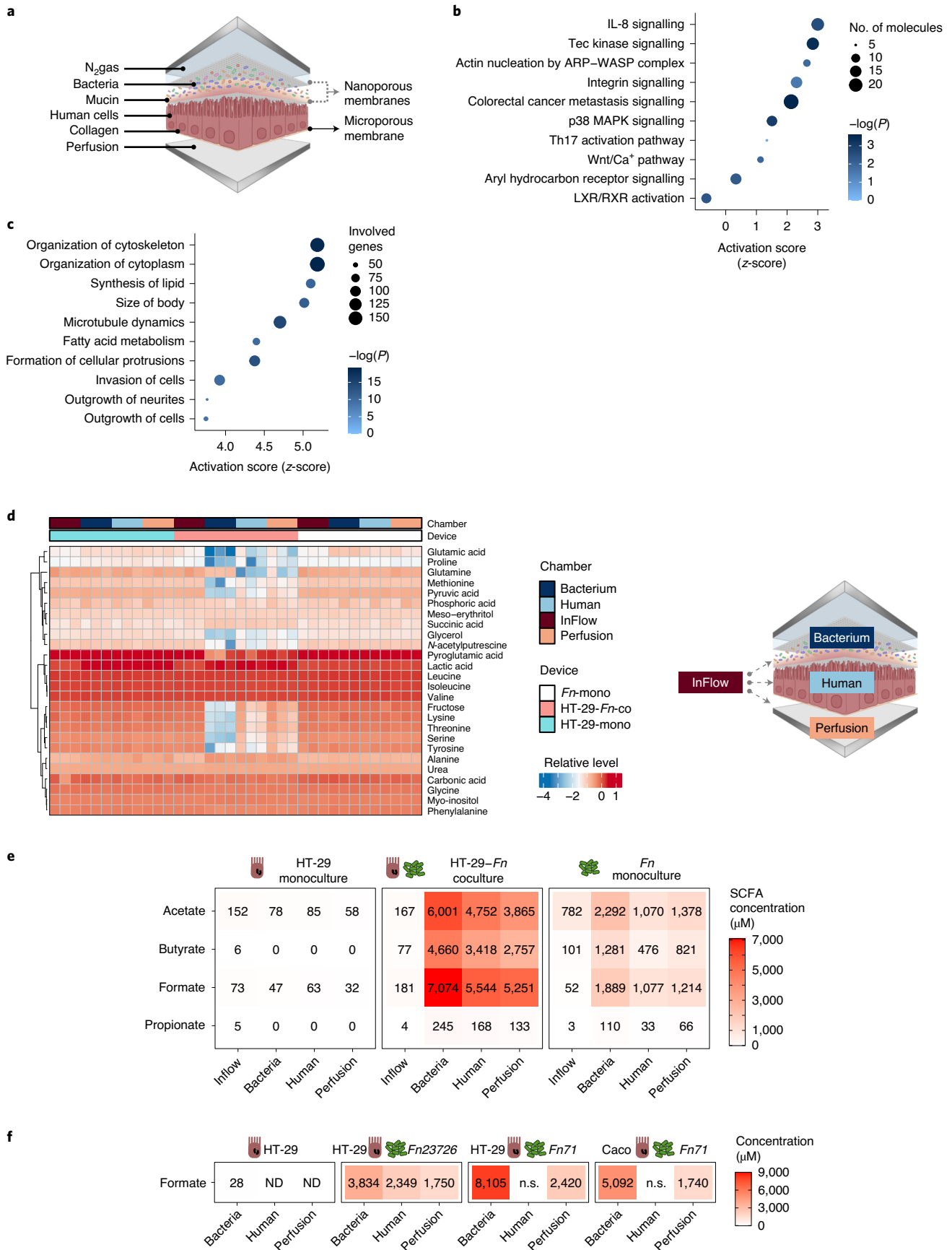
relevance of formate secretion in human disease. We then focused on the most interesting metabolites in the untargeted metabolomics analysis identified in Fig. 2d and Extended Data Fig. 3. Most of the altered metabolites including glutamine, lysine, pyruvate, threonine and serine, were reduced in HT-29 cells after coculture with *Fn* (Fig. 2d), independently of the strain used (Extended Data Fig. 3e). In addition, we observed an increase in genes related to the AhR signalling pathway and cancer stemness in CRC cells upon coculture with the various *Fn* strains described above (Extended Data Fig. 3f). Altogether, these data indicate that our findings are generalizable among CRC cells and *Fn* strains.

Fn induces a central carbon metabolism shift in tumour cells.

To study the microbe–host crosstalk further, we used the *Fn* ATCC 25586 genome-scale reconstruction model from the AGORA resource²⁹ as a basis for model curation and host–microbe pairwise interaction simulations. A Recon 2-based conditioned specific model of HT-29, grown in HuMiX, served as a host model⁶. The simulations of *Fn*–HT-29 cometabolism via flux variability analysis (FVA) predicted an increased production of formate when compared to the monoculture model. Notably, formate secretion to the extracellular space showed the biggest increase in maximum secretion flux, followed by aspartic acid and butanol (reflected in Fig. 3a, left cell, also Extended Data Fig. 4a). Acetate, succinate and glutamic acid were further metabolites, whose maximum secretion increased for *Fn*. These metabolites could serve as a carbon and nitrogen source, feeding into the cancer cells' tricarboxylic acid (TCA) cycle. A predicted lower net production of lactate in HT-29 cells, together with a reduced gene expression of glycolysis enzymes and lower levels of lactate detected in the HT-29 HuMiX chamber (compared to its monoculture control), indicate a lower glycolytic activity of the CRC cells in coculture (reflected in Fig. 3a, right cell, also Extended Data Fig. 4b). In addition, the main glucose transporter GLUT1 was transcriptionally downregulated and the *in silico* model even allowed for an increased maximum glucose export flux, an effect potentially compensated by increased fluxes in glutamic acid and glutamine metabolism (Extended Data Fig. 4c). As indicated by RNA-sequencing (RNA-seq) and metabolomics data, TCA cycle enzyme activities were increased and pools of amino acids were reduced in HT-29 cell HuMiX chambers (Extended Data Fig. 4d,e). The *in silico* model supported this effect by predicting increased uptake rates for amino acids such as glutamine, glutamic acid, cysteine, aspartic acid, alanine and glycine, as the pools of these amino acids were depleted in HuMiX cocultures. Notably, *Fn* could provide an alanine source for HT-29 cells. Moreover, an additional pool of amino acids could be accessed by the HT-29 cells in coculture due to the intra and extracellular proteolytic activity of *Fusobacterium*³⁰. Thus, serum protein breakdown could render additional nitrogen sources accessible to cancer cells.

To verify the patient relevance of our predicted and measured metabolite changes, we reanalysed metagenomic and metabolomic data from a Japanese cohort (Yachida; metadata for this cohort can be found in Supplementary Table 4)³¹. We compared how well the

Fig. 2 | Gut-on-chip cocultures (HuMiX) of *Fn* with patient-derived primary CRC cells reveal protumorigenic and pro-invasive effects along with an altered metabolic profile. **a**, Schematic representation of the HuMiX setup. **b**, IPA canonical pathway analysis of HuMiX HT-29 co- versus monoculture differential gene expression analysis. Plot shows z-scores, $-\log(P)$ values and the number of molecules per pathway. Selected significant pathways are shown ($P < 0.05$), Fisher's Exact Test. **c**, IPA Diseases or Functions analysis of HuMiX HT-29 co- versus monoculture differential gene expression analysis. Plot shows z-scores, P values, and the number of genes per pathway. Top ten significant pathways are shown ($P < 0.05$): Fisher's exact test. **d**, Exometabolite profile of HuMiX cocultures. Untargeted GC–MS metabolite detection in chamber supernatants was used to measure relative, normalized metabolite levels. Heatmap shows relative log intensities of extracellular metabolites (rows) across $n = 3$ independent experiments, each with one device per condition (main columns) composed of four chambers (subcolumns). **e**, SCFA levels in the HuMiX after HT-29–*Fn* coculture. Heatmap shows mean concentrations of extracellular SCFAs (rows) from $n = 3$ independent experiments, each with one HuMiX device per condition, composed of four chambers (columns). **f**, Levels of formate in the HuMiX chambers using different *Fn* isolates (*Fn* 23726 and one clinical isolate *Fn* ssp. *animalis* 7_1) in coculture with HT-29 or Caco-2 cells ($n = 1$ experiment with one HuMiX device per condition). ND, not detected and n.s., not sampled.



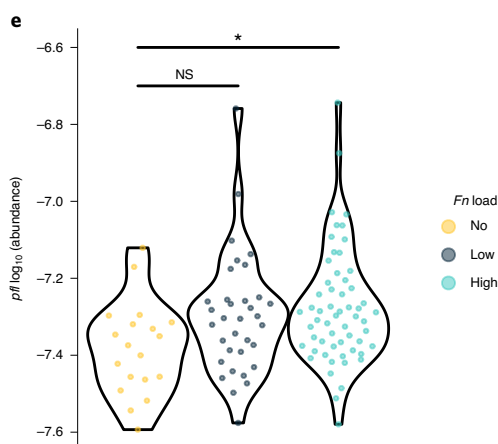
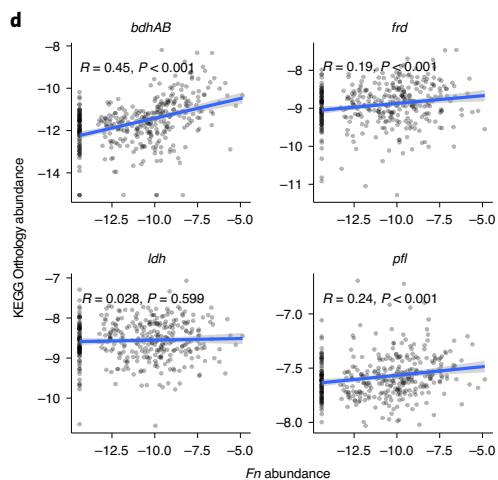
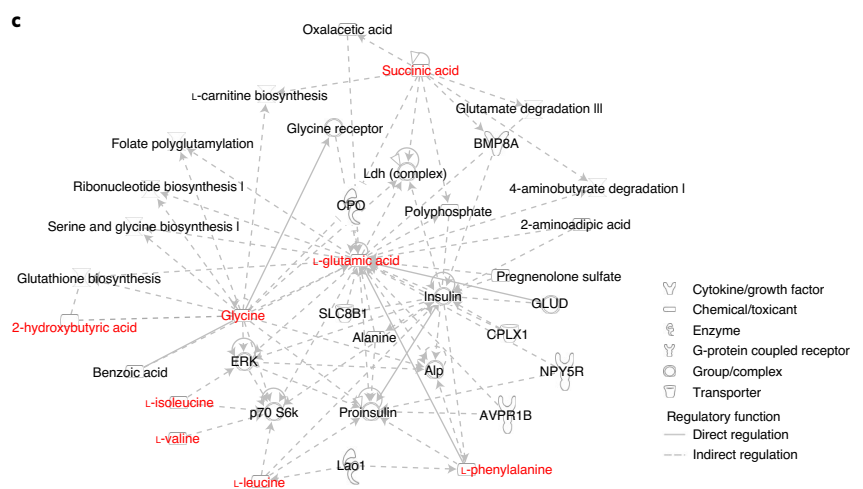
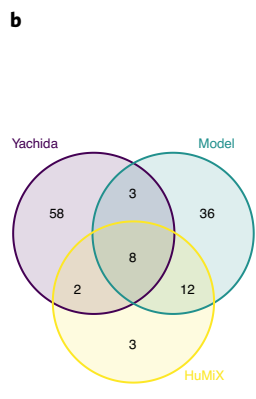
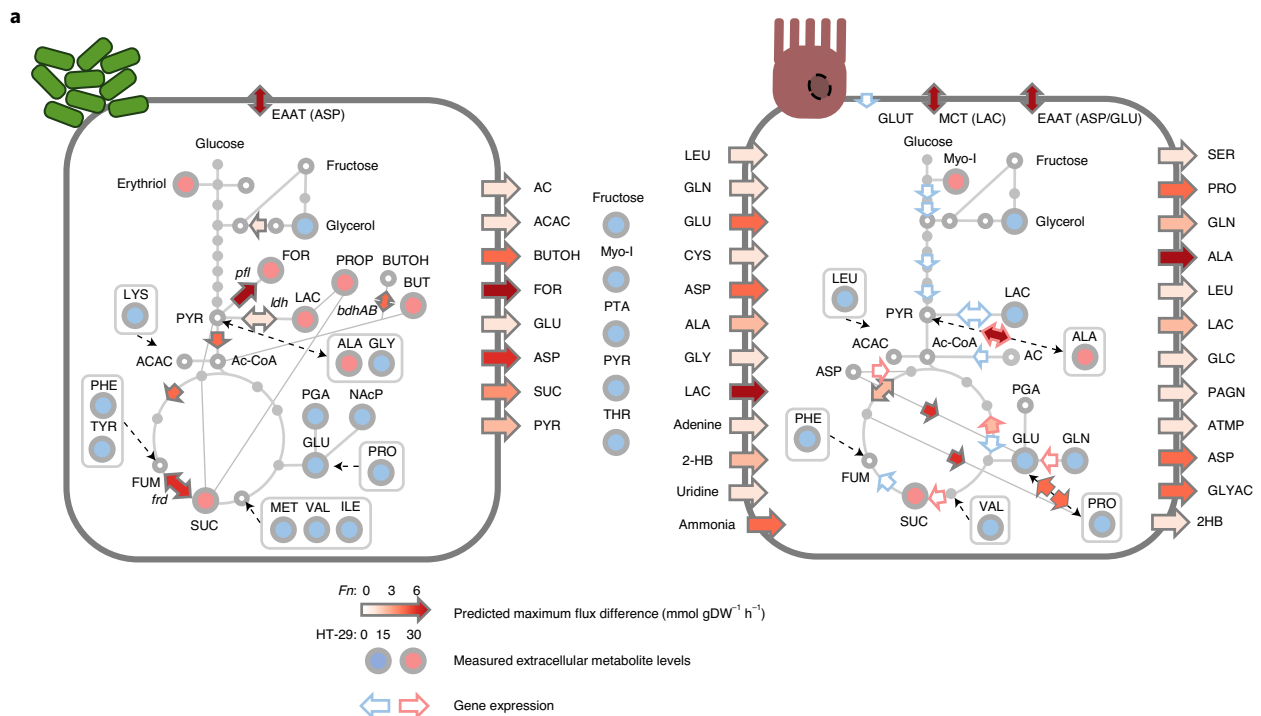


Fig. 3 | In silico modelling delineates enhanced formate metabolism of *Fn* in CRC context, which is also observed in *Fusobacterium*-high patients with CRC. **a**, Metabolic models of the central carbon metabolism of *Fn* (left) and HT-29 cells (right) in coculture versus monocultures. Biomass production was 0.20476 for *Fn* and 0.01 for HT-29. Flux constraints were set according to the metabolite secretion profiles of the HuMiX monoculture controls and pairwise interactions were calculated via FVA. Heatmap arrows show maximum flux changes (mmol dGW⁻¹ h⁻¹) of depicted reactions as predicted by FVA. Arrows at the cells' boundaries represent intra-extracellular exchange reactions. Measured differentially abundant exometabolites in the respective HuMiX coculture chambers are shown as filled circles (blue, decreased; red, increased). Extracellular filled circles show unassigned metabolites. Coloured arrow borders show upregulated (red) or downregulated (blue) genes involved in the respective reaction fluxes, on the basis of RNA-seq data (refer to Fig. 2b). Dashed arrows show indirect connections between nodes. 2HB, 2-hydroxybutyrate; AC, acetate; ACAC, acetoacetate; ATPMP, adenylylthiomethyl-pentose; BUTOH, butanol; EAAT, excitatory amino acid transporters; GLYAC, glyceric acid; LAC, lactate; MCT, monocarboxylate transporters; NAcP, *n*-acetylputrescine; PAGN, phenylacetylglutamine; PGA, pyrroglutamic acid; and *bdhAB*, butanol dehydrogenase. **b**, Core set of *Fn*-related metabolites. Venn diagram shows overlapping, differentially abundant metabolites in the Yachida dataset (purple), in HuMiX (yellow) and in the metabolic model (green). **c**, IPA network analysis of the gene-regulatory role of *Fn* core metabolites (red) in connection with host gene-regulatory nodes of stemness and invasion. **d**, KEGG orthology gene abundances of *Fn*-related genes identified by the model in patients with CRC with different fusobacterial load in stool samples. Two-tailed Spearman's rho correlation testing was used. **e**, *pfl* in stool metagenomes of stage I/II, *Fn*^{high} patients with CRC. $n^{Fn-no} = 19$, $n^{Fn-low} = 34$, $n^{Fn-high} = 58$ (NS, not significant, $P = 0.0219$ for *Fn*-no versus *Fn*-high; one-way analysis of variance (ANOVA) with Tukey's honestly significant differences test). * $P < 0.05$.

in silico model predicted the metabolic changes observed in the HuMiX experiments and in patients with high fusobacterial levels. A Venn diagram shows the overlapping differentially abundant metabolites from the three datasets (Fig. 3b). The constraint-based modelling approach covered 80% (20/25) of the differentially abundant metabolites in the HuMiX cocultures. An overlap of eight metabolites was shared among all three datasets: glutamic acid, 2-hydroxybutyrate, glycine, isoleucine, leucine, phenylalanine, succinic acid and valine. Of note, formate was left out in the analyte library of the Yachida dataset and is therefore not part of the list of the metabolites that are common between the three datasets. The IPA analysis for the downstream roles of these metabolites suggested a highly connected network of gene-regulatory nodes and metabolites involved in controlling stem cells and cellular invasion, as defined by IPA Disease and Function (Fig. 3c).

On the basis of the relative abundance of Kyoto Encyclopedia of Genes and Genomes (KEGG) orthology genes in patient stool metagenomes of the Yachida dataset, we checked for a correlation between previously identified bacterial enzymes (Extended Data Fig. 4d,e) and *Fn* abundance (Fig. 3d). Patients with higher *Fn* loads exhibited higher levels of butanol dehydrogenase (*bdhAB*), fumarate reductase (*frd*) and pyruvate-formate lyase (*pfl*) ($\rho > 1$, $P < 0.05$). Lactate dehydrogenase (*ldh*) did not correlate with fusobacterial abundance. These results indicate that the genes for *Fn*-related metabolic interactions with tumour cells and especially formate production were increasingly present in patients with higher *Fn* levels.

Accordingly, the KEGG orthology gene for microbial formate production *pfl* was significantly higher in stage I/II CRC patients who had high fusobacterial presence (Fig. 3e).

These results demonstrate that in coculture, *Fusobacterium* metabolism is altered towards the higher production of formate and the secretion of metabolites that potentially affect the function and regulation of malignant cells in CRC. At the same time, HT-29 cell central carbon metabolism relied more on TCA cycle linked metabolites. While model predictions such as increased maximum secretion flux for glucose or glutamine require further investigation, the results showed that the applied constraints were able to predict a range of metabolic processes that were subsequently observed in the experimental data from the HuMiX cocultures. In addition, a predicted core set of metabolites and formate may play a role in fusobacterial-host interactions, not only in vitro cocultures, but also in CRC patients. In our next experiments, we continued to use the well-described CRC-associated bacterium *Fn* as a model to study the mechanisms behind microbe-host crosstalk in the context of CRC.

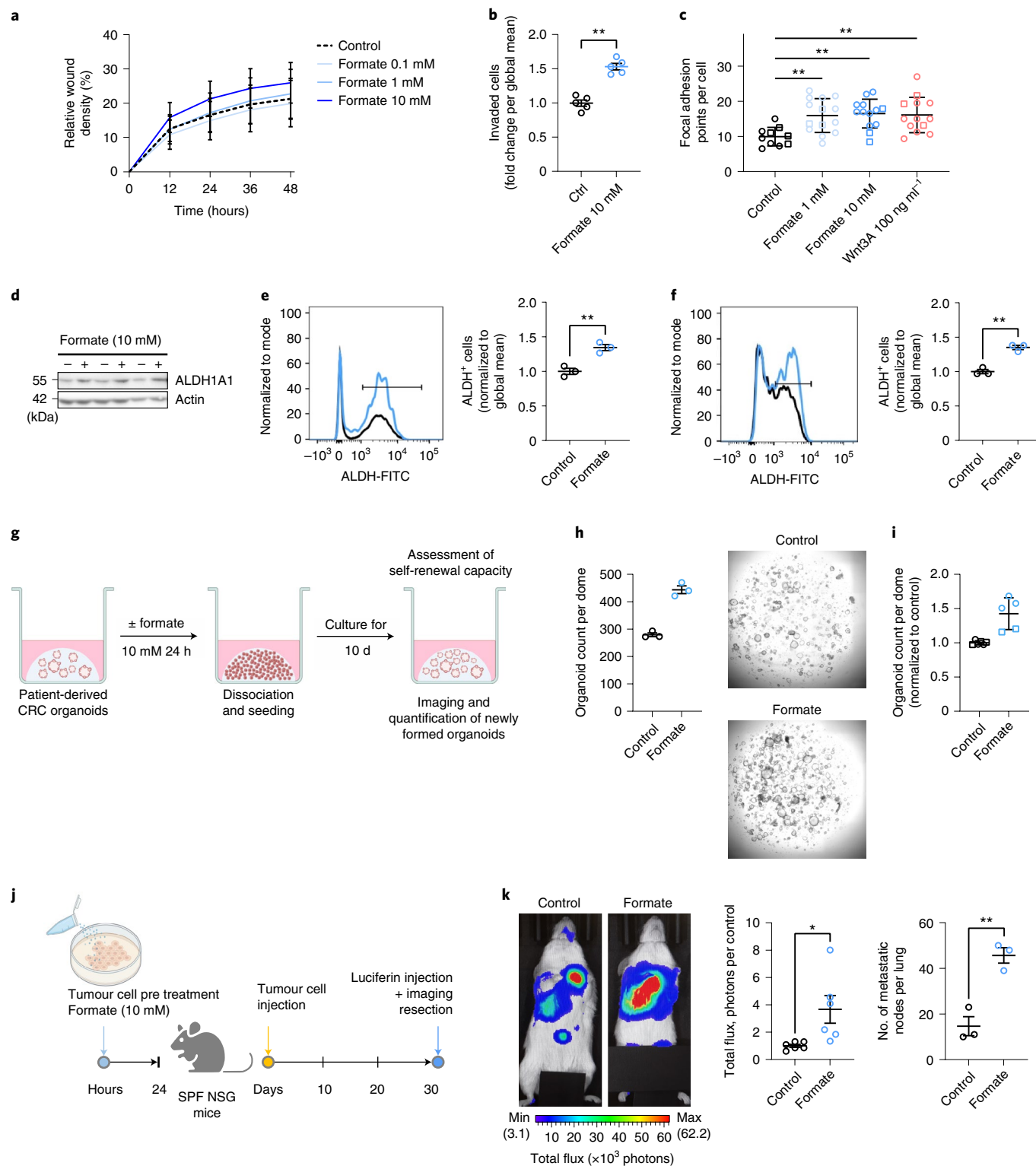
Formate increases cancer stemness, driving CRC cell invasion. In our HuMiX experiments, microbiome–host crosstalk was found to induce different signalling pathways associated with cell invasion. Formate was identified as a common and main *Fn* fermentation product (Fig. 2e,f). Since formate is also common to other bacteria³² and since endogenous formate was previously shown to affect the

Fig. 4 | Formate drives CRC cell invasion by increasing cancer stemness. **a**, Scratch wound healing capacity of HCT116 CRC cells over 48 h, $n = 3$ biologically independent experiments. Gut formate physiological dose is considered to be near 10mM. **b**, Transwell invasion of HCT116 CRC cells at 48 h. Data show the means of replicates from $n = 5$ independent experiments, $P = 0.0079$, two-tailed Mann-Whitney test. **c**, Focal adhesion formation in HCT116 CRC cells after formate or rhWnt3A exposure for 24 h. Data show means of technical replicates with at least 100 cells per condition from two independent experiments (indicated by different shapes). $P = 0.0150$, $P = 0.0037$ and $P = 0.0067$ for formate 1mM, formate 10mM and rhWnt3A versus control, respectively, ordinary two-way ANOVA with Tukey's multiple comparisons test. **d**, ALDH1A1 protein levels in formate (+) treated versus untreated (–) HT-29 cells. **e,f**, ALDH activity in human (**e**) or APCmin mouse colonic organoids (**f**) after formate stimulation (5 mM, 24 h; FACS). Left, representative histogram of the ALDH⁺ cell population (crossbar). Right, quantification of ALDH activity, $n = 3$ independent experiments. $P = 0.0053$ in **e** and $P = 0.0129$ in **f**, unpaired two-tailed *t*-test. FITC, fluorescein isothiocyanate. **g**, Schematic representation of experimental setup. Derived control or formate-treated organoids from a CRC patient were dissociated and reseeded at 24 h. **h**, The organoid formation capacity on day 10 after reseeding. Left, the number of organoids counted in one representative experiment out of two, with three technical replicates, Right, representative images of one well per condition, **i**, Normalized organoid formation, $n = 2$ biologically independent experiments (different data point shapes) with two or three technical replicates per condition. **j**, Schematic overview of the intravenous metastatic dissemination model. HT-29-Luc tumour cells were treated with formate before intravenous injection into NSG mice (1×10^6 cells per 200 μ l per injection). **k**, IVIS imaging after luciferin injection on day 30 after tumour cell injection. Left, representative images from one mouse per group. Middle, the quantified tumour cell signal from lungs. Data show the total luciferase signal (reported as photon flux per mg of tissue), normalized to control, $n = 6$ biologically independent animals per condition, pooled from two independent experiments. Right, the number of lung macroscopic metastatic nodes, $n = 3$ biologically independent animals per condition from one representative experiment out of two. $P = 0.0257$ and $P = 0.0045$ in the middle and right panels, respectively, unpaired two-tailed *t*-test. **b**, **c**, **e**, **f** and **k** are shown as mean \pm s.e.m. * $P < 0.05$, ** $P < 0.01$.

invasion of glioblastoma cells *in vitro*³³, we proposed that exogenous microbiome-secreted formate could also affect colon cancer cell invasion and ultimately metastatic dissemination. Scratch invasion and transwell assays showed that formate increased the invasion of CRC cells in both experimental setups (Fig. 4a,b). This was accompanied by the formation of a higher number of focal adhesion points per cell (Fig. 4c and Extended Data Fig. 5a). Of the other *Fn*-derived metabolites, only acetate showed an effect on cancer cell invasion, while propionate, succinate and alanine did not (Extended Data

Fig. 5b, ref. ³⁴). Lactate was used as a positive control for invasion rather than an *Fn*-derived metabolite. Indeed, lactate levels tended to decrease in the HT-29 HuMiX chamber after coculture (Extended Data Fig. 3a).

Tumour metastasis has long been associated with CSCs. Indeed, several studies^{35,36}, including one of our own³⁷ and a recent one using time-lapse intravital microscopy following cells during dissemination³⁸, showed that only cells with CSC activity are able to outgrow into macroscopic metastases. In addition, we observed an increase



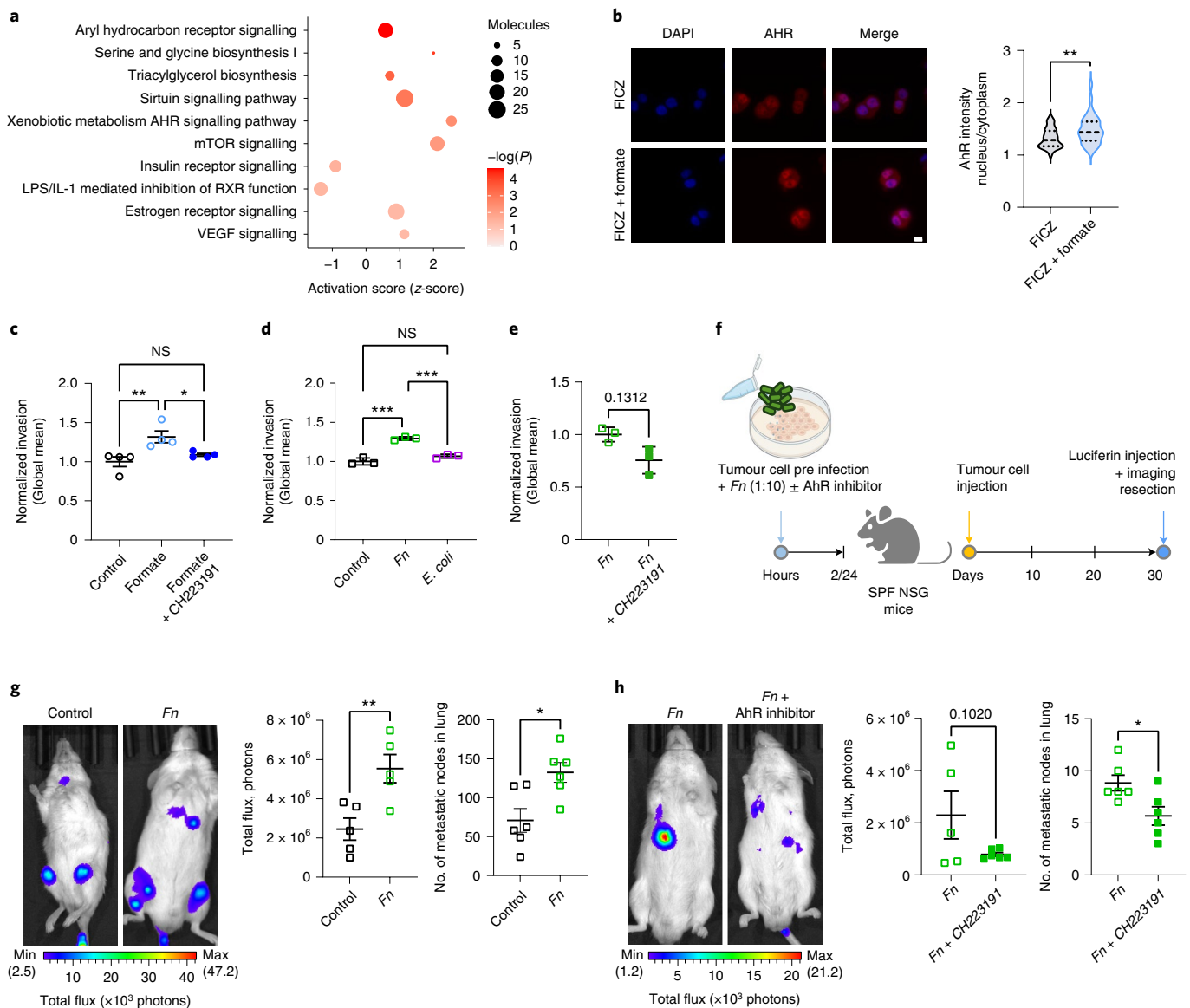


Fig. 5 | Microbiome-derived formate drives metastatic dissemination through the activation of the AhR signalling pathway. **a**, IPA analysis of differentially expressed genes in formate (10 mM) versus PBS treated HT-29 cells. Plot shows z-scores, P values and the number of molecules per selected significant pathway ($-\log_{10}(P) > 1.3$), Fisher's Exact Test. **b**, AhR nuclear translocation in HT-29 cells treated with FICZ (known AhR ligand, 100 nM) alone or combination with formate (10 mM) for 6 h. Data shows technical replicates from two independent experiments, $P = 0.00113$, two-tailed nested ANOVA. Dashed lines represent the medians and dotted lines represent the upper and lower quartiles. **c**, Transwell invasion of HCT116 CRC cells after formate stimulation (10 mM, 48 h) alone, or in presence of an AhR signalling inhibitor (CH223191, 0.5 μ M); $P = 0.0417$ for formate versus control; ordinary one-way ANOVA. **d**, Transwell invasion of HCT116 CRC cells after *Fn* and *Escherichia coli* coculture (MOI 10, 2 h). $***P < 0.001$; ordinary one-way ANOVA. **e**, Transwell invasion of HCT116 cells upon *Fn* preinfection alone or in presence of an AhR signalling inhibitor (CH223191, 0.5 μ M) for 24 h: paired two-sided t -test. **c–e**, Data show pooled means of replicates from three (**d** and **e**) or four (**c**) independent experiments. **f**, Schematic representation of experimental setup of **g** and **h**: NSG mice were intravenously injected with 1×10^6 cells per 200 μ l (**g**) or 0.5×10^6 cells per 200 μ l (**h**) HT-29-Luc cells, preinfected with *Fn* (MOI 10) for 2 h and pretreated with (**h**) or without (**g**) AhR inhibitor for 24 h (CH223191, 0.5 μ M). After 30 d, signals of HT-29-Luc cells were determined using the IVIS. **g,h**, Left, representative image of one mouse per group. Middle, data show the total luciferase signal (reported as photon flux per mg of tissue) normalized to control. Right, the number of lung macroscopic metastatic nodes, $n = 6$ biologically independent animals per condition (one mouse failed luciferin intraperitoneal injection). Regions of interest in **g**: control, $1\text{--}11.001 \times 10^6$ and *Fn*, $2\text{--}3.379 \times 10^6$. $P = 0.0094$, $P = 0.0110$ and $P = 0.0210$ in **g** (middle and right) and **h** (right), respectively, unpaired two-tailed t -test. Total flux was measured in the chest area. Data in **c–e**, **g** and **h** are represented as mean \pm s.e.m. $*P < 0.05$, $**P < 0.01$, $***P < 0.001$.

in genes related to cancer stemness upon coculture with various *Fn* strains (Extended Data Fig. 3f). Therefore, to understand whether the formate-induced invasion of cancer cells could be explained by the acquisition of CSC traits, we first looked for conventional CSC markers such as aldehyde dehydrogenase 1 (ALDH1)³⁹. Accordingly,

ALDH1A1 protein levels were higher in formate-treated HT-29 cells (Fig. 4d). ALDH2 levels were also slightly induced (Extended Data Fig. 5c). Next, we took advantage of the more physiological colonic organoid model. Treatment with formate resulted in increased ALDH activity in both human organoids (Fig. 4e) and

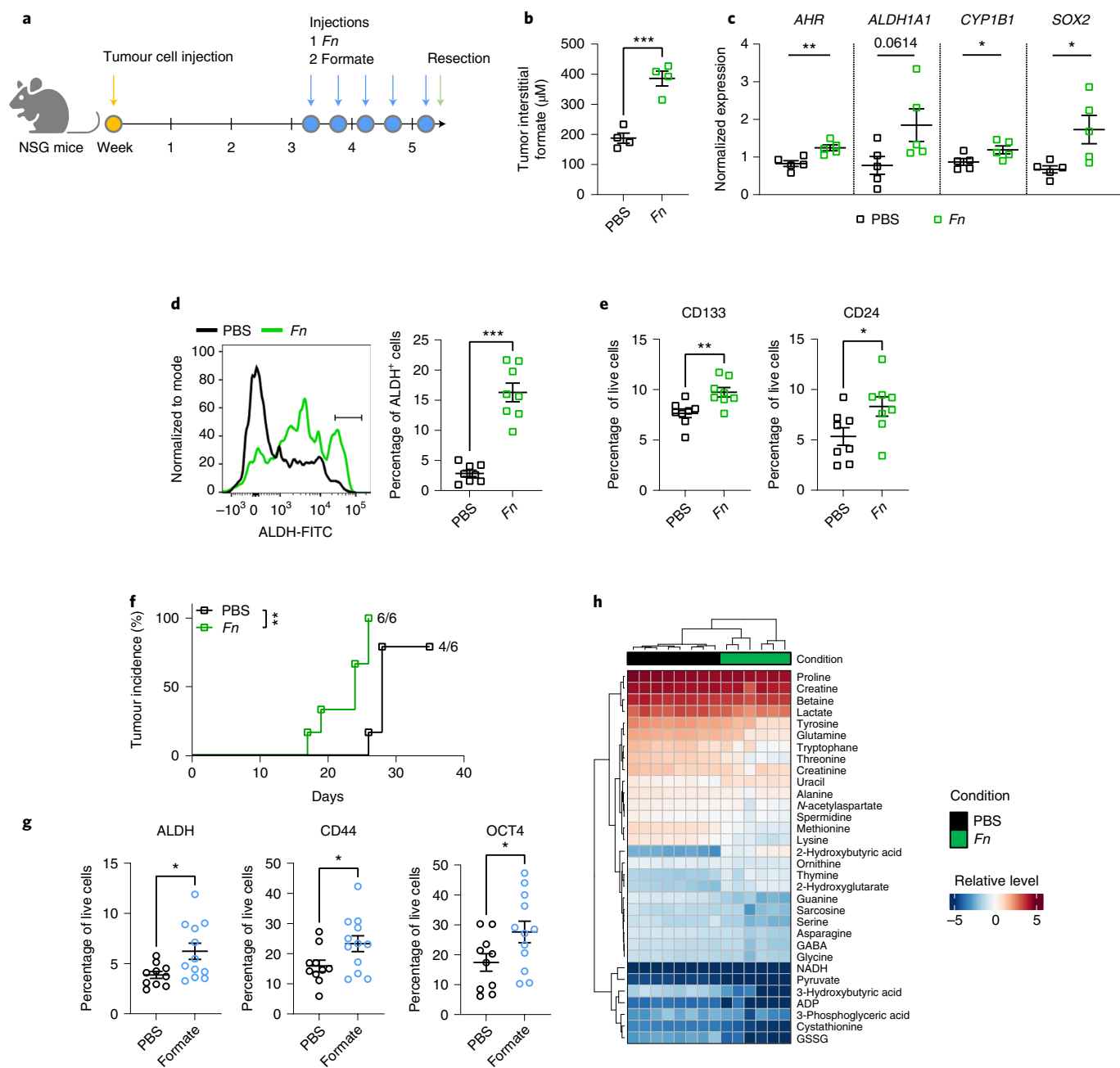


Fig. 6 | *Fn* and formate increase cancer stemness in vivo. **a**, Schematic overview of the tumour infection mouse model. SPF NSG mice were subcutaneously injected with HT-29 cells subcutaneously (1×10^6 cells per flank). After tumour formation, tumours were injected with *Fn* (MOI 10) or PBS (ctrl) or formate (10 mM, 60 μl) for five consecutive injections over 11 d. **b**, Formate levels in TIF of *Fn*-injected tumours. TIF volume was only sufficient for processing from $n=4$ independently treated tumours per group. **c**, Gene expression levels of *AHR* (far left), *ALDH1A1* (left), *CYP1B1* (right) and *SOX2* (far right) as assessed by rt-qPCR for $n=5$ independently treated tumours per group. **d**, ALDH activity assay of *Fn*-/PBS-infected tumours. Left, representative histogram of ALDH⁺ cell populations (crossbar) in mouse tumours assessed by FACS. Right, quantification of ALDH activity in tumours on *Fn* infection. **e**, CD133 (left) and CD24 (right) expression in mouse tumours after intratumoral injection, as assessed by FACS, $n=8$ independently treated tumours per group in **d** and **e**. **f**, Serial transplantation of *Fn*-treated xenografts. Tumours were explanted at endpoint, dissociated in culture and re-injected subcutaneously in secondary recipient mice at 5,000 cells per flank, $n=6$ biologically independent animals per group. Kaplan-Meier analysis of tumour incidence was performed using a Mantel-Cox test, $P=0.0024$. **g**, Expression of ALDH (left), CD44 (middle) and OCT4 (right) in formate-treated and control xenografts as assessed by FACS, $n=10$ and $n=12$ independently treated tumours for the control and treated conditions, respectively. **h**, Untargeted metabolite analysis of TIF samples from *Fn*-/PBS-infected xenografts. Heatmap shows normalized intensities, $n=8$ and $n=6$ independently treated tumours for the control and treated conditions respectively. Data in **b-d** (right), **e** and **g** are shown as mean \pm s.e.m., $P=0.0006$ in **b**, $P=0.0051$ in **c** (far left), $P=0.0482$ in **c** (right), $P=0.0262$ in **c** (far right), $***P<0.001$ in **d** (right), $P=0.0052$ in **e** (left), $P=0.0380$ in **e** (right), $P=0.0209$ in **g** (left), $P=0.0409$ in **g** (middle) and $P=0.0450$ in **g** (right), unpaired two-sided *t*-test. $*P<0.05$, $**P<0.01$, $***P<0.001$.

APCmin mouse organoids (Fig. 4f). A similar trend was observed in HT-29 cells upon *Fn* coculture (Extended Data Fig. 5d), altogether suggesting that microbiome-derived formate induces CSC traits in cancer cells. We further used the organoid model to confirm that microbiome-derived formate consequently potentiates the activity and function of CSC. We treated patient-derived organoids with formate for 24 hours. Treated organoids were then single-cell dissociated and reseeded to assess the long-term effects of formate on organoid formation capacity, which is known to be dependent on CSCs (Fig. 4g). After 10 days, we observed enhanced organoid formation upon reseeded formate-treated organoids compared to control organoids (Fig. 4h,i). These results clearly demonstrate that formate induces a higher CSC activity. Finally, we chose the tail vein dissemination assay to validate the formate-induced CSC activity in vivo. This assay focuses on the metastatic outgrowth of cells, which is, as mentioned above, highly dependent on cells possessing CSC traits^{35,36,38}. Formate pretreatment of HT-29 cells led to elevated formation of lung metastasis (Fig. 4j,k).

Formate activates the AhR signalling pathway. To better understand the underlying molecular mechanisms of formate-induced CSC activity and metastatic dissemination, we performed RNA-seq on formate-treated HT-29 cells. The IPA analysis revealed that AhR signalling was the most significantly activated pathway (Fig. 5a, $z=0.6$, $P<0.001$), similarly to the previous observations in patients bearing *Fn* and in HuMiX (Extended Data Fig. 1e,g, Fig. 2b and Extended Data Fig. 3f). Most of all, AhR not only plays an important role in xenobiotic metabolism, but it also plays a role in the regulation of CSC-like properties via ALDH⁴⁰ or downstream β -catenin/Wnt signalling activation⁴¹. We therefore proposed the AhR signalling pathway to be responsible for the formate-induced CSC activity described in Fig. 4. Accordingly, formate induced AhR nuclear translocation in HT-29 cells (Fig. 5b). We were able to rescue the effects of formate on CRC cell invasion with an AhR inhibitor (Fig. 5c), suggesting that *Fn*-derived formate induces cell invasion by activating AhR-induced cancer stemness. We propose that this effect is specific to some CRC-associated bacteria such as *Fn*, as Gram-negative *E. coli* did not induce increased tumour cell invasion (Fig. 5d). Concordantly, the pro-invasive effects of *Fn* tended to be rescued by adding an AhR inhibitor (CH223191) in vitro (Fig. 5e). We additionally observed induction of MAPK and Wnt signalling on formate treatment (pathways observed in Fig. 2b in HuMiX and Extended Data Fig. 5e,f), and the formate-induced invasion tended to be inhibited when Wnt signalling was blocked by the small-molecule inhibitor F535 (Extended Data Fig. 5g). It remains unclear how the AhR and Wnt signalling pathways are interlinked in the formate-induced invasion of CRC cells. In breast cancer models, the AhR pathway was described to control CSC

proliferation, development, self-renewal and chemoresistance through activation of the β -catenin pathway⁴². Along those lines, *Fn*-induced Wnt signalling was inhibited by AhR inhibition, suggesting that AhR activation can further lead to Wnt activation in our model (Extended Data Fig. 5h).

Finally, we observed a higher metastatic dissemination in the lungs after tail vein injection of cells pretreated with *Fn* compared to controls (Fig. 5f,g). Furthermore, the addition of an AhR inhibitor was able to revert the *Fn*-induced metastatic dissemination (Fig. 5f,h). These results demonstrate that *Fn* induces cancer stemness and thereby metastatic dissemination by activation of the AhR signalling pathway.

Altogether, our data indicate that AhR inhibitors might be of potential therapeutic interest in *Fn*-bearing CRC, who present bleaker outcomes than *Fn*-negative CRC patients²¹.

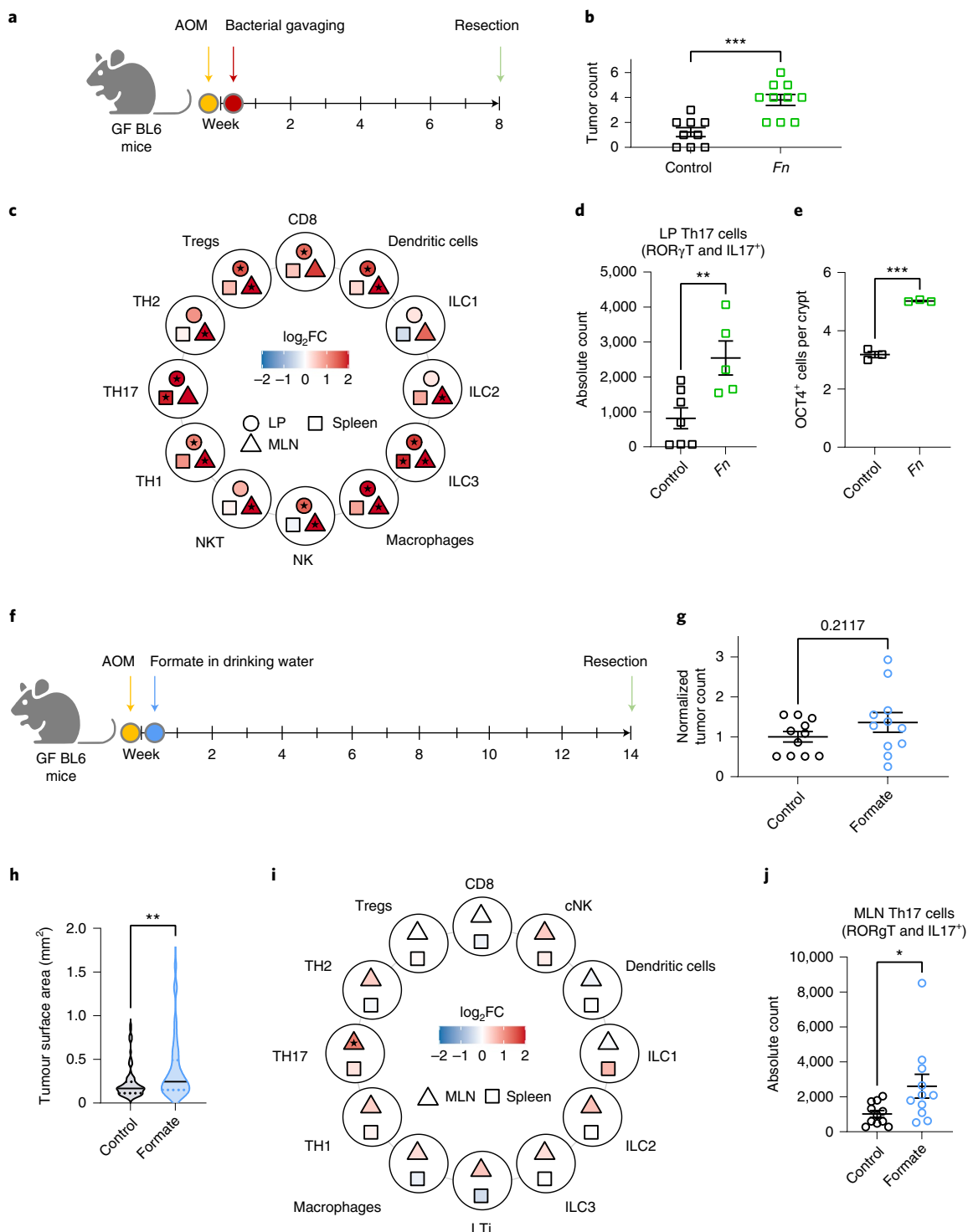
***Fn* and formate increase cancer stemness in mice.** To evaluate the effects of *Fn*-host cometabolism in vivo, we injected subcutaneous mouse tumours and xenografts with *Fn* (Fig. 6a(1)) or with formate (Fig. 6a(2)). Intratumoral *Fn* injection led to an increase of formate levels within tumour interstitial fluids (TIFs) (Fig. 6b), accompanied by an increased *AHR*, *CYP1B1*, *SOX2* and *ALDH1A1* (trendwise) gene expression (Fig. 6c). *Fn* intratumoral administration also led to higher numbers of ALDH⁺ cells in the tumours (Fig. 6d). These results, together with *Fn*-induced expression of CSC markers, such as CD133 and CD24 (Fig. 6e) (with the exception of CD44, Extended Data Fig. 5i), indicate that *Fn* plays a role in CSC regulation, adding to *Fn*-induced metastatic formation observed in colorectal^{43,44} and breast cancer⁴⁵. Of note, we observed that *Fn*-injected xenografts sustained tumorigenic potential in serial transplantation assays, showing increased frequency and shorter latency compared to PBS-injected xenografts (Fig. 6f and Extended Data Fig. 5j). Altogether, these results further support the role of *Fn* in regulating the potential for tumoral self-renewal. To further provide proof that the microbiome-derived formate induces cancer stemness in vivo, we additionally assessed CSC markers after xenograft exposure to formate. Similar to the *Fn* injections, we observed an effect on cancer stemness in tumours injected with formate (Fig. 6g and Extended Data Fig. 5l). Finally, the metabolic profile of the *Fn*-infected TIFs revealed decreased glutamine, lactate and pyruvate levels (Fig. 6h), as previously observed in HuMiX (Fig. 2). Our data therefore indicate that microbiome-derived formate induces CSC activity in tumours.

***Fn* and formate promote CRC formation and expand Th17 cells.** To study the role of *Fn* and its secretion product formate on tumour and immune cells during carcinogenesis, we used a chemically induced germ-free CRC mouse model and administered

Fig. 7 | *Fn* administration and formate treatment lead to an increase in tumour incidence or tumour size and an increase in Th17 cells. **a**, Schematic overview of the experimental setup. Germ-free mice received a single dose of AOM intraperitoneal (10 mg kg⁻¹). After 3 days, mice were gavaged with bacteria (10⁸ CFU per mouse) and euthanized at 8 weeks postinjection. **b**, Colonic tumour incidence, $n=9$ and $n=10$ biologically independent animals in the control and treated groups respectively, pooled from two independent experiments. **c**, Immune cell phenotyping of *Fn* versus PBS gavaged mice. Heatmap shows estimate (log₂ fold change (FC)) of normalized immune cell counts in mouse lamina propria (LP), MLNs and spleens. Filled star indicates $P<0.05$, two-tailed least squared means method. **d**, Th17 cell counts (CD4⁺IL-17⁺ROR γ T⁺ T cells) in LP from **c**. One representative experiment is shown with $n=7$ and $n=5$ biologically independent animals in the control and treated groups respectively, in **c** and **d**. **e**, OCT4⁺ cells per colonic crypt, $n=3$ biologically independent mice, with at least eight crypts per mouse. **f**, Schematic overview of the experimental setup. Germ-free (GF) mice received a single dose of AOM intraperitoneal (10 mg kg⁻¹) and formate was administered via the drinking water (250 mM) for the duration of the experiment (14 weeks). **g,h**, Tumour incidence (**g**) and surface areas (**h**) in colons of formate-treated mice and controls. **i**, Immune cell phenotyping of formate-treated mice and controls. Heatmap shows estimate (log₂ fold change) of normalized immune cell counts in the MLNs or spleens. Filled star indicates $P<0.05$, two-tailed least squared means method. **j**, Th17 counts (CD4⁺IL-17⁺ROR γ T⁺) in the MLNs from **i**. $n=11$ biologically independent animals per group from two independent experiments pooled for representation in **g,i** and **j**. **h**, One representative experiment with all tumour surface areas from four independent biological animals per condition. The solid line represents the median and the dotted line represents the upper and lower quartiles. Data in **b,d,e,g** and **j** are shown as mean \pm s.e.m. $P=0.0004$ in **b**, $P=0.0095$ in **d**, $***P<0.0001$ in **e**, $P=0.0363$ in **j**, unpaired two-tailed t -test, $P=0.00259$ in **h**, two-tailed nested ANOVA (factor mouse nested within condition). $*P<0.05$, $**P<0.01$, $***P<0.001$.

Fn via oral gavage (Fig. 7a). Significantly higher tumour counts were observed after *Fn* gavage (Fig. 7b and Extended Data Fig. 6a) compared to the gavage control, without effects on colon length and mouse weight (Extended Data Fig. 6b). These results were in line with data on similar mouse models⁴². *Fn* led to the expansion and activation of a range of proinflammatory immune cell types (for example, IL4-producing Th2 cells or ILC3s) (Fig. 7c and Extended Data Fig. 6c for the gating strategy). In parallel, we observed an expansion of the regulatory T cell (Treg) compartment upon fusobacterial gavage. An expansion of Treg populations in the spleen of *Fn*-injected mice has been previously described⁴⁶

and shown to correlate with *Fn* in patients, but this phenotype has not been previously observed in the lamina propria^{47,48}. Lamina propria NK cell levels were increased after *Fn* gavage (Fig. 7c). A similar effect of *Fn* on NK cells has also been observed following subcutaneous chamber infections⁴⁹. Notably, an expansion of CD4⁺IL-17⁺RORγT⁺ T cells was observed in the lamina propria of *Fn*-treated tumour bearing mice (Fig. 7d and Extended Data Fig. 6c). We also observed a higher number of OCT4-positive cells in the colons of *Fn*-treated mice compared to the control (Fig. 7e), indicating increased cancer stemness, potentially driven by AhR (Extended Data Fig. 6d).



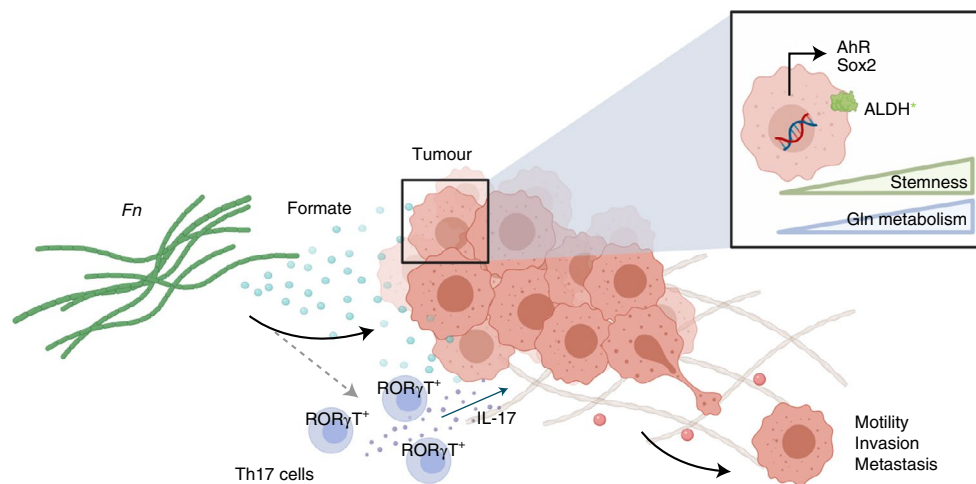


Fig. 8 | Microbiome-derived formate drives CRC invasion via secreted formate. The *Fn* metabolism product formate promotes AhR signalling in vitro and in vivo, which results in increased Th17 cell infiltration and ALDH activity, cancer stemness and metastatic dissemination of tumour cells.

We subsequently used a chemically induced germ-free mouse model (to avoid bias introduced by host microbiome formate production) and administered formate via the drinking water (Fig. 7f). Although tumour incidence was not significantly increased on formate treatment (Fig. 7g), average tumour size was higher compared to that of the control group (Fig. 7h). Similar to the *Fn* administration, formate treatment led to an expansion of CD4⁺IL-17⁺RORγT⁺ T cells in the mesenteric lymph nodes (MLNs) of formate-treated, tumour bearing mice (Fig. 7i,j), suggesting that the effect of *Fn* on Th17 cells is mediated via formate.

Discussion

In this study, we investigated the role of microbiome–host metabolic crosstalk in CRC by using *Fn*, a CRC-associated bacterium, as a model system. Using a metagenomics approach, we showed that tumours with high levels of *Fusobacterium* display an aberrant regulation of genes involved in the metabolism of tumour cells. Using in silico, in vitro and in vivo approaches, we identified formate as a bacterial oncometabolite involved in CRC progression (Fig. 8, graphical abstract).

Our HuMiX-based cometabolism experiments revealed a metabolic shift in tumour cells towards an increased glutamine metabolism. Metabolic reprogramming through increased glutamine metabolism can contribute to cancer therapy resistance^{48,49} and more invasive and metastatic cancer cells tend to be highly glutamine-dependent⁵⁰. Given these findings, the effect of *Fn* on chemoresistance⁵¹ might additionally be driven by altered glutamine metabolism. We observed increases in pyruvate, succinate and alanine metabolism in *Fn* upon coculture with tumour cells, as well as a reduction in a range of amino acids, suggesting fusobacterial use. An increased capacity of the microbiota to use amino acids via the putrefaction pathway resulting in polyamines, has been shown to promote colon carcinogenesis⁵². Yachida and colleagues observed increased levels of isoleucine, phenylalanine, tyrosine, valine in early-stage CRC³¹, presenting a potential niche to attract *Fusobacterium* and other microbes during these stages. Along this line, besides a depletion of fibre degradation genes, CRC is marked by the enrichment of genes for the use of host carbohydrates and amino acids, together with higher levels of *Fusobacteria*, *Peptostreptococcus*, *Porphyromonas* and others⁵³. Of note, lysine fermentation, along with reduced butyrate production, was recently shown to be *Fn*-linked in CRC patients⁵⁴.

Most importantly, we observed elevated levels of formate in our HuMiX coculture experiments. Higher formate levels have

previously been associated to chemotherapy non-responders in lung cancer, along with some bacterial species such as *Leuconostoc lactis* and *Eubacterium siraeum*⁵⁵. Thus, the effect of microbial changes on CRC chemoresistance^{56,57}, such as increased *Fn*⁵¹, could be driven by formate. Furthermore, we performed HuMiX experiments using different strains of *Fn* and observed a similar increase in formate levels upon coculture, independent of the strain used. Of note, *Fn* 7_1 is a clinical isolate and showed the highest upregulation in formate upon coculture, underlining the clinical relevance of our findings. This is further supported by a recent study showing high formate production by a clinical isolate of *Bacteroides faecalis*⁵⁸.

The integral results of our in vitro and in silico approaches suggest that a glucose-rich environment allowed for an increased short-chain fatty acid (SCFA) production by *Fn*, when in coculture with tumour cells, with formate as a major metabolic product. Since highly vascularized tumours gain more access to glucose from the bloodstream, the endothelial response in relation to *Fn* could favour formate metabolism. Along this line, studies have described the release of vascular endothelial growth factor in response to *Fn*⁵⁹. In addition, fusobacterial Fada binds VE-cadherin on endothelial cells, and causes VE-cadherin relocation from cell–cell junctions⁶⁰. The resulting increased endothelial permeability facilitates bacterial translocation and can ultimately lead to a hyperpermeable state, a hallmark of pathological angiogenesis found in tumours⁶¹. Higher tumour vascularization promotes oxidative metabolism, which can lead to a formate overflow in cancer cells³³, in addition to enhanced secretion of bacterial formate. Indeed, our in vitro and in vivo data hint at increased one-carbon metabolism, marked by serine consumption (Fig. 2d and Extended Data Fig. 3d). This is in line with a recent study showing elevated expression of MTHFD1L, an enzyme involved in formate production and one-carbon metabolism, in CRC patients⁶². Metagenomics data from CRC patients supported the hypothesis of an elevated formate metabolism via *pfl* along with a *Fusobacterium*-specific protumorigenic metabolic footprint. Nevertheless, the association of *pfl* activity and formate levels in *Fusobacterium*^{high} patients requires further analysis by using formate-deficient mutants in vitro and in vivo. Of note, *Fusobacterium* is significantly tumour-enriched, which may explain its role in specifically increasing formate at the tumour site.

Previous studies on metagenomics data have revealed bacterial formate oxidation as an overrepresented metabolic pathway in a chemically induced colitis mouse model³². Formate is known to be mainly produced by the gut microbiome, as formate cannot be detected in germ-free animals³². It is produced and secreted by

several different bacteria^{58,63}. In specific pathogen free (SPF) mice, dysbiosis was accompanied by an increase in luminal formate levels in a millimolar range³². These metabolic shifts can ultimately bridge the colonization of, and provide fitness advantages to, further opportunistic pathogens (for example, pathogenic *E. coli* strains) possessing formate dehydrogenase and terminal oxidase genes, as shown in murine models of colitis³². Furthermore, a recent study associated formate, propionate, and acetate levels to *Faecalibacterium* and *Bifidobacterium* in the faecal microbiomes of inflammatory bowel syndrome patients⁶⁴. Taken together, these studies underline the variety of formate-producing microbes, further extending its importance beyond *Fn*. Nevertheless, the protumorigenic effect of formate might be context dependent and regulated by other metabolic factors in the tumour niche. Further studies are required to explain these complex interactions surrounding formate and its function in the tumour microenvironment.

SCFA have been shown to control important signalling pathways responsible for AhR signalling and stem cell proliferation and maintenance: either as inducers or inhibitors^{65–67}. Using various *in vitro* and *in vivo* models, we now show that formate is able to induce invasion of cancer cells potentially by inducing AhR signalling and its associated CSC regulation. These results agree with our previous study in which formate increased the tumour-initiating capacity of cancer cells⁶⁸. Taken together, ALDH activity, as well as signalling activation of the genomic and non-genomic AhR pathways with functional consequence for other pathways (β -catenin/Wnt, NF- κ B, MAPK and FAK)⁴⁰, indicate that microbiome-secreted formate might play an important role in the malignant transformation of intestinal cells and the control of their stemness properties.

Furthermore, immune cell regulation might be an additional factor in formate-driven tumorigenesis. In line with other previous studies⁴², our *in vivo* data showed *Fn*-accelerated tumour incidence. In general, we observed a proinflammatory effect of *Fn* through the expansion of CD4⁺IL-17⁺ROR γ T⁺ T cells in MLNs. It has been shown that activation of AhR during murine Th17 cell development increases Th17 T cell populations and their production of cytokines⁶⁹. Consequently, formate, secreted by *Fn* or other tissue invading microbes, may affect the Th17 T cell compartment in CRC. Accordingly, we observed that supplementation of formate in mice led to an increase in CD4⁺IL-17⁺ROR γ T⁺ T cells. Hence, AhR inhibitors, which have already been described as a potential new therapeutic target in cancer⁷⁰ may be used for patients with *Fusobacterium*-positive tumours or with elevated levels of formate and formate-producing bacteria.

Altogether, our study offers an integrative approach on studying the role of the microbiota in CRC with a focus on host–microbe cometabolism. A protumorigenic phenotype that comes with enhanced fusobacterial metabolite secretion has not been observed before in the CRC context. The results indicate implications for the role of microbiome-derived metabolites, especially formate, as a key player responsible for *Fusobacterium*'s protumorigenic properties via enhanced AhR signalling and CSC regulation. Of note, our study demonstrates the role of formate in cancer stemness, and in conjunction with other previously discussed studies, it further raises the question of community-wide formate production, not only by *Fusobacterium*, but also by other bacterial species. Altogether, we identified microbiome-derived formate as an oncometabolite in CRC.

Methods

Human sample acquisition and DNA extraction. Patient samples were donated willingly under informed consent and were handled in accordance with institutional guidelines. Ethical approval was obtained from the Comité National d'Éthique de Recherche, Luxembourg (reference 201009/09) followed by institutional approval by the Ethics Review Panel of the University of Luxembourg (ERP-16-032). Patients underwent surgery at the Centre Hospitalier Emile Mayrisch (Esch-sur-Alzette, Luxembourg), the Centre Hospitalier de Luxembourg

(Luxembourg) or the Zitha Klinik (Luxembourg). A healthy control cohort (ND collection) was enrolled by the National Centre of Excellence in Research on Parkinson's disease (NCER-PD) and 16S rRNA gene sequencing data were shared according to a biomaterial cooperation agreement. The average ages of stool sample donors were 70.29 \pm 8.74, with 12 female and 40 male participants (in-house CRC collection) and 62.05 \pm 7.51, with 30 female and 33 male participants (ND collection). All patients were enrolled between the years 2010 and 2019. No participant compensation was provided. Stool samples were collected in OmniGut tubes (DNA Genotek) at the time of surgery (patients with CRC) or at home (healthy donors). Patients received a mild phosphate enema the day before surgery. Metadata can be found in (Supplementary Table 1).

Primary colon cancer tissue and matched distant non-neoplastic colon tissue (at the farthest longitudinal surgical margin) were collected following the standard preanalytical practices for biospecimens, established by the Integrated Biobank of Luxembourg (IBBL). Immediately after surgery, fresh samples were transported to the laboratory on ice for further downstream processing (generation of primary cell cultures and organoids within one hour after resection). A pathologist from the National Center of Pathology, Laboratoire National de Santé provided the clinical and histopathological data.

RNA and DNA were extracted from stool or tissue with the AllPrep kit for RNA, DNA and protein extraction (Qiagen).

16S rRNA gene sequencing analysis. 16S rRNA gene sequencing was performed at the IBBL. Sequencing of the V3 and V4 regions of prokaryotic 16S rRNA gene was performed on the Illumina MeSeq platform using 2 \times 300 bp paired-end reads. Raw 16S rRNA gene sequences (FASTAQ files) were cleaned and clustered as operational taxonomic units using NgTax or the Dada2 pipeline. Bacterial abundances were plotted as base means.

Bacterial strains and growth conditions. Bacterial strains were obtained from the German Collection of Microorganisms and Cell Cultures (DSMZ) or through collaborations and are listed in Supplementary Table 5. Adjusted from the DSMZ culture recommendations, bacterial strains were revived from glycerol stocks in 3 ml of anoxic brain heart infusion broth (Merck). Bacterial inoculums were maintained in an anaerobic chamber at 37°C, 95% N₂ and 5% CO₂.

Cell lines. HCT116, HT-29 and Caco-2 CRC cell lines were obtained from ATCC and maintained in DMEMF12 with 10% [v/v] foetal bovine serum (FBS) and 1% [v/v] penicillin/streptomycin. RKO-7TGP from the University of Vienna (Dolznig laboratory) were maintained as previously described⁷¹. HT-29 cells were obtained from ATCC. Firefly luciferase-green fluorescent protein (GFP) lentiviral particles (PLV-10172-50, Cellomics Technology) were used to transduce HT-29 cells (HT-29-Luc) at 1 \times 10⁸ TU per ml per followed by puromycin selection. Cell lines (purchased from ATCC) were frequently checked to be mycoplasma-free and were authenticated before and after their use in this study (STR analysis, DSMZ).

Primary organoid generation. Muscle and mucus were removed from tissue samples (human or C57BL/6J-*Apc*^{Min}/J mouse colon) using surgical scissors. The tissue was cut into small pieces and washed 3–4 times with cold PBS until all visible debris was removed. The tissue pieces were resuspended in 5 ml of Gentle Cell Dissociation Reagent (STEMCELL Technologies) and placed at 4°C on a benchtop roller for 30–60 min. After incubation, the tissue pieces were washed three times in 10 ml of cold PBS (vigorous shaking) and discarded. The washes were filtered through a 70- μ m pore-sized filter and crypts were resuspended at 300 crypts per 20 μ l of Matrigel matrix per dome (Corning) and seeded into 48-well plates. Matrigel was left to solidify for 10 min at 37°C, then 250 μ l of Human or Mouse Organoid Growth Medium (hOGM/mOGM, STEMCELL Technologies) with 10 μ M Y-27632 (STEMCELL Technologies, catalogue no. 72304) was added to each well. Organoids were passaged after 7–14 d with total medium changes (without Y-27632) every 3–4 d. Organoids were passaged using Gentle Cell Dissociation Reagent and reseeded as described above at a ratio of 1:1.5–1:2.

HuMiX cocultures. The HuMiX device is a modular bioreactor perfusion system allowing for the coculture of bacteria and human cells, as described previously. HuMiX devices are assembled and primed 1 d before the beginning of the experiments. In short, the top chamber was perfused with N₂ gas (0.1 l min⁻¹ flow rate), establishing anoxia (roughly 0.4–0.6% O₂). It was separated from the underlying microbial chamber by a nanoporous membrane (0.05 μ m) coated with porcine mucin (Sigma-Aldrich), dissolved in ddH₂O (0.025 mg ml⁻¹). Another nanoporous membrane separated the microbial chamber from the underlying epithelial chamber, which in turn was coated with rat-tail collagen solution (50 μ g ml⁻¹, Corning). A microporous membrane (1.0 μ m) separated the epithelial chamber from the bottom chamber, which supplied fresh oxigen (roughly 6% oxygen) growth medium to the human cells. The constant perfusion flow rate was 25 μ l min⁻¹ (0.5 r.p.m., peristaltic pump). Epithelial cells (HT-29, Caco-2) were inoculated 6 d before bacterial inoculation (*Fn* 25586, *Fn* 23726 and one clinical isolate *Fn* ssp. *animalis* 7_1). Bacterial–human coculture lasted 24 h before the devices were opened and supernatants and cells were collected for downstream

analysis (RNA, DNA, proteins, extracellular metabolites). Control devices with bacterial or human cell monocultures were run in parallel.

RNA-seq. For library preparation, 1 µg of RNA was poly(A) selected, fragmented and reverse transcribed by using the Elute, Prime, Fragment Mix (Illumina). End repair, poly(A)-tailing, adaptor ligation and library enrichment were performed as described in the Low Throughput protocol of the TruSeq RNA Sample Prep Guide. RNA library quantity and quality were assessed with the Agilent 2100 BioAnalyzer (Agilent). RNA libraries were sequenced as 100 bp paired-end runs on an Illumina HiSeq2500 platform.

Metabolomics. Before untargeted semiquantitative GC–MS analysis (Agilent MassHunter, v.B.08.00), sample derivatization was carried out using an automated sample preparation robot in two steps: methoxamine hydrochloride ketone stabilization and *tert*-butyldimethylsilyl group replacement. After derivatization, we carried out GC–MS analysis in an Agilent 7890A GC coupled to an Agilent 5975C inert XL Mass Selective Detector. For precision quantification, GC–MS measurements of selected derivatives were performed using a selected ion monitoring mode. The automated in-house designed Metabolite Detector software, in combination with an in-house library, were used for peak assignment of all chromatographs and metabolites⁷².

For highly volatile short chain fatty acid detection (targeted and absolute quantitative), the derivatization was performed in the organic phase using diethyl ether. A dilution series of a volatile free acid mix was prepared for the quantitative measurement. Then 10 µl of 200 mM ethylbutyric acid (internal standard), 1 ml of 99% diethyl ether and 10 µl of 1 mM hydrochloric acid were added to each reaction tube primed with 190 µl of fluid (fresh medium, ultrapure water as blanks, for each time point of the dilution series). The mix was agitated for 10 min at room temperature and centrifuged at maximum speed for 5 min at 15°C. The upper phase was transferred to a new 2-ml reaction tube. Then 1 ml of diethyl ether was added to the tube with the lower phase and a second agitation/centrifugation/transfer step was performed. 250 µl of each sample were then distributed into GC vials in triplicates and 25 µl of MTBSTFA (with 1% *tert*-butyldimethylsilyl chloride) were added to them. Finally, the samples were analysed using GC–MS. Quantification of metabolite levels in mouse TIF by liquid chromatography–mass spectrometry as previously described (minimal volume 50 µl required, samples with yields under 50 µl were not processed, Xcalibur software, v.4.3.73.11)⁷³. All resulting raw data were analysed in R.

Western blot. Cells were washed with 4°C PBS and lysed with 4°C RIPA buffer containing 1% SDS and protease inhibitor. After addition of Lämmli buffer, lysates were vortexed and supernatants were heated to 95°C for 5 min. Then 20 µg of proteins were resolved by SDS–PAGE on 12% gels, transferred onto a polyvinylidene difluoride–PLUS Transfer Membrane (0.2 µm). Blocking was performed with TBS-N + 10% BSA for 1 h at room temperature, followed by an overnight incubation at 4°C for probing with specific primary antibodies (MAB1501, Millipore and sc-32293, Santa Cruz at 1:5,000; 12035, 18818S, 4370S, 4694, 9154S, 9212, 4511, 3033T, Cell Signalling and sc-93, sc-109, Santa Cruz at 1:1,000) diluted in TBS-N + 5% BSA. Membranes were washed three times with TBS-N and incubated for 1 h at room temperature with the corresponding horseradish peroxidase (HRP)-labelled secondary antibody (7076s and 7074 at 1:500). The blots were revealed using enhanced chemoluminescence on a Fusion FX imaging platform.

Real-time quantitative PCR (qPCR). RNA extractions, reverse transcriptions and real-time qPCRs were carried out according to previously reported protocols²⁸. Reference genes used for qPCR were *ACT1* and *GAPDH* (for mouse) and *YWHAZ* and *EEF1A1* (for human). Primer pairs used for qPCR with reverse transcription (RT–qPCR) are listed in either Supplementary Table 5 or described previously²⁸. rt–qPCR was analysed using qBase+ v.3.2 (Biogazelle) according to The Minimum Information for Publication of Quantitative Real-Time PCR Experiments guidelines.

Reporter gene assay. Here, 50,000 7TGP-RKO β-catenin/Wnt signalling reporter cells were seeded in 24-well plates. Bacterial suspensions were added to the cell growth medium at a multiplicity of infection (MOI) of 1:20 for 1 h. Cells were then washed twice with medium, supplemented with 1% [v/v] penicillin/streptomycin and further incubated in antibiotic-containing media for 48 h. For supernatant, formic acid 10 mM + 1% [v/v] 1 N NaOH or control treatments, 10% [v/v] cell media was supplemented accordingly and cells were incubated for 48 h. After trypsinization, cells were stained with LIVE/DEAD Fixable Near-IR Dead Cell Stain (ThermoFisher) and viable GFP⁺ cells were acquired by fluorescence-activated cell sorting (FACS). Standards of rhWnt3A protein were used.

AhR nuclear translocation assay. HT-29 cells were treated with 6-formylindolo(3,2-b)carbazole (FICZ) alone (100 nM), or FICZ (100 nM) and formate (10 mM) for 6 h. After treatment, cells were fixed in 4% paraformaldehyde at room temperature for 10 min and permeabilized with 0.1% Triton-X. Cells were stained with 1:500 dilutions of a polyclonal anti-AhR antibody (Enzo

Life Sciences, BML-SA210), followed by donkey anti-rabbit AF647 secondary antibody (ThermoFisher, A-31573). Nuclei were visualized using 4,6-diamidino-2-phenylindole (DAPI). Fluorescence micrographs of cells were obtained with a fluorescent microscope (OLYMPUS IX83) at ×60 magnification. Intensity of AhR staining in nuclei and cytoplasm was assessed using ImageJ, and ratio of AhR intensity between nucleus and cytoplasm was calculated for single cells.

Scratch wound invasion assay. Cancer cell lines were seeded on rat-tail collagen type I-coated (200 µg ml⁻¹) 96-well Imagelock plates (Sartorius) at approximately 90% confluence in cell culture medium containing rat-tail collagen type I (1 mg ml⁻¹). A wound was scratched in the cell monolayer across each well with the CellplayerTM 96-well WoundMaker (Sartorius). Confluence of the wound area was monitored with the IncuCyte imaging system (Sartorius) every 3 h for a total of 72 h. At the end of the experiment, the IncuCyte ZOOM software (2018B, Sartorius) was used to calculate a relative wound density for each time point by measuring the spatial cell density in the wound area relative to the spatial cell density outside the wound area.

Transwell invasion assay. Here, 50,000 HCT116 cells were seeded in serum-free medium in the upper compartment of ThinSert transwells (8 µm, Greiner). Medium containing 10% FBS [v/v] was added to the bottom compartment as a chemoattractant. Cells on transwells were treated with formate (10 mM) ± AhR inhibitor (CH223191, 0.5 µM), or preinfected with *Fn* (MOI 1:10) for 2 h, washed twice with medium, supplemented with 1% [v/v] penicillin/streptomycin and further incubated in antibiotic-containing media for 24 h ± AhR inhibitor, at 37°C for 48 h. Cells were fixed with 4% PFA and stained with the cytological dye crystal violet (0.125% [v/v] in PBS; Sigma-Aldrich). Residual non-invasive cells were manually removed from the apical membrane side with a cotton swab. The transwells were dried overnight at room temperature and examined under a microscope. ImageJ software was used to and measure contrast-based cell confluence of treated cells relative to control (untreated) cells.

Focal adhesion assay. Here, 50,000 cells were seeded on 1 cm Ø glass cover slips in 12-well plates and treated with formate (1 or 10 mM) or rhWnt3A (100 ng ml⁻¹, PeproTech) for 48 h. Cells were fixed in 3.5% PFA for 30 min, washed with PBS and PBS-Triton 1%, blocked with 5% milk and washed again in PBS-Triton 0.5%. Focal adhesions were stained overnight with mouse-anti-Vinculin (primary antibody; V4505, Sigma-Aldrich), before being further washed with PBS-Triton 0.5% and stained with goat-anti-mouse-DyLight488 (ab96879, Abcam), Phalloidin AlexaFluor594 (A12381, ThermoFisher) and DAPI (ThermoFisher). Pictures were randomly taken under a confocal microscope and images were analysed using ImageJ (version 1.53K).

Organoid assay. Organoids embedded in Matrigel were treated with 5 mM sodium formate (Carl Roth, 4404.1), for 48 h at 37°C. ALDH activity was assessed according to the manufacturer's protocol (ALDEFLUOR Kit, Stemcell Technologies, 01700). CD133 (Miltenyi Biotec, 130-090-826), CD44 (BD Biosciences, 555428) and CD24 (BD Biosciences, 560533) labelling was done at a 1:500 dilution. LIVE/DEAD cell viability staining was performed at a 1:1,000 dilution (Invitrogen, L34975). Samples were acquired on a FACS CantoII Cell Analyzer (BD Biosciences). Results were analysed with the FlowJo software (BD Biosciences).

For functional assays, organoids derived from patients with CRC embedded in Matrigel (Corning, catalogue no. 356231), were treated with 10 mM sodium formate (Carl Roth, catalogue no. 4404.1) or PBS as a control in hOGM (STEMCELL Technologies, catalogue no. 06010) for 24 h at 37°C. Matrigel domes were dissociated and reseeded as described above, at 40 × 10³ dissociated single cells per well. Organoids were maintained for 10 d as described above. Plates were imaged with Cytation 5 (BioTek). Organoids were counted using ImageJ.

Constraint-based modelling. The microbial reconstructions from the AGORA resource²⁹ were downloaded from the Virtual Metabolic Human website (www.vmh.life)⁷⁴. The models were manually curated and analysed through functions implemented in the COBRA 3.0 Toolbox⁷⁵ using IBM ILOG Cplex (IBM, Inc.). Pairwise models were built and interrogated using the pairwise modelling module in the Microbiome Modelling Toolbox⁷⁶. The *Fn* model for the interaction analysis was curated on the basis of the Fusobacterium_nucleatum_subsp_nucleatum_ATCC_25586 AGORA model (version November 2019 from vmh.life; addtoFNmodel.m script: addition of 18 reactions, for example for the uptake and hydrolysis of dipeptides, the production of *N*-acetylputrescine or the transport of lactate and 2-hydroxybutyrate). The simulations were performed as previously described⁶⁸. The HT-29 model was set up as a conditioned, cell type specific model on the basis of RNA-seq gene expression data of HT-29 monocultures in HuMiX with a Recon 2.0 background as previously described⁸.

Animal models. All animal experiments were performed according to all applicable laws and regulations (EU Directive 2010/63/EU and Grand-Ducal Regulation of 11 January 2013 on the protection of animals used for scientific purposes), after receiving approval from the Animal Experimentation Ethics

Committee at the University of Luxembourg (AEEC) and the Ministry of Agriculture, Viticulture and Rural Development (LUPA 2019/99 and LUPA 2019/60). Mice were housed in a SPF or a germ-free facility at a relative humidity of 40–70%, at 22 °C and in 12 h dark/light cycles. They received food (SAFE A40) and water ad libitum.

Dissemination/metastasis model (SPF). Here, 7–8-week-old in-house bred male and female nod scid gamma (*Mus musculus* NSG) mice were intravenously injected with $0.5\text{--}1 \times 10^6$ per 200 μl of HT-29-Luc cells, pretreated with or without 10 mM formate (24 h) or preinfected with *Fn* (MOI 1:10) for 2 h, washed twice with medium, supplemented with 1% [v/v] penicillin/streptomycin and further incubated in antibiotic-containing media for 24 ± 4 h AhR inhibitor (CH223191, 0.5 μM). After 30 d, mice were intraperitoneally injected with luciferin (15 mg ml^{-1} , 100 μl per 10 g mouse weight). Then 10 min postinjection, signals were detected using the In Vivo Imaging System (IVIS). Macroscopic counting of metastasis nodes of the lungs was performed blindly.

Xenograft infection models (SPF). For this, 7–8-week-old in-house-bred male and female nod scid gamma (*M. musculus* NSG) mice were subcutaneously injected with 1×10^6 HT-29 tumour cells. When tumours reached a size of approximately 200 mm^3 , 5×10^7 *Fn* ATCC 25586 or formate (10 mM) was injected intratumorally (three injection sites per tumour) every other day for 10 d (for a total of five injection time points). Xenografts were harvested at endpoint, weighed, subjected to TIF extraction as previously described⁷³ and dissociated (Human Tumour Dissociation Kit, Miltenyi) for further FACS-based protein expression analysis. For serial transplantation of *Fn*-injected xenografts, primary tumours were explanted, dissociated in culture and re injected in secondary recipient mice, as previously described⁷⁸. Tumour incidence and growth was monitored over time blindly by palpation of the flanks. The maximal total tumour volume of 2,000 or 1,500 mm^3 per single tumour permitted by the AEEC was not exceeded.

Treatment models (bacteria and formate in germ-free). Here, 11–15-week-old germ-free in-house-bred C57BL/6Ntac mice (*M. musculus*) received a single 10 mg kg^{-1} intraperitoneal azoxymethane (AOM) injection (A5486, Sigma-Aldrich). Three days later, male and female mice received 10^8 colony forming units (CFU) *Fn* suspended in 200 μl of PBS via oral gavage (PBS alone as control) or formate (250 mM) was provided in the drinking water (male mice only) for the course of the experiment (pure drinking water as control)⁷⁷. At endpoint (8 or 14 weeks after AOM injection, for bacterial gavage or formate treatment respectively), mice were euthanized via intraperitoneal injection of medetomidine (1 mg kg^{-1}) and ketamine (150 mg kg^{-1}). Organs were harvested and snap-frozen in liquid nitrogen on collection and stored at -80 °C until analysis. A 0.5 cm snip of the colon containing a stool pellet was embedded in TissueTek and snap-frozen in liquid nitrogen. Another 0.5 cm snip was directly snap-frozen in liquid nitrogen. The rest of the colon was cut open longitudinally, macroscopic tumours were counted blindly by two independent researchers, and the colon was subjected to immune cell phenotyping as described below. Photographs of the tumours were taken and tumour size was quantified using ImageJ from photographs acquired on a Wild M10 (Leica), at $\times 8$ magnification.

Immune cell phenotyping. Colons were opened longitudinally, washed with cold PBS, cut into smaller sections and digested first in Roswell Park Memorial Institute medium (RPMI) containing glutamine/Glutamax, 3% FBS, 1% P/S, 5 mM EDTA and 0.154 mg ml^{-1} DTT, then in serum-free RPMI with Liberase (0.1 mg ml^{-1}) and DNase (0.5 mg ml^{-1}), both in a shaking incubator at 800 r.p.m. and 37 °C for 20 and 30 min, respectively. Spleens were digested in serum-free RPMI with Liberase (0.1 mg ml^{-1}) and DNase (0.5 mg ml^{-1}) for 30 min. Digestions were quenched with RPMI medium containing 10% FBS and cell suspensions were filtered through a 70- μm filter. MLNs were mashed through a 70- μm filter. Single-cell suspensions were restimulated in RPMI containing 10% FBS, 0.1 $\mu\text{g} \text{ml}^{-1}$ PMA and 1.5 $\mu\text{g} \text{ml}^{-1}$ ionomycin for 4 h at 37 °C. Brefeldin A was added 1 h after the beginning of the restimulation. Cells were stained with Near-IR fixable live/dead dye and with the corresponding antibody panels (Supplementary Table 6) using the Cytofix/Cytoperm kit (554714, BD Bioscience). An additional Fc blocking step preceded antibody staining for macrophages and dendritic cells. Acquisition was performed on a CantoII and a Fortessa FACS machine and data were analysed using BD FACSDiva v.8.0.1 and FlowJo software v.10.6.1 (gating strategies can be found in Extended Data Figs. 7 and 8). Total cell counts were determined by Cedex and multiplied by frequency of subpopulations to get absolute counts.

Animal tissue sections and staining. Animal tissues were embedded in TissueTek (Sakura) and snap-frozen in liquid nitrogen immediately on collection. Cryosections were fixed with ice-cold methanol for 10 min at -20 °C, blocked with a 3% BSA solution and incubated overnight at 4 °C with primary antibodies (Supplementary Table 5). Secondary antibodies (Supplementary Table 5) were applied for 1 h at room temperature. Sections were mounted with DAPI Fluoromount-G (ThermoFisher), viewed under a confocal microscope and processed with ImageJ.

For OCT4 staining, 7- μm cryosections were fixed in acetone and stained with 1:500 dilutions of anti-OCT4 (primary antibody, Abcam, ab18976) and goat anti-rabbit AF647 (secondary antibody, Invitrogen, A-21245). Sections were mounted using DAPI Fluoromount-G (SouthernBiotech, 0100–20). Images were acquired using a Cytation 5 microscope (BioTek Instruments) and analysed using ImageJ.

Computational tools, data and statistical analysis. R studio (v.1.2.5042, R studio Team 2020) with R CRAN (v.4.0.3, R Core Team 2020) was used for the analysis of 16S rRNA gene and RNA/WX sequencing, as well as for metabolomics data analysis. Plots were made using ggplot2. MATLAB (R2018B) and the COBRA toolbox (3.0) were used for the constraint-based modelling. PathSeq analysis was conducted as described in²². 16S and RNA-seq data were analysed using the dada2. RT-qPCR data were analysed using qBase+ 3.2 (Biogazelle). We used the CMS subtypes published by Guinney et al. for TCGA samples. As EGA samples were not characterized in the Guinney report, we used the R CMScaller package to characterize the subtypes⁷⁸. Data were analysed through the use of IPA (QIAGEN Inc., <https://www.qiagenbioinformatics.com/products/ingenuity-pathway-analysis>) with input lists of differentially expressed genes, filtered for \log_2 fold change > 1.5 and $P < 0.05$. GraphPad Prism or R was used for statistical analysis, which is further described in the respective figure legends.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Sequence data from the EGA are available under accession number EGAS00001000288. Sequence data from the patients with CRC and healthy donor samples can be found on ENA (EBI) under the accession number PRJEB46665. The TCGA analysis was based on data (TCGA-COAD, TCGA-READ) generated by the TCGA Research Network: <https://www.cancer.gov/tcga>. All scripts and models can be found under https://gitlab.lcsb.uni.lu/mdm/ternes_et_al_2022/. Microbial reconstructions are modified from the Virtual Metabolic Human website (www.vmh.life) and available on gitlab. Source data are provided with this paper.

Received: 15 July 2021; Accepted: 25 February 2022;
Published online: 18 April 2022

References

- Bray, F. et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **68**, 394–424 (2018).
- Pleguezuelos-Manzano, C. et al. Mutational signature in colorectal cancer caused by genotoxic *pks+* *E. coli*. *Nature* **580**, 269–273 (2020).
- Kim, H. J., Huh, D., Hamilton, G. & Ingber, D. E. Human gut-on-a-chip inhabited by microbial flora that experiences intestinal peristalsis-like motions and flow. *Lab Chip* **12**, 2165–2174 (2012).
- Shah, P. et al. A microfluidics-based in vitro model of the gastrointestinal human–microbe interface. *Nat. Commun.* **7**, 11535 (2016).
- Bauer, E. & Thiele, I. From network analysis to functional metabolic modeling of the human gut microbiota. *mSystems* **3**, e00209–e00217 (2018).
- Pacheco, M. P. et al. Identifying and targeting cancer-specific metabolism with network-based drug target prediction. *EBioMedicine* **43**, 98–106 (2019).
- Ternes, D. et al. Microbiome in colorectal cancer: how to get from meta-omics to mechanism? *Trends Microbiol.* **28**, 401–423 (2020).
- Abed, J. et al. Fap2 mediates *Fusobacterium nucleatum* colorectal adenocarcinoma enrichment by binding to tumor-expressed Gal-GalNAc. *Cell Host Microbe* **20**, 215–225 (2016).
- Casasanta, M. A. et al. *Fusobacterium nucleatum* host-cell binding and invasion induces IL-8 and CXCL1 secretion that drives colorectal cancer cell migration. *Sci. Signal.* **13**, eaba9157 (2020).
- Gur, C. et al. Binding of the Fap2 protein of *Fusobacterium nucleatum* to human inhibitory receptor TIGIT protects tumors from immune cell attack. *Immunity* **42**, 344–355 (2015).
- Rubinstein, M. R. et al. *Fusobacterium nucleatum* promotes colorectal carcinogenesis by modulating E-cadherin/ β -catenin signaling via its Fada adhesin. *Cell Host Microbe* **14**, 195–206 (2013).
- Kaplan, C. W. et al. *Fusobacterium nucleatum* outer membrane proteins Fap2 and RadD induce cell death in human lymphocytes. *Infect. Immun.* **78**, 4773–4778 (2010).
- Martin-Gallausiaux, C., Malabirade, A., Habier, J. & Wilmes, P. *Fusobacterium nucleatum* extracellular vesicles modulate gut epithelial cell innate immunity via FomA and TLR2. *Front. Immunol.* **11**, 3298 (2020).
- Zanzoni, A., Spinelli, L., Braham, S. & Brun, C. Perturbed human sub-networks by *Fusobacterium nucleatum* candidate virulence proteins. *Microbiome* **5**, 89 (2017).
- Richardson, A. R. Virulence and metabolism. *Microbiol. Spectr.* <https://doi.org/10.1128/microbiolspec.GPP3-0011-2018> (2019).

16. Wei, Q., Qian, Y., Yu, J. & Wong, C. C. Metabolic rewiring in the promotion of cancer metastasis: mechanisms and therapeutic implications. *Oncogene* **39**, 6139–6156 (2020).
17. Bullman, S. et al. Analysis of *Fusobacterium* persistence and antibiotic response in colorectal cancer. *Science* **358**, 1443–1448 (2017).
18. Dieter, S. M. et al. Distinct types of tumor-initiating cells form human colon cancer tumors and metastases. *Cell Stem Cell* **9**, 357–365 (2011).
19. Qureshi-Baig, K., Ullmann, P., Haan, S. & Letellier, E. Tumor-initiating cells: a critical review of isolation approaches and new challenges in targeting strategies. *Mol. Cancer* **16**, 40 (2017).
20. Wang, Q., Yu, C., Yue, C. & Liu, X. *Fusobacterium nucleatum* produces cancer stem cell characteristics via EMT-resembling variations. *Int. J. Clin. Exp. Pathol.* **13**, 1819–1828 (2020).
21. Yan, X., Liu, L., Li, H., Qin, H. & Sun, Z. Clinical significance of *Fusobacterium nucleatum*, epithelial–mesenchymal transition, and cancer stem cell markers in stage III/IV colorectal cancer patients. *OncoTargets Ther.* **10**, 5031–5046 (2017).
22. Kostic, A. D. et al. PathSeq: a comprehensive computational tool for the identification or discovery of microorganisms by deep sequencing of human tissue. *Nat. Biotechnol.* **29**, 393–396 (2011).
23. Warren, R. L. et al. Co-occurrence of anaerobic bacteria in colorectal carcinomas. *Microbiome* **1**, 16 (2013).
24. Purcell, R. V., Visnovska, M., Biggs, P. J., Schmeier, S. & Frizelle, F. A. Distinct gut microbiome patterns associate with consensus molecular subtypes of colorectal cancer. *Sci. Rep.* **7**, 11590 (2017).
25. Guinney, J. et al. The consensus molecular subtypes of colorectal cancer. *Nat. Med.* **21**, 1350–1356 (2015).
26. Salvucci, M. et al. Patients with mesenchymal tumours and high *Fusobacteriales* prevalence have worse prognosis in colorectal cancer (CRC). [Gut **https://doi.org/10.1136/GUTJNL-2021-325193**](https://doi.org/10.1136/GUTJNL-2021-325193) (2021).
27. Thanki, K. et al. Consensus molecular subtypes of colorectal cancer and their clinical implications. *Int. Biol. Biomed. J.* **3**, 105–111 (2017).
28. Qureshi-Baig, K. et al. What do we learn from spheroid culture systems? Insights from tumorspheres derived from primary colon cancer tissue. *PLoS ONE* **11**, e0146052 (2016).
29. Magnúsdóttir, S. et al. Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota. *Nat. Biotechnol.* **35**, 81–89 (2017).
30. Ogawa, A. T. et al. Characterization of proteolytic activities of *Fusobacterium nucleatum*. *J. Endod.* **32**, 521–523 (2006).
31. Yachida, S. et al. Metagenomic and metabolomic analyses reveal distinct stage-specific phenotypes of the gut microbiota in colorectal cancer. *Nat. Med.* **25**, 968–976 (2019).
32. Hughes, E. R. et al. Microbial respiration and formate oxidation as metabolic signatures of inflammation-associated dysbiosis. *Cell Host Microbe* **21**, 208–219 (2017).
33. Meiser, J. et al. Increased formate overflow is a hallmark of oxidative cancer. *Nat. Commun.* **9**, 1368 (2018).
34. Fung, K. Y. C., Cosgrove, L., Lockett, T., Head, R. & Topping, D. L. A review of the potential mechanisms for the lowering of colorectal oncogenesis by butyrate. *Br. J. Nutr.* **108**, 820–831 (2012).
35. Ye, X. et al. Distinct EMT programs control normal mammary stem cells and tumour-initiating cells. *Nature* **525**, 256–260 (2015).
36. Mani, S. A. et al. The epithelial–mesenchymal transition generates cells with properties of stem cells. *Cell* **133**, 704–715 (2008).
37. Ullmann, P. et al. The miR-371–373 cluster represses colon cancer initiation and metastatic colonization by inhibiting the TGFBR2/ID1 signaling axis. *Cancer Res.* **78**, 3793–3808 (2018).
38. Fumagalli, A. et al. Plasticity of Lgr5-negative cancer cells drives metastasis in colorectal cancer. *Cell Stem Cell* **26**, 569–578.e7 (2020).
39. Dzhenkova, E. A. et al. Association between cancer stem cell marker ALDH1 and clinical and morphological factors of colorectal cancer prognosis. *J. Clin. Oncol.* **39**, 3522–3522 (2021).
40. Stanford, E. A. et al. The role of the aryl hydrocarbon receptor in the development of cells with the molecular and functional characteristics of cancer stem-like cells. *BMC Biol.* **14**, 20 (2016).
41. Larigot, L., Juricek, L., Dairou, J. & Coumou, X. AhR signaling pathways and regulatory functions. *Biochim. Open* **7**, 1–9 (2018).
42. Tomkovich, S. et al. Locoregional effects of microbiota in a preclinical model of colon carcinogenesis. *Cancer Res.* **77**, 2620–2632 (2017).
43. Chen, Y. et al. *Fusobacterium nucleatum* promotes metastasis in colorectal cancer by activating autophagy signaling via the upregulation of CARD3 expression. *Theranostics* **10**, 323–339 (2020).
44. Chen, S. et al. *Fusobacterium nucleatum* promotes colorectal cancer metastasis by modulating KRT7-AS/KRT7. *Gut Microbes* **11**, 511–525 (2020).
45. Parhi, L. et al. Breast cancer colonization by *Fusobacterium nucleatum* accelerates tumor growth and metastatic progression. *Nat. Commun.* **11**, 3259 (2020).
46. Jia, Y. et al. TLR2/TLR4 activation induces Tregs and suppresses intestinal inflammation caused by *Fusobacterium nucleatum* in vivo. *PLoS ONE* **12**, e0186179 (2017).
47. Kostic, A. D. et al. *Fusobacterium nucleatum* potentiates intestinal tumorigenesis and modulates the tumor-immune microenvironment. *Cell Host Microbe* **14**, 207–215 (2013).
48. Saito, T. et al. Two FOXP3(+)/CD4(+) T cell subpopulations distinctly control the prognosis of colorectal cancers. *Nat. Med.* **22**, 679–684 (2016).
49. Chaushu, S. et al. Direct recognition of *Fusobacterium nucleatum* by the NK cell natural cytotoxicity receptor NKP46 aggravates periodontal disease. *PLoS Pathog.* **8**, e1002601 (2012).
50. Yang, L. et al. Metabolic shifts toward glutamine regulate tumor growth, invasion and bioenergetics in ovarian cancer. *Mol. Syst. Biol.* **10**, 728 (2014).
51. Yu, T. et al. *Fusobacterium nucleatum* promotes chemoresistance to colorectal cancer by modulating autophagy. *Cell* **170**, 548–563.e16 (2017).
52. Johnson, C. H. et al. Metabolism links bacterial biofilms and colon carcinogenesis. *Cell Metab.* **21**, 891–897 (2015).
53. Zeller, G. et al. Potential of fecal microbiota for early-stage detection of colorectal cancer. *Mol. Syst. Biol.* **10**, 766 (2014).
54. Hertel, J., Heinken, A., Martinelli, F. & Thiele, I. Integration of constraint-based modeling with fecal metabolomics reveals large deleterious effects of *Fusobacterium* spp. on community butyrate production. *Gut Microbes* **13**, 1–23 (2021).
55. Zhao, Z. et al. Metagenome association study of the gut microbiome revealed biomarkers linked to chemotherapy outcomes in locally advanced and advanced lung cancer. *Thorac. Cancer* **12**, 66–78 (2021).
56. Viaud, S. et al. The intestinal microbiota modulates the anticancer immune effects of cyclophosphamide. *Science* **342**, 971–976 (2013).
57. Iida, N. et al. Commensal bacteria control cancer response to therapy by modulating the tumor microenvironment. *Science* **342**, 967–970 (2013).
58. Yu, S. Y. et al. *Bacteroides faecalis* sp. nov., isolated from human faeces. *Int. J. Syst. Evol. Microbiol.* **69**, 3824–3829 (2019).
59. Mendes, R. T. et al. Endothelial cell response to *Fusobacterium nucleatum*. *Infect. Immun.* **84**, 2141–2148 (2016).
60. Fardini, Y. et al. *Fusobacterium nucleatum* adhesin FadA binds vascular endothelial cadherin and alters endothelial integrity. *Mol. Microbiol.* **82**, 1468–1480 (2011).
61. Nagy, J. A., Benjamin, L., Zeng, H., Dvorak, A. M. & Dvorak, H. F. Vascular permeability, vascular hyperpermeability and angiogenesis. *Angiogenesis* **11**, 109–119 (2008).
62. Agarwal, S. et al. MTHFD1L, a folate cycle enzyme, is involved in progression of colorectal cancer. *Transl. Oncol.* **12**, 1461–1467 (2019).
63. Catlett, J. L. et al. Metabolic feedback inhibition influences metabolite secretion by the human gut symbiont *Bacteroides thetaiotaomicron*. *mSystems* **5**, e00252–20 (2020).
64. Jia, Q. et al. Fecal microbiota of diarrhea-predominant irritable bowel nm syndrome patients causes hepatic inflammation of germ-free rats and berberine reverses it partially. *BioMed Res. Int.* <https://doi.org/10.1155/2019/4530203> (2019).
65. Kaiko, G. E. et al. The colonic crypt protects stem cells from microbiota-derived metabolites. *Cell* **165**, 1708–1720 (2016).
66. Marinelli, L. et al. Identification of the novel role of butyrate as AhR ligand in human intestinal epithelial cells. *Sci. Rep.* **9**, 643 (2019).
67. Lazarova, D. L., Chiaro, C. & Bordonaro, M. Butyrate induced changes in Wnt-signaling specific gene expression in colorectal cancer cells. *BMC Res. Notes* **7**, 226 (2014).
68. Greenhalgh, K. et al. Integrated *in vitro* and *in silico* modeling delineates the molecular effects of a synbiotic regimen on colorectal-cancer-derived cells. *Cell Rep.* **27**, 1621–1632 (2019).
69. Veldhoen, M. et al. The aryl hydrocarbon receptor links TH17-cell-mediated autoimmunity to environmental toxins. *Nature* **453**, 106–109 (2008).
70. Paris, A., Tardif, N., Galibert, M.-D. & Corre, S. AhR and cancer: from gene profiling to targeted therapy. *Int. J. Mol. Sci.* **22**, E752 (2021).
71. Kramer, N. et al. Autocrine WNT2 signaling in fibroblasts promotes colorectal cancer progression. *Oncogene* **36**, 5460–5472 (2017).
72. Hiller, K. et al. MetaboliteDetector: comprehensive analysis tool for targeted and nontargeted GC/MS based metabolome analysis. *Anal. Chem.* **81**, 3429–3439 (2009).
73. Sullivan, M. R., Lewis, C. A. & Muir, A. Isolation and quantification of metabolite levels in murine tumor interstitial fluid by LC/MS. *Bio-Protoc.* **9**, e3427 (2019).
74. Noronha, A. et al. The Virtual Metabolic Human database: integrating human and gut microbiome metabolism with nutrition and disease. *Nucleic Acids Res.* **47**, D614–D624 (2019).
75. Heirendt, L. et al. Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. *Nat. Protoc.* **14**, 639–702 (2019).
76. Baldini, F. et al. The microbiome modeling toolbox: from microbial interactions to personalized microbial communities. *Bioinforma* **35**, 2332–2334 (2019).

77. Dhayade, S. et al. Impact of formate supplementation on body weight and plasma amino acids. *Nutrients* **12**, 2181 (2020).
78. Eide, P. W., Bruun, J., Lothe, R. A. & Sveen, A. CMScaller: an R package for consensus molecular subtyping of colorectal cancer pre-clinical models. *Sci. Rep.* **7**, 16618 (2017).
79. Varrette, S., Bouvry, P., Cartiaux, H. & Georgatos, F. Management of an academic HPC cluster: the UL experience, in *Proc. 2014 International Conference on High Performance Computing Simulation (HPCS)* 959–967 (Institute of Electrical and Electronics Engineers, 2014).

Acknowledgements

We thank the patients who kindly donated their samples and made this study possible. We thank members of the Dolznig laboratory (Medical University of Vienna) for the 7TGP-RKO cell line and culturing advice. We also thank D. Kuhn, P. May, C. Martin-Gallausiaux, K. Greenhalgh and T. Bintener for their help. We thank the LCSB Metabolomics Platform, especially C. Jäger, X. Dong and F. Vanhalle, for providing technical and analytical support. We thank all the contributing surgeons and nurses from the Centre Hospitalier Emile Mayrisch, the Centre Hospitalier du Luxembourg, the Zitha Klinik and the Clinical and Epidemiological Investigation Centre of the Luxembourg Institute of Health for their work with the patients. We thank the Fondation Cancer for its support during the setup of the Luxembourgish patients with CRC cohort. We also thank M. Mittelbronn and the pathologists and macroscopy team from NCP/LNS. We also thank their collaborators at the IBBL, F. Betsou, N. Goncharenko, C. Bahlawane and A. Gaignaux for the overall setup of the patient sample collection and the management of the cohort. We are also grateful to D. Coowar for managing the animal facility as well as the veterinarian service (M. Schmit and J. Behm) of the University of Luxembourg for their assistance and guidance in animal welfare during animal experimentation. The experiments presented in this paper were carried out using the high-performance computing facilities at the University of Luxembourg (ref. ⁷⁹, <https://hpc.uni.lu>). Biorender (BioRender.com) was used for graphical illustration. Funding This work was supported by the Luxembourg National Research Fund (FNR) (grant nos. CORE/C16/BM/11282028 (E.L.), PoC/18/12554295 (E.L.), PRIDE17/11823097 (M.T.), PRIDE19/14254520 (S.A.) and CORE/15/BM/10404093 (P.W.), ATTRACT Programme grant (nos. FNR/A12/01 to I.T. and A18/BM/11809970 to J.M.)), by the Luxembourg National Research Fund and the Fondation Cancer Luxembourg (grant no. CORE/C20/BM/14591557 (E.L.)) as well as by the Fondation du Pélican de Mie and Pierre Hippert-Faber under the aegis of the Fondation de Luxembourg ('Pelican Grant'; D.T., M.T. and J.K.), a FNRS-Télévie grant to M.M., no. 7.4565.21-40007364), an Internal Research Project at the University of Luxembourg (MiDiCa—integrated analysis of the effects of microbiome-diet interactions on human colorectal adenocarcinoma enterocytes; E.L., P.W., S.H. and D.T.), the Fondation Cancer and the Fondation Kriibskrank Kanner Luxembourg ('Effects of pesticides on the gut and their role in cancer development, PestGut', V.I.P. and S.H.), the Action LIONS Vaincre le Cancer Luxembourg and a European Research Council grant under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 757922 to I.T. and no. 863664 to P.W.). This project was also supported by the Doctoral School in Science and Engineering (D.T., J.K. and M.T.) and the Department of Life Sciences and

Medicine at the University of Luxembourg. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions

Conceptualization was done by D.T., M.T., V.I.P., P.W. and E.L. Methodology was devised by D.T., V.I.P., M.T., M.M., S.A., J.M., P.W. and E.L. Clinical methodology was developed by P.M., N.Z., J.Y.F. and M.G. Formal analysis was carried out by D.T., M.T., V.I.P., M.M., A. Gaigneaux, A. Ginolhac, E.K., A.H. and J.M. The investigation was carried out by D.T., V.I.P., M.T., M.M., S.A., J.K., L.G., M. Schmitz, M. Schmoetten, F.R., C.D., T.D.N., A.F.-B. and C.M.-G. Resources were obtained by N.Z., J.Y.F., M.G., L.N.-C., T.D.N. and D.J.S. Data curation was done by D.T., E.K., A. Gaigneaux and A. Ginolhac. Visualization was done by D.T., M.T., V.I.P., M.M., E.K. and S.A. The original draft was written by D.T. and E.L. Review and editing of the draft was done by D.T., E.L., M.T., T.S., M.M., V.I.P., E.K., M.P., A.H., I.T., J.M., A.F.-B., C.M.-G., S.H. and P.W. Project administration was carried out by E.L. Funding was acquired by, J.M., I.T., P.W. and E.L. Finally, S.H., I.T., P.W. and E.L. supervised the project.

Competing interests

The authors declare no competing interests.

Additional information

Extended data are available for this paper at <https://doi.org/10.1038/s42255-022-00558-0>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s42255-022-00558-0>.

Correspondence and requests for materials should be addressed to Elisabeth Letellier.

Peer review information *Nature Metabolism* thanks Livio Trusolino and the other, anonymous, reviewers for their contribution to the peer review of this work. Primary Handling Editor: Ashley Castellanos-Jankiewicz.

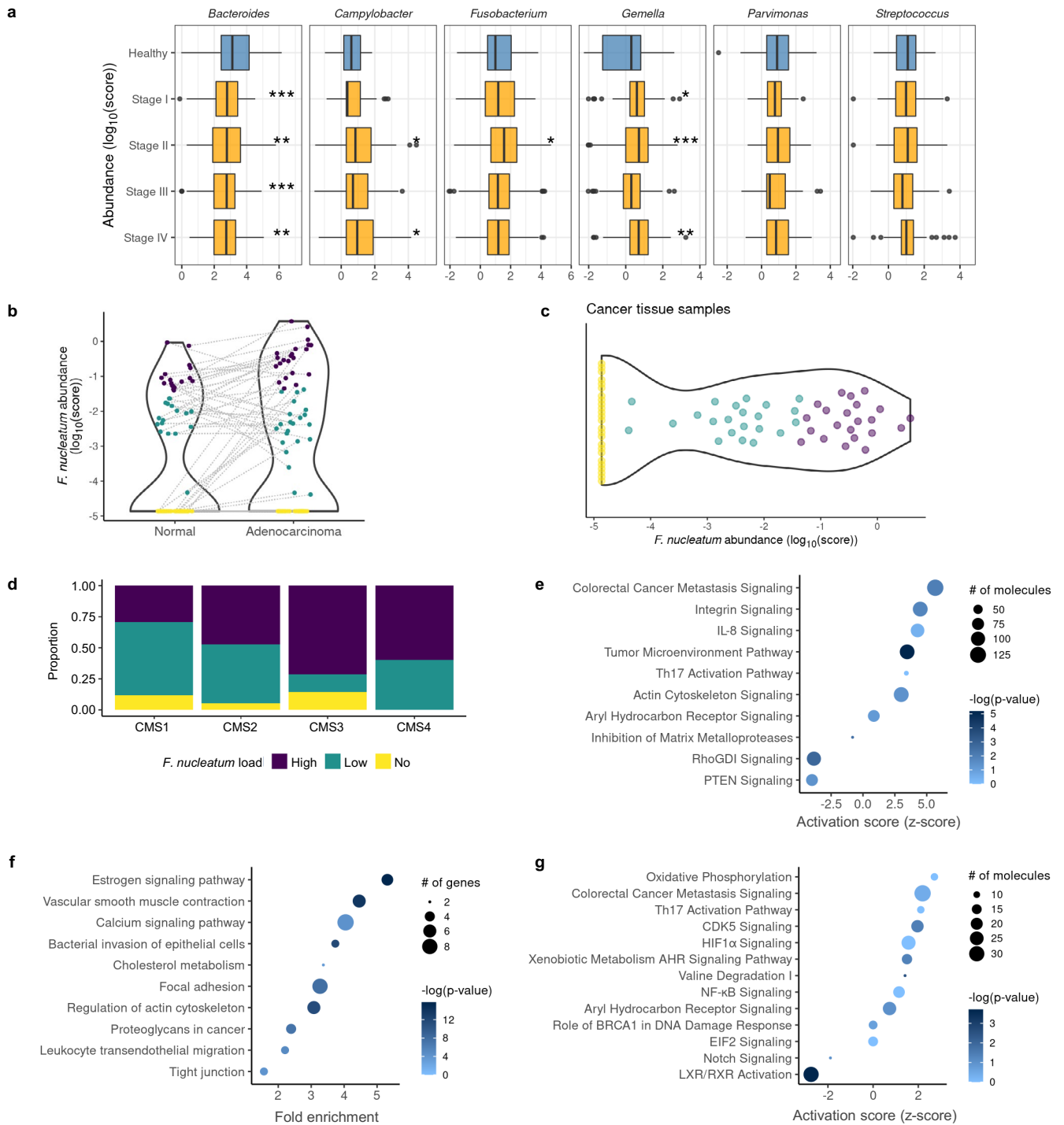
Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



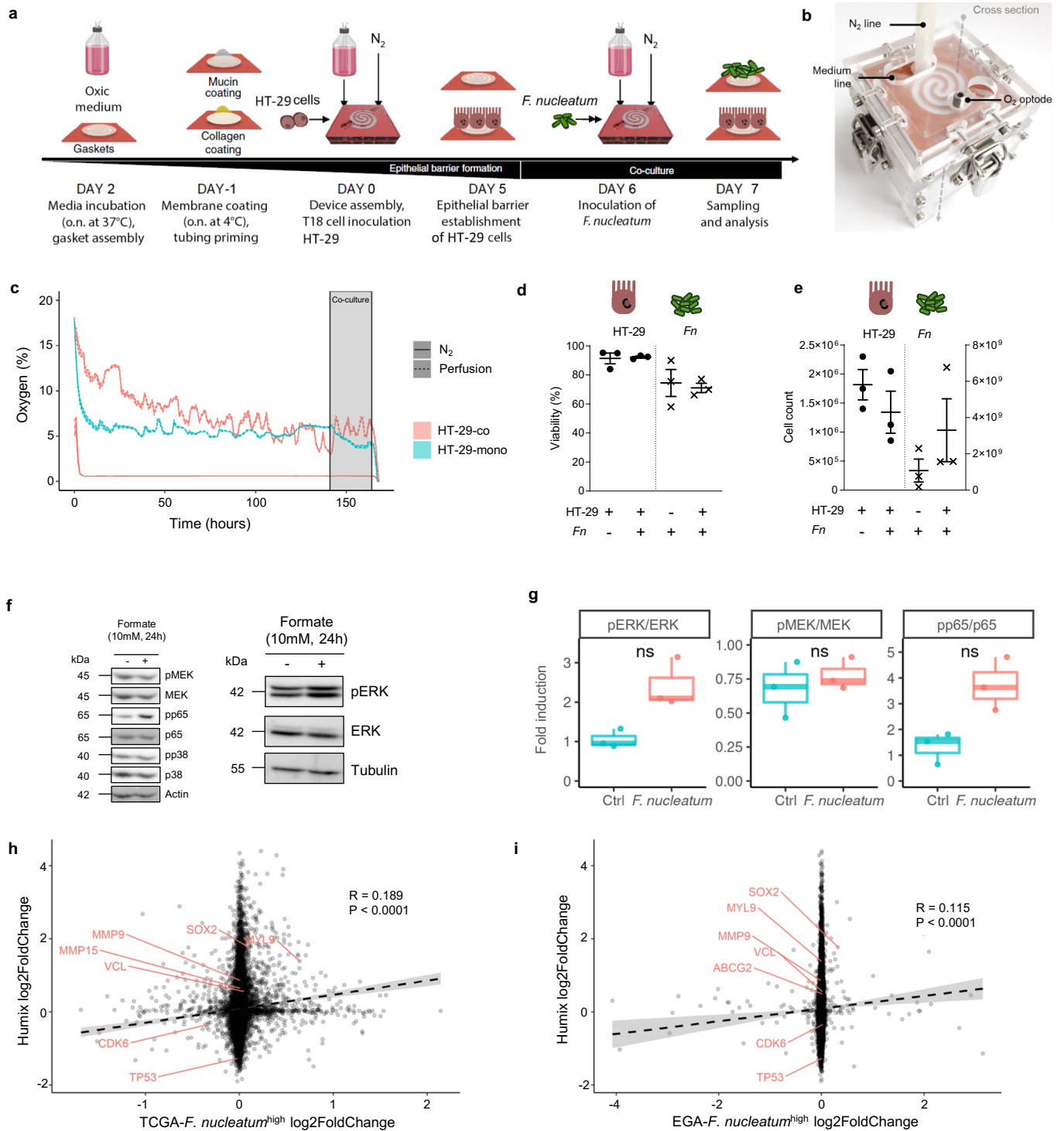
Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022, corrected publication 2023



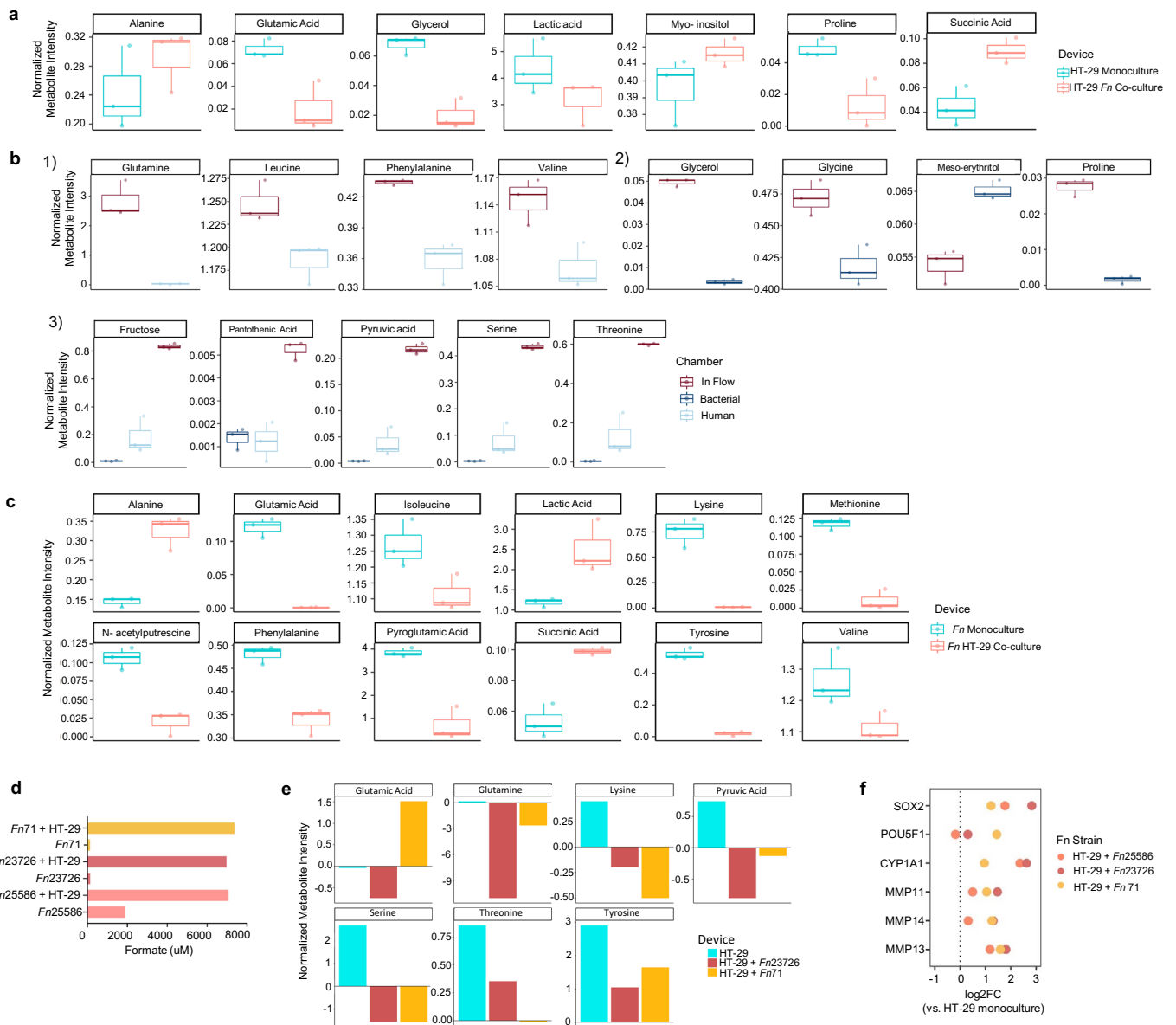
Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | Microbiome composition in CRC patients and pathways identified in *Fusobacterium*^{high} patients. **a**, Abundance of top CRC-enriched and differentially abundant bacteria in tissue per stage (TCGA). Boxplots shows medians with 1st and 3rd quantiles. The whiskers from the hinges to the smallest/largest values represent 1.5*inter-quartile range (IQR). n=108 healthy, n=102 Stage I, n= 209 Stage II, n=163 Stage III and n=86 Stage IV biologically independent donor samples, p=0.000705, p=0.00101, p=0.0000154 and p=0.00153 for *Bacteroides* in Stages I-IV vs. Healthy respectively, p=0.0112 and p=0.0402 for *Campylobacter* in Stages II and IV vs. Healthy respectively, p=0.0223 for *Fusobacterium* in Stage II vs Healthy, p=0.0293, p=0.000608 and p=0.00267 for *Gemella* in Stages I, II and IV vs. Healthy respectively, pairwise t test. **b**, *Fn* abundance in the EGA cohort. PathSeq-analysis for *Fn* abundance in matching normal vs. adenocarcinoma tissue on RNA-seq data, p=0.00165, paired two-tailed t test. **c**, *Fn* tissue abundance distribution in CRC patients of the EGA cohort. *Fn* was detected via RNA-seq and analyzed for the logarithmic score bacterial distribution (n=69). Quantile-based classification (color code) was applied for the target bacterium (quantcut function from the *gtools* R package). **d**, Correlation of *Fn* abundance with consensus molecular subtypes in CRC. The cohort in **c** was subjected to gene expression analysis and further classified via the *CMScaller* R package into CMS as described⁶⁷. Colored, segregated bars show the proportion of patients with differing fusobacterial loads per CMS. Chi-squared tests were performed for comparing the fusobacterial across all CMS. No significant differences were observed. n^{CMS1}=17, n^{CMS2}=19, n^{CMS3}=7, n^{CMS4}=5. **e**, IPA analysis of *Fusobacterium*^{high vs. no} differential gene expression analysis of the TCGA dataset. Plot shows z-scores, p-values, and the number of molecules per pathway. Selected significant pathways are shown (-log(p-value)>1.3). **f**, KEGG-based GSEA of *Fusobacterium*^{high vs. no} differential gene expression analysis of the TCGA dataset (pathfindR R package). Plot shows fold enrichment, p-values, and the number of genes per pathway. All significant pathways are shown (p<0.05). **g**, IPA analysis of *Fusobacterium*^{high vs. no} differential gene expression analysis of the EGA dataset. Plot shows z-scores, p-values, and the number of enriched molecules per pathway. Selected significant pathways are shown (-log(p-value)>1.3). *p<0.05, **p<0.01, ***p<0.001.

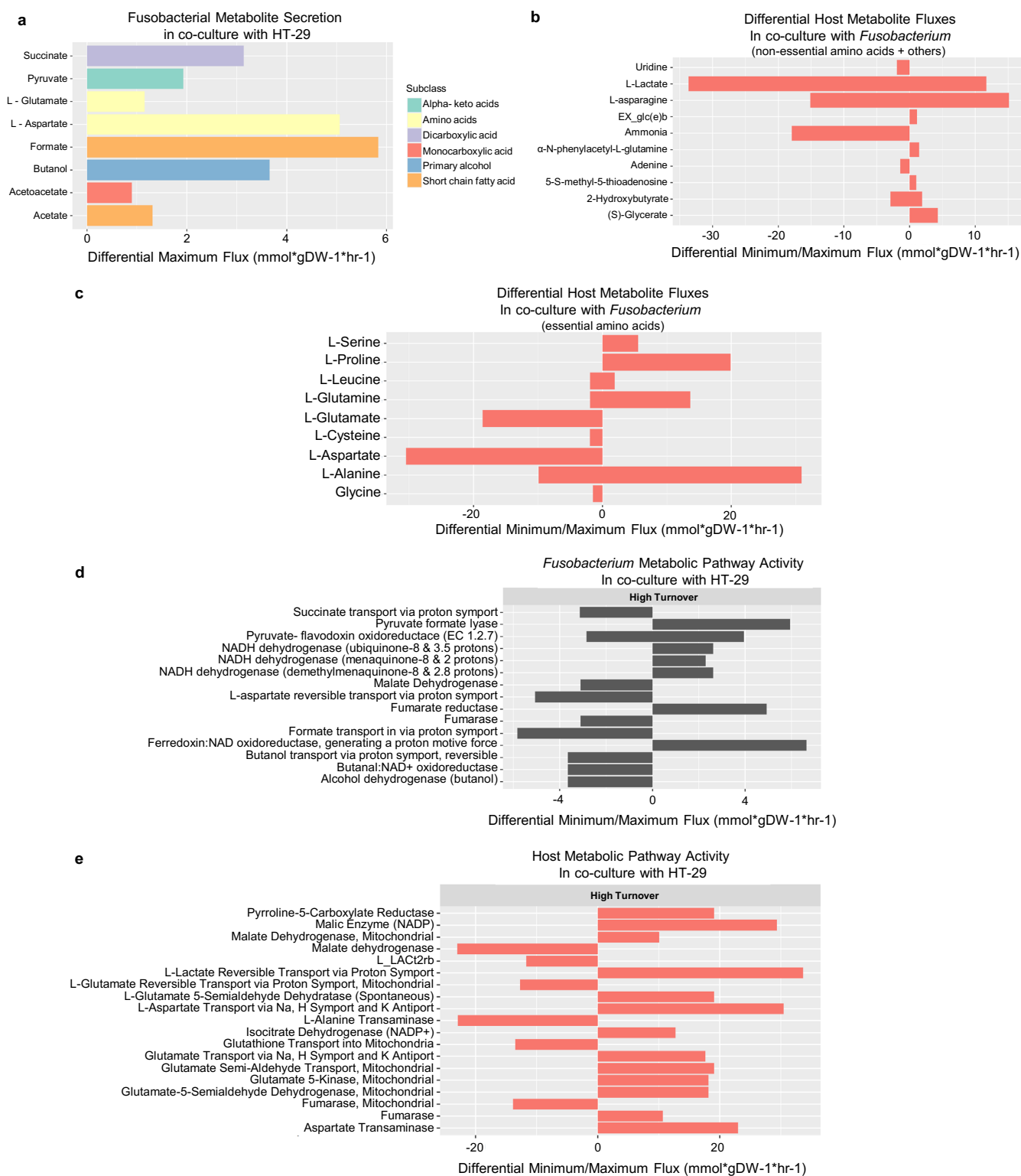


Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | Microbiome cross-talk studies in HuMiX. **a**, Experiment layout of HuMiX experiments. Devices assembly and medium priming for 1 day; epithelial cell inoculation and monolayer formation for 6 days; bacterial-human coculture for 24h. Control devices with either bacterial or human cell mono-cultures were run in parallel. At the endpoint, devices were opened for chamber metabolite sampling and cell harvesting. **b**, Clamped HuMiX device, showing the HuMiX device (spiral), the N₂ and medium lines as well as free optode pockets (cf. Figure 2A). **c**, Oxygen levels (%) as monitored in the perfusion and N₂ gas microchambers of HuMiX HT-29 mono-/ and cocultures. The grey box marks the the co-culture time. **d**, *Fn* and HT-29 cell viability in HuMiX mono-/ and co-cultures. Human cell viability was assessed by TrypanBlue staining and FACS-based L/D NearIR staining. Bacterial cell viability was assessed by BacLight L/D staining under a microscope. Three representative images were taken and cell viability was assessed with ImageJ. **(c)**, **(d)** and **(e)** show data from n=three biologically independent experiments. **e**, Cell counts of human cells and bacteria under co-culture. Cell counts of HT-29 cells were taken after TrypanBlue staining. Bacteria cell counts were assessed using CASY cell counter with the 45µm capillary. **f**, Western Blot detection of MAPK signaling, as well as NF-κB signaling on HuMiX HT-29 cell lysates. phosphorylation of MEK at Ser217/221, p65 at Ser536, pERK at thr 202/tyr204, and p38 at thr180/tyr182 under HT-29 monoculture (-) and HT-29-*Fn* co-culture (+) conditions in HT-29 cells. Data shows one representative of three biologically independent experiments. **g**, Quantified ratios of phosphorylated protein per total target protein levels from **(f)**, n=three biologically independent experiments. All boxplots show medians with 1st and 3rd quantiles. The whiskers from the hinges to the smallest/largest values represent 1.5*inter-quartile range (IQR). ns=not significant; two-sided paired t-tests. **h-i**, Correlation of global transcriptomic profile of HuMiX co-cultures with *Fusobacterium*^{high vs. no} patient's gene expression profile of the TCGA **(h)** and EGA **(i)** datasets. Interesting gene targets are marked in red. Two-tailed Spearman's rho correlation testing was used to assess significance between the two data sets.

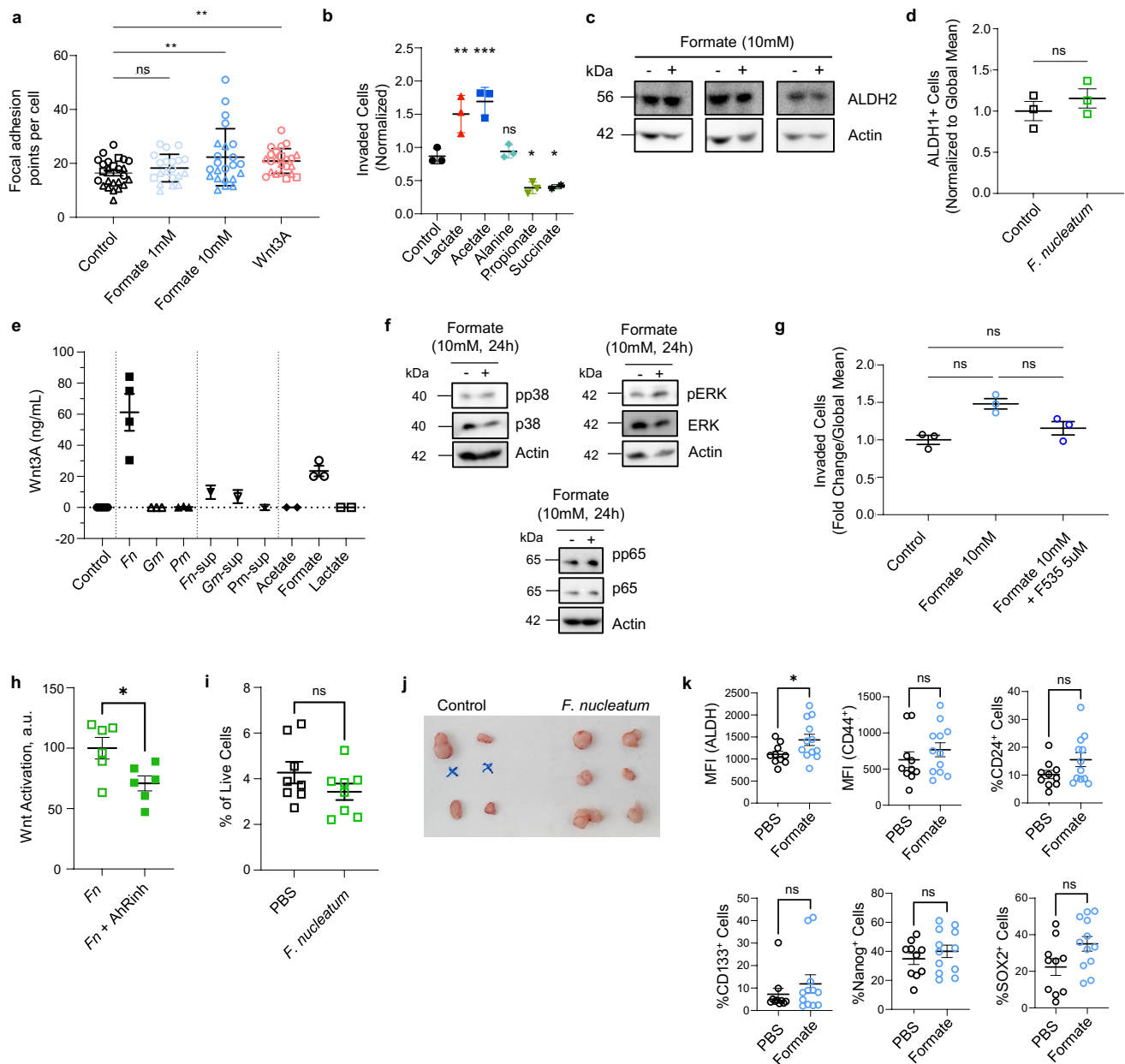


Extended Data Fig. 3 | Metabolome composition in HuMiX microbiome-host crosstalk. **a**, Significant differentially abundant metabolites in HT-29 mono-/ vs. HT-29-co-cultures, human cell microchamber supernatants. **b**, Significant differentially abundant metabolites in (1) HT-29-co-cultures comparing inflow metabolite levels with human chamber metabolite levels, (2) *Fn*-co-cultures, comparing inflow metabolite levels with bacterial microchamber supernatants. (3), HT-29-co-cultures and *Fn*-co-cultures, comparing inflow metabolite levels with human and bacterial microchamber supernatants. These were ambiguous metabolites that could not be assigned to any of the two cell types as being produced or secreted. **c**, Significant differentially abundant metabolites in *Fn* mono-/ vs. *Fn*-co-cultures, bacteria microchamber supernatants. **a**, **b** and **c** show data from $n=$ three biologically independent experiments. All boxplots (in **a**, **b** and **c**) show normalized median metabolite levels (log intensities). Boxplots show medians with 1st and 3rd quantiles. The whiskers from the hinges to the smallest/largest values represent 1.5*inter-quartile range (IQR). $p < 0.05$ for **a**, **b**, and **c**; unpaired, two-sided t -test. **d**, Levels of formate in different *Fn* isolates (*Fusobacterium nucleatum* 23726 and *Fusobacterium nucleatum* ssp. animalis 7_1, clinical isolate, one chamber per isolate) in monoculture and co-cultures with T 18 cells in HuMiX. **e**, Levels of metabolites in HT-29 cells alone or in co-culture with different *Fn* isolates (*Fusobacterium nucleatum* 23726 and *Fusobacterium nucleatum* ssp. animalis 7_1, clinical isolate, one chamber per isolate). **f**, Genes related to AhR signaling and cancer stemness in HT-29 cells after co-culture with different *Fn* isolates in HuMiX.



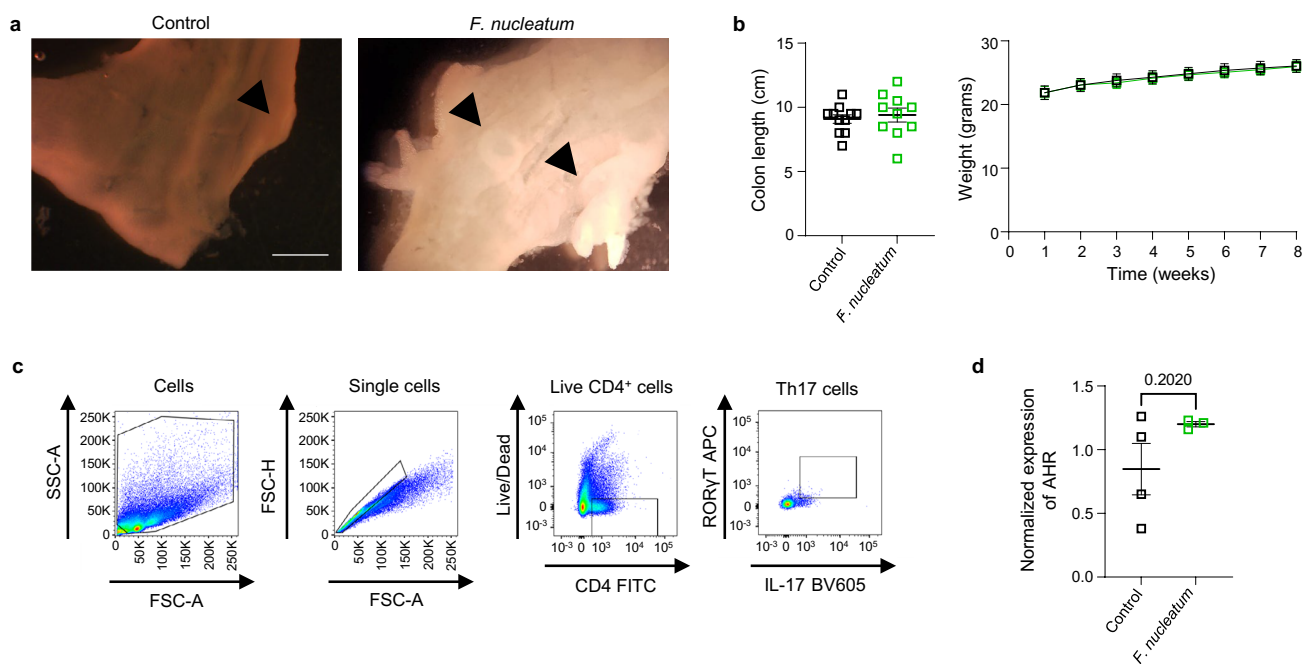
Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | *In silico* metabolic modelling of Fn-host tumor cell interaction. **a**, Maximum flux changes for exchange reactions of *F. nucleatum* secreted metabolites. Calculated flux ranges from the FVA results of the *F. nucleatum* monoculture model were subtracted from FVA results of the HT-29-*Fn* co-culture model. No minimum flux changes were observed. **b**, Minimum and maximum flux changes for exchange reactions of HT-29 secreted/uptaken non-essential amino acids and other metabolites. Calculated flux ranges from the FVA results of the HT-29 mono-culture model were subtracted from flux ranges of the HT-29-*Fn* co-culture model. **c**, Minimum and maximum flux changes for exchange reactions secreted/uptaken essential amino acids in the HT-29 model. Calculated flux ranges from the FVA results of the HT-29 mono-culture model were subtracted from flux ranges of the HT-29-*Fn* co-culture model. **d**, Minimum and maximum flux changes for intracellular reactions of *F. nucleatum*. Calculated flux ranges from the FVA results of the *F. nucleatum* mono-culture model were subtracted from FVA results of the HT-29-*Fn* co-culture model. Cutoff for displayed reactions was -2 and 2 mmol*dGW⁻¹*hr⁻¹. **e**, Minimum and maximum flux changes for intracellular reactions of HT-29. Calculated flux ranges from the FVA results of the HT-29 mono-culture model were subtracted from flux ranges of the HT-29-*Fn* co-culture model. Cutoff for displayed reactions was -10 and 10 mmol*dGW⁻¹*hr⁻¹.

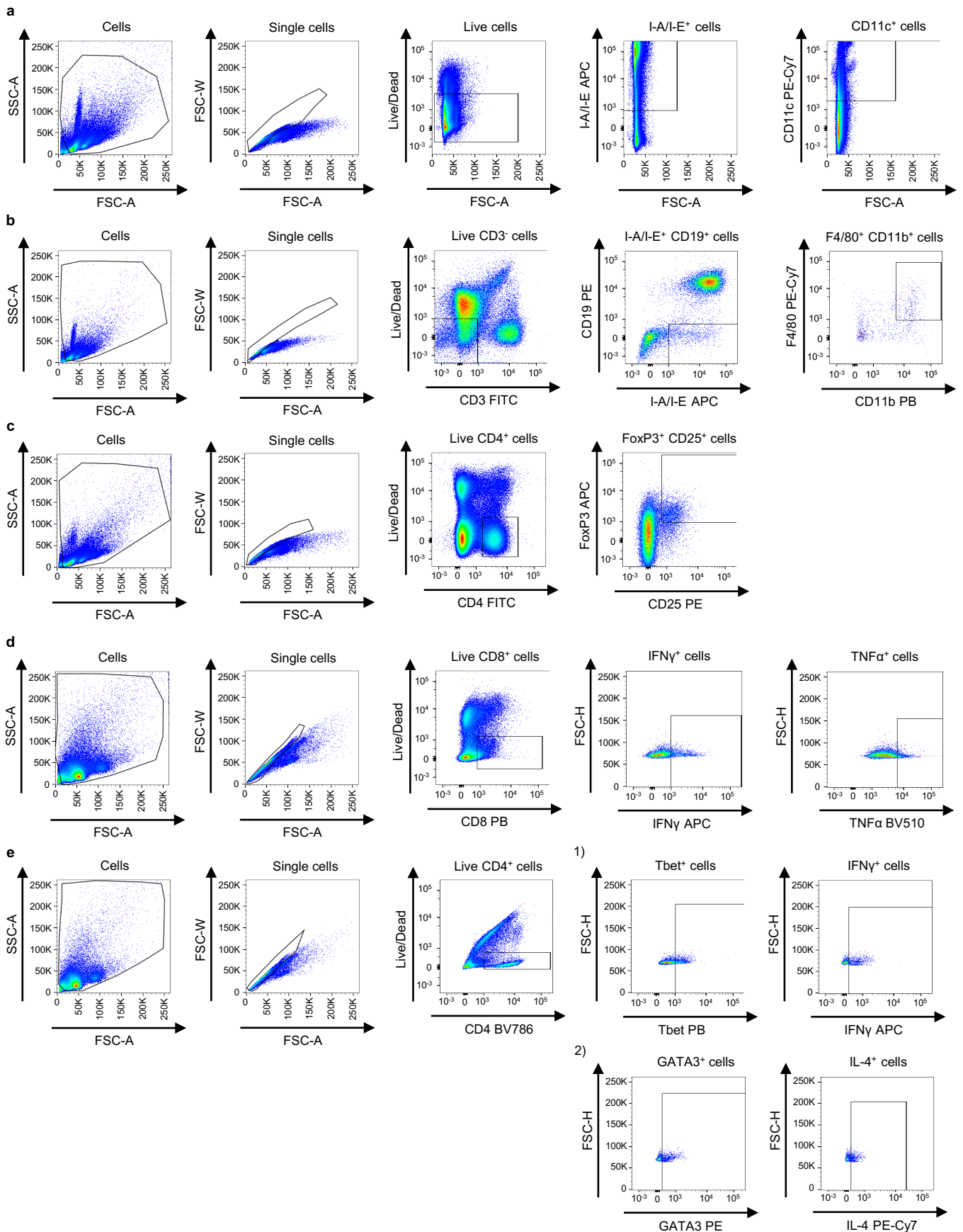


Extended Data Fig. 5 | See next page for caption.

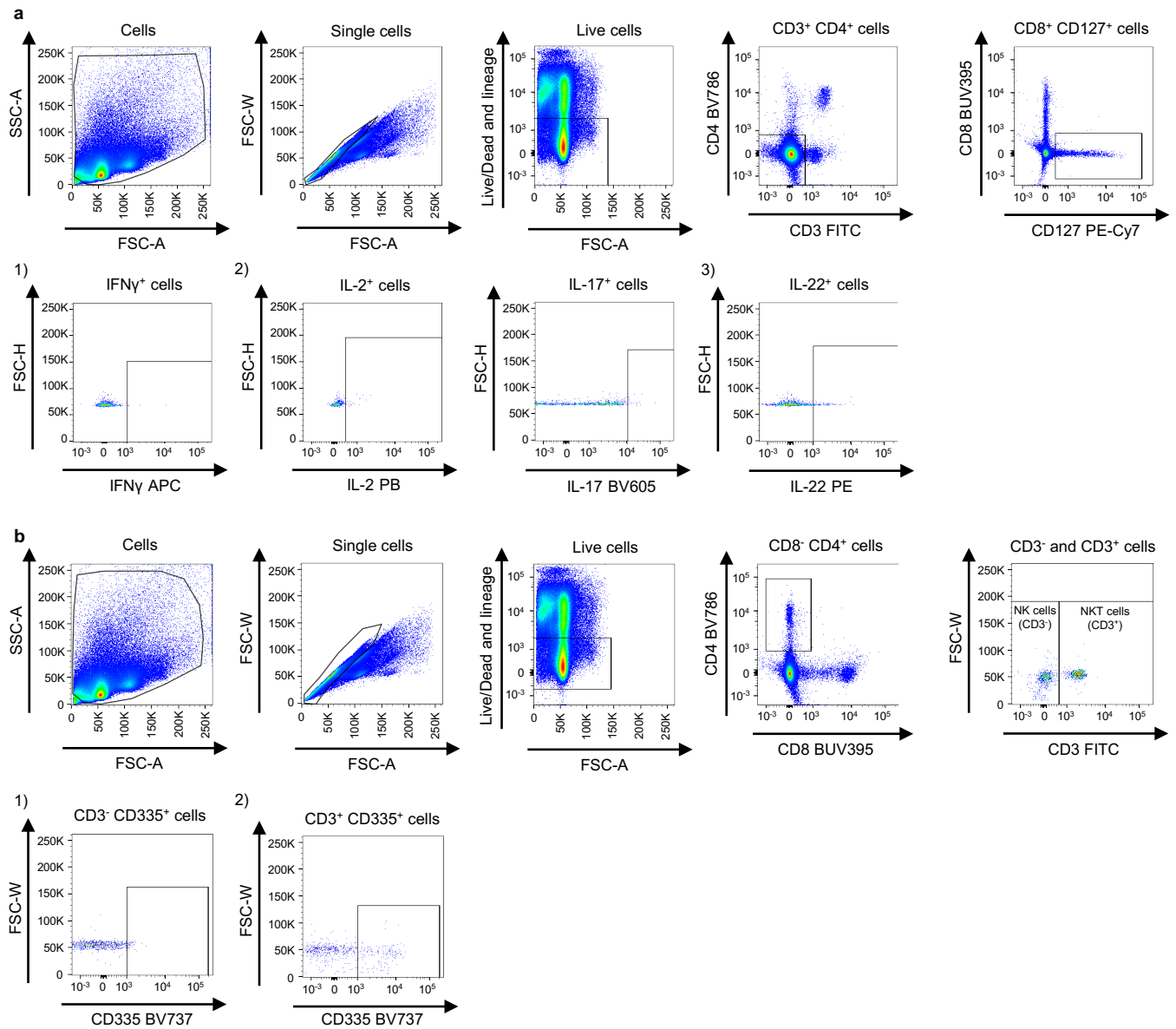
Extended Data Fig. 5 | *Fusobacterium nucleatum* and formate induced pathways. **a**, Focal adhesion formation in HT-29 cells after formate or rhWnt3A (100ng/mL) exposure for two days. Data shows technical replicates with at least 100 cells per condition from three independent experiments (different shapes), $p=0.0029240$ and $p=0.0024919$ for formate 10mM and Wnt3A vs. control respectively, two-tailed nested ANOVA, with Tukey multiple comparison of means. **b**, Transwell invasion of HCT116 CRC cells after treatment with lactate (positive control), acetate, propionate (10mM), alanine and succinate (5mM), $n=$ three biologically independent experiments, each with three technical replicates, $ns=$ not significant, $p=0.0030$, $p=0.0004$, $p=0.0217$, $p=0.0479$ for lactate, acetate, propionate and succinate vs. control respectively, ordinary one-way ANOVA Dunnett testing. **c**, Protein levels of ALDH2 in formate-treated (+) and untreated (-) cells. **d**, ALDH activity in human colonic organoids after treatment with *Fn*. Data shows mean \pm SEM of $n=$ three biologically independent experiments ($ns=$ not significant). **e**, Wnt activity in 7TGO-RKO Wnt-reporter cells stimulated with bacteria *Fn*, *Gemella morbillorum* (*Gm*) and *Parvimonas micra* (*Pm*) or with 10% [v/v] cell-free bacterial culture supernatants (-sup) or with bacterial secretion products (10mM) for 1h. Bars represent mean \pm SD of $n=$ one (bacterial supernatants), two (lactate, acetate), three (*Gm*, *Pm*, formate, Wnt), four (*Fn*) or six (control) biological replicates. **f**, Protein levels of pERK/ERK, pp65/p65 and pp38/p38 in formate treated (+) and untreated (-) cells. **g**, Transwell invasion of HCT116 CRC cells after 48 hours of stimulation with formate alone or in presence of a Wnt-signaling inhibitor (F535). Data shows mean \pm SEM of $n=$ three biologically independent experiments. Control and formate condition alone are also shown in Fig. 4b. $ns=$ not significant, repeated measures one-way ANOVA. **h**, Wnt activity in *Fn*-supernatant treated 7TGP-RKO Wnt-reporter cells alone or with an AhR inhibitor. Data shows mean \pm SEM of two biologically independent experiments, $p=0.0228$. **i**, Stemness marker CD44 in mouse tumours after intratumoral injection of *Fn*. Data shows mean \pm SEM percentage of CD44⁺ live cells ($ns=$ not significant), $n=8$ independently treated tumors. **j**, Photograph of explanted subcutaneous xenografts from secondary recipient mice from Fig. 5f. **k**, Expression of stemness markers in formate-injected and control tumors, $n=10$ and $n=12$ independently treated tumors for the control and formate conditions respectively. Data shows mean \pm SEM for mean fluorescence intensity (MFI) of ALDH and CD44 and percentage of CD24, CD133, Nanog and Sox2, expressing cells. ($ns=$ not significant, $p=0.0460$ for ALDH); unpaired, two-sided t-test in (**d**), (**i**), (**j**), and (**l**). * $p<0.05$, ** $p<0.01$, *** $p<0.001$.



Extended Data Fig. 6 | *Fusobacterium nucleatum* and formate induced pathways. **a**, Representative images of open colons of PBS or *Fn* treated mice. Black arrows mark tumors, scalebar=5mm. **b**, Colon length in centimeters (measured upon resection, left) and mouse body weight over time in grams (right) of germ-free mice gavaged with *Fn* and control mice. Data shows mean \pm SEM, n=11 and n=10 biologically individual animals for the control and *Fn*-treated groups respectively, pooled from two independent experiments. **c**, Representative gating strategy of Th17 cells. Th17 cell gating was performed on live (APC-Cy7⁻) single cells with FITC(CD4)⁺, BV605(IL-17)⁺ and APC(ROR γ T)⁺ staining. **d**, Gene expression levels of *AHR* as assessed by RT qPCR in *Fn*-gavaged mice. Data shows mean \pm SEM of n=4 biological replicates in control and n=3 biological replicates in *Fn*, two-sided t-test.



Extended Data Fig. 7 | Flow cytometry gating strategies. **a**, Representative gating strategy of dendritic cells. **b**, Representative gating strategy of macrophages. **c**, Representative gating strategy of T regulatory cells. **d**, Representative gating strategy of CD8 cells. **e**, Representative gating strategy of T helper 1 (subpanel 1) and T helper 2 (subpanel 2) cells.



Extended Data Fig. 8 | Flow cytometry gating strategies. **a**, Representative gating strategy of innate lymphoid cells (ILC); IL1 (subpanel 1), IL2 (subpanel 2), IL3 (subpanel 3). **b**, Representative gating strategy of natural killer (NK, subpanel 1) and natural killer T (NKT, subpanel 2) cells.

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

GC-MS chromatograms were processed using Agilent MassHunter Quantitative Analysis for GC-MS, Version B.08.00
 LC-MS Data were acquired with Thermo Xcalibur software (version 4.3.73.11)
 WB: detection was performed with the Fusion FX imaging platform
 Flow Cytometric analysis was performed using Cantoll and a Fortessa FACS machine
 Incucyte Live-Cell Analysis System (Essen Bioscience)
 Real-Time PCR system (Applied Biosciences)
 Sequencing of the V3 and V4 regions of prokaryotic 16S rRNA gene was performed on the Illumina MeSeq platform using 2x300 bp paired-end reads.
 RNA sequencing: RNA library quantity and quality were assessed with the Agilent 2100 BioAnalyzer (Agilent, USA). RNA libraries were sequenced as 100bp paired-end runs on an Illumina HiSeq2500 platform.

Data analysis

Metabolite Detector software (version 3.220180913) used for mass spectrometric data post processing, quantification, MID calculations,
 Image studio lite software for image analysis (WB)
 Image J 1.53K
 MATLAB (R2018B)
 Cobra 3.0
 FlowJo software 10.6.1
 BD FACSDiva (8.0.1)
 Raw 16S rRNA gene sequences (FASTAQ files) were cleaned and clustered as operational taxonomic units (OTUs) using NgTax or the Dada2 pipeline.
 Cedex
 R studio (1.2.5042, Rstudio Team 2020) with R CRAN (4.0.3, R Core Team 2020)
 Incucyte ZOOM software (2018B, Sartorius, Germany)

RT-qPCR data was analyzed using qBase+ 3.2 (Biogazelle) according to MIQE guidelines
GraphPad Software version8
All scripts, codes and models can be found under https://gitlab.lcsb.uni.lu/mdm/ternes_et_al_2022/.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

All data that support the findings of this study are available in the source data provided with the manuscript. Additionally, sequence data from the EGA is available under accession number EGAS00001000288. Sequence data from the CRC patients and healthy donor samples can be found on ENA (EBI) under the accession number PRJEB46665. The TCGA analysis was based upon data (TCGA-COAD, TCGA-READ) generated the TCGA Research Network: <https://www.cancer.gov/tcga>. All scripts and models can be found under https://gitlab.lcsb.uni.lu/mdm/ternes_et_al_2022/. Microbial reconstructions are modified from the virtual Metabolic Human website (www.vmh.life) and available on gitlab.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	For in vitro studies a minimum of triplicates per condition were used and each experiment was verified in at least three independent experiments. For in vivo work, the number of mice per group was calculated based on a power calculation in liaison with a bio-statistician. Please refer to the Figure legends and Method sections for further details.
Data exclusions	no data was excluded unless clear technical issues were identified
Replication	All in vitro experiments were verified in at least three independent experiments with same overall outcome. Specific number of repetitions is provided in the manuscript. For in vivo groups, adequate group size was calculated beforehand. As stated in the manuscript, for some experiments, we pooled two independent experiments to reach the number calculated by the bioinformatician. All in vivo replication were successful.
Randomization	Before starting all of the in vivo experiments, mice were randomly allocated to different groups. All Mass Spec samples were randomized during the analysis to avoid bias from eg. instrument drift.
Blinding	Analyses of in vivo samples (tumor counts) were blinded and analyzed by two researchers independently. IDs were uncovered after final result was obtained. Results from both persons lead to the same overall result.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

- | | |
|-------------------------------------|---|
| n/a | Involved in the study |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> Antibodies |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> Eukaryotic cell lines |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Palaeontology and archaeology |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> Animals and other organisms |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> Human research participants |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Clinical data |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Dual use research of concern |

Methods

- | | |
|-------------------------------------|--|
| n/a | Involved in the study |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> ChIP-seq |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> Flow cytometry |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> MRI-based neuroimaging |

Antibodies

Antibodies used	Actin, OCT4, Tubulin, ALDH1A1, ALDH2, ERK, pERK1/2, MEK, pMEK, p38, pp38, p65, pp65, CD8a, CD4, CD3e, Nkp46, TNF α , CD127, IL-17a, IFN γ , D4, CD16/32, CD25, IL-4, CD19, CD86, CD11c, CD80, CD40, IL-22, IL-2, TER119, CD11b, Tbet, Gata3, MHC II, CD11b, CD86, CD19, CD5, Ly6G, ROR γ T, FoxP3, F4/80, Ahr, donkey anti-rabbit Alexa Fluor 647. For further information, please refer to Supplementary Table 5.
Validation	All used antibodies are well described and used across research laboratories as well as used in published manuscripts.

Eukaryotic cell lines

Policy information about [cell lines](#)

Cell line source(s)	The HT-29, HCT116 and CACO-2 commercial CRC cell lines were obtained from ATCC and maintained under the in DMEMF12 with 10% [v/v] FBS and 1% [v/v] penicillin/streptomycin. RKO-7TGP were received from the University of Vienna (Dolznig lab) and maintained as described in the manuscript.
Authentication	All cell lines except the primary cultures established have been received from ATCC and authenticated via STR profiling.
Mycoplasma contamination	Cells have been checked for mycoplasma every two weeks and all cell lines used in the manuscript were mycoplasma negative.
Commonly misidentified lines (See ICLAC register)	HCT116 are part of the list but have been authenticated via STR profiling at DSMZ before the study.

Animals and other organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research

Laboratory animals	mouse musculus C57Bl6N and NSG mice both male and female aged 11-15 and 7-8 respectively.
Wild animals	No wild animals were used
Field-collected samples	No field collected samples were used
Ethics oversight	All animal experiments were performed according to all applicable laws and regulations, after receiving approval by the institution's Animal Experimentation Ethics Committee at UL (AEEC) and the Ministry of Agriculture, Viniculture & Rural Development (LUPA 2019/99 and LUPA 2019/60). They ensure that care and use of animals for research purposes was conducted according to the EU Directive 2010/63/EU, as well as the Grand-Ducal Regulation of January 11, 2013 on the protection of animals used for scientific purposes. These included the justification of the use of animals, their welfare and the incorporation of the principles of the 3R's (Replacement, Reduction and Refinement). A biostatistician reviewed all animal protocols.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics	The average ages of stool sample donors were 70.29 +/- 8.74, with 12 female and 40 male subjects (in-house CRC collection) and 62.05 +/- 7.51, with 30 female and 33 male subjects (ND collection). All patients were enrolled between the years 2010 and 2019.
Recruitment	All CRC patients were enrolled at diagnosis by the Centre Hospitalier Emile Mayrisch (Esch-sur-Alzette, Luxembourg), the Centre Hospitalier de Luxembourg (Luxembourg, Luxembourg), or the Zitah Klinik (Luxembourg, Luxembourg). A healthy control cohort (ND collection) was enrolled by the National Centre of Excellence in Research on Parkinson's disease (NCER-PD) and 16S rRNA gene sequencing data was shared according to a biomaterial cooperation agreement.
Ethics oversight	Patient samples were donated willingly under informed consent and were handled in accordance with institutional guidelines. Ethical approval was obtained from the Comité National d'Ethique de Recherche, Luxembourg (Reference 201009/09) followed by institutional approval by the Ethics Review Panel of the University of Luxembourg (ERP-16-032).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Plots

Confirm that:

- The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- All plots are contour plots with outliers or pseudocolor plots.
- A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation

Organoids embedded in Matrigel were treated with 5mM sodium formate (Carl Roth, 4404.1), for 48h at 37°C. ALDH activity was assessed according to the manufacturer's protocol (ALDEFLUOR Kit, Stemcell Technologies, 01700). CD133 (Miltenyi Biotec, 130-090-826), CD44 (BD Biosciences, 555428) and CD24 (BD Biosciences, 560533) labeling was done at a 1:50 dilution. LIVE/DEAD cell viability staining was performed at a 1:1000 dilution (Invitrogen, L34975). Samples were acquired on a FACS Cantoll Cell Analyzer (BD Biosciences). Results were analyzed with the FlowJo software (BD Biosciences).

Xenografts were harvested at endpoint and dissociated (Human Tumor Dissociation Kit, Miltenyi) for further FACS-based protein expression analysis.

Colons were opened longitudinally, washed with cold PBS, cut into smaller sections and digested first in RPMI containing glutamine/GlutamMAX, 3% FBS, 1% P/S, 5mM EDTA and 0.154mg/mL DTT, then in serum-free RPMI with Liberase (0.1mg/mL) and DNase (0.5mg/mL), both in a shaking incubator at 800rpm and 37°C for 20min and 30min respectively. Spleens were digested in serum-free RPMI with Liberase (0.1mg/mL) and DNase (0.5mg/mL) for 30min. Digestions were quenched with RPMI medium containing 10% FBS and cell suspensions were filtered through a 70µm filter. Mesenteric lymph nodes were mashed through a 70µm filter. Single-cell suspensions were restimulated in RPMI containing 10% FBS, 0.1µg/mL PMA and 1.5µg/mL ionomycin for 4h at 37°C. Brefeldin A was added 1h after the beginning of the restimulation. Cells were stained with Near IR fixable live/dead dye and with the corresponding antibody panels (Supplementary Table 6) using the Cytotfix/Cytoperm kit (554714, BD Bioscience). An additional Fc blocking step preceded antibody staining for macrophages and dendritic cells. Acquisition was performed on a Cantoll and a Fortessa FACS machine and data was analyzed using BD FACSDiva 8.0.1 and FlowJo software 10.6.1 (gating strategies can be found in Extended Data Figures 7 and 8).

Instrument

BD FACS Cantoll and BD LSR Fortessa FACS machine

Software

BD FACS Diva 8.0.1 and FlowJo 10.6.1 softwares

Cell population abundance

Total cell counts were determined by Cedex and multiplied by frequency of subpopulations to get absolute counts.

Gating strategy

FSC/SSC selection of cells, exclusion of doublets by FSC-A/FSC-H
Unstained controls and single stained controls were acquired to identify negative and positive cell populations for each marker.
Detailed gating strategy is explained in the extended data figure 7-8

- Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.