

# Effect of the intratumoral microbiota on spatial and cellular heterogeneity in cancer

<https://doi.org/10.1038/s41586-022-05435-0>

Received: 9 March 2022

Accepted: 10 October 2022

Published online: 16 November 2022

Open access

 Check for updates

Jorge Luis Galeano Niño<sup>1</sup>, Hanrui Wu<sup>1,9</sup>, Kaitlyn D. LaCourse<sup>1,9</sup>, Andrew G. Kempchinsky<sup>1</sup>, Alexander Baryames<sup>1</sup>, Brittany Barber<sup>2</sup>, Neal Futran<sup>2</sup>, Jeffrey Houlton<sup>2,8</sup>, Cassie Sather<sup>3</sup>, Ewa Sicinska<sup>4</sup>, Alison Taylor<sup>5</sup>, Samuel S. Minot<sup>6</sup>, Christopher D. Johnston<sup>7</sup>✉ & Susan Bullman<sup>1</sup>✉

The tumour-associated microbiota is an intrinsic component of the tumour microenvironment across human cancer types<sup>1,2</sup>. Intratumoral host–microbiota studies have so far largely relied on bulk tissue analysis<sup>1–3</sup>, which obscures the spatial distribution and localized effect of the microbiota within tumours. Here, by applying in situ spatial-profiling technologies<sup>4</sup> and single-cell RNA sequencing<sup>5</sup> to oral squamous cell carcinoma and colorectal cancer, we reveal spatial, cellular and molecular host–microbe interactions. We adapted 10x Visium spatial transcriptomics to determine the identity and in situ location of intratumoral microbial communities within patient tissues. Using GeoMx digital spatial profiling<sup>6</sup>, we show that bacterial communities populate microniches that are less vascularized, highly immuno-suppressive and associated with malignant cells with lower levels of Ki-67 as compared to bacteria-negative tumour regions. We developed a single-cell RNA-sequencing method that we name INVADEseq (invasion–adhesion-directed expression sequencing) and, by applying this to patient tumours, identify cell-associated bacteria and the host cells with which they interact, as well as uncovering alterations in transcriptional pathways that are involved in inflammation, metastasis, cell dormancy and DNA repair. Through functional studies, we show that cancer cells that are infected with bacteria invade their surrounding environment as single cells and recruit myeloid cells to bacterial regions. Collectively, our data reveal that the distribution of the microbiota within a tumour is not random; instead, it is highly organized in microniches with immune and epithelial cell functions that promote cancer progression.

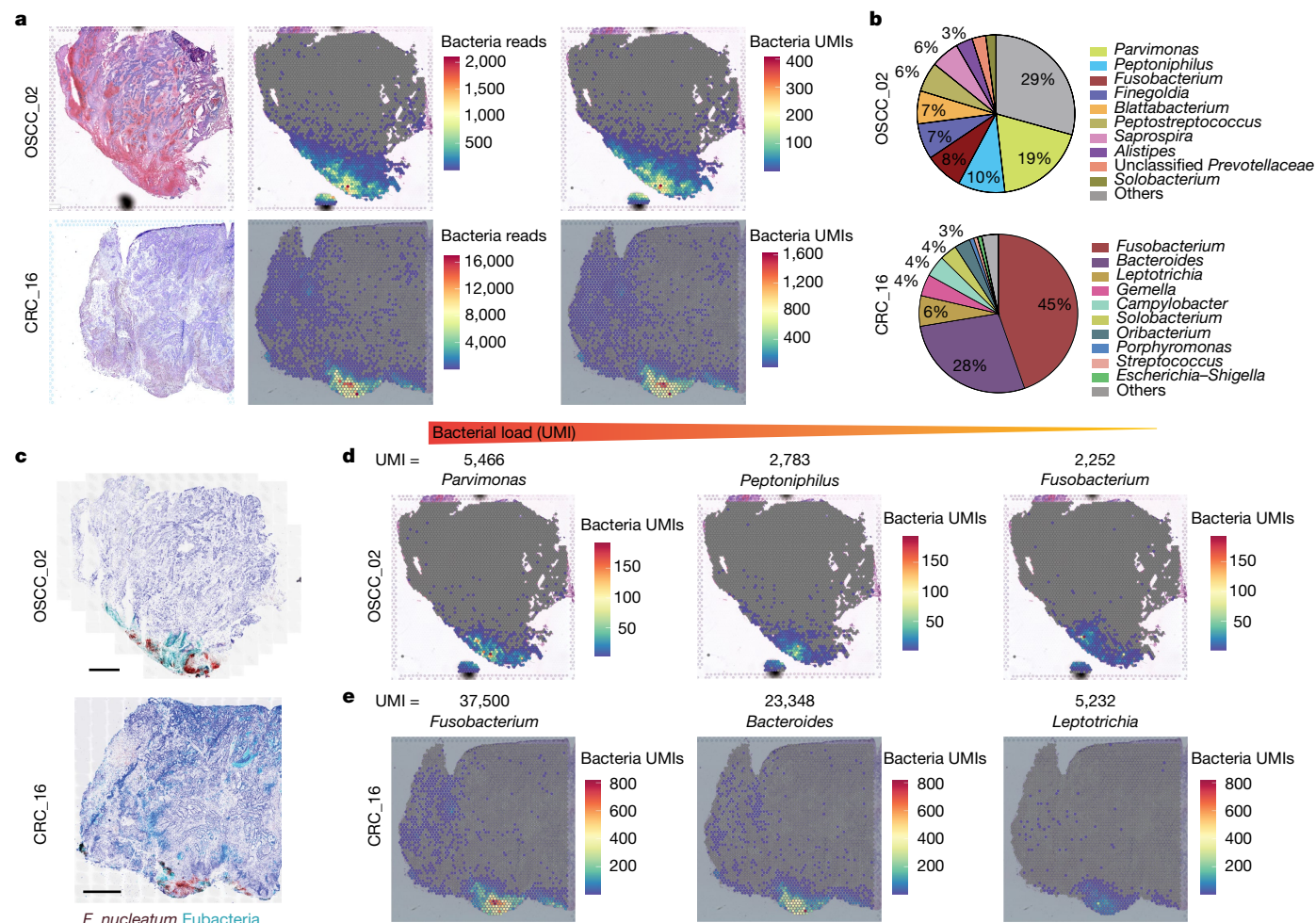
In the tumours of patients with cancer, malignant cells are surrounded by a complex network of non-malignant cells that may have pro- or anti-tumorigenic effects depending on their cell type and abundance. In vitro and preclinical animal models indicate that bacteria in the tumour-associated microbiota have a role in cancer development<sup>7</sup>, metastasis<sup>8–10</sup>, immunosurveillance<sup>11–13</sup> and chemoresistance<sup>14,15</sup>. There is strong molecular evidence of an intratumoral microbiota across at least 33 major cancer types<sup>2,12,13,16</sup>, as well as imaging data that show the co-localization of pan-bacterial markers with immune and epithelial cell targets, suggesting that the intratumoral microbiota can be intracellular<sup>2,8,13</sup>. However, the precise identity of these cell-associated organisms and the specific host cell types with which they interact in patient tumours have yet to be fully revealed. In addition, whether the spatial distribution of the intratumoral microbiota and specific host–microbial cellular interactions affect distinct functional capabilities within the tumour microenvironment (TME) is largely unknown. Here, focusing on cancers at the extremes of the gastrointestinal tract—oral squamous cell carcinoma (OSCC) and colorectal cancer (CRC)—we modify

in situ spatial-profiling technologies and single-cell RNA sequencing (scRNA-seq) to concurrently map host–bacterial spatial, cellular and molecular interactions within the TME. Our results reveal how the intratumoral microbiota contributes to tumour heterogeneity.

## Heterogeneity of the intratumoral microbiota

We performed 16S rRNA gene sequencing on 44 pieces of tissue from the tumours of 11 patients with CRC (Extended Data Fig. 1a), and observed that the composition of the intratumoral microbiota at the phylum and the genus level (Extended Data Fig. 1a and Supplementary Table 2), including *Fusobacterium* (Extended Data Fig. 1b), varied within individual patient tumours. Principal component analysis with beta diversity clustering (Extended Data Fig. 1c) and dendrogram analysis (Extended Data Fig. 1d) showed that over one third of the patients assessed ( $n = 4$  out of 11) had relatively stable microbiome compositions; however, most patients ( $n = 7$  out of 11) exhibited varying levels of heterogeneity in the intratumoral microbiome. This suggests a

<sup>1</sup>Human Biology Division, Fred Hutchinson Cancer Center, Seattle, WA, USA. <sup>2</sup>University of Washington Medical Center, Seattle, WA, USA. <sup>3</sup>Genomics Core, Fred Hutchinson Cancer Center, Seattle, WA, USA. <sup>4</sup>Department of Pathology, Dana-Farber Cancer Institute, Boston, MA, USA. <sup>5</sup>Herbert Irving Comprehensive Cancer Center, Columbia University, New York, NY, USA. <sup>6</sup>Data Core, Fred Hutchinson Cancer Center, Seattle, WA, USA. <sup>7</sup>Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Center, Seattle, WA, USA. <sup>8</sup>Present address: Head and Neck Specialists, Sarah Cannon Cancer Institute, Charleston, SC, USA. <sup>9</sup>These authors contributed equally: Hanrui Wu, Kaitlyn D. LaCourse. ✉e-mail: [johnston@fredhutch.org](mailto:johnston@fredhutch.org); [sbullman@fredhutch.org](mailto:sbullman@fredhutch.org)



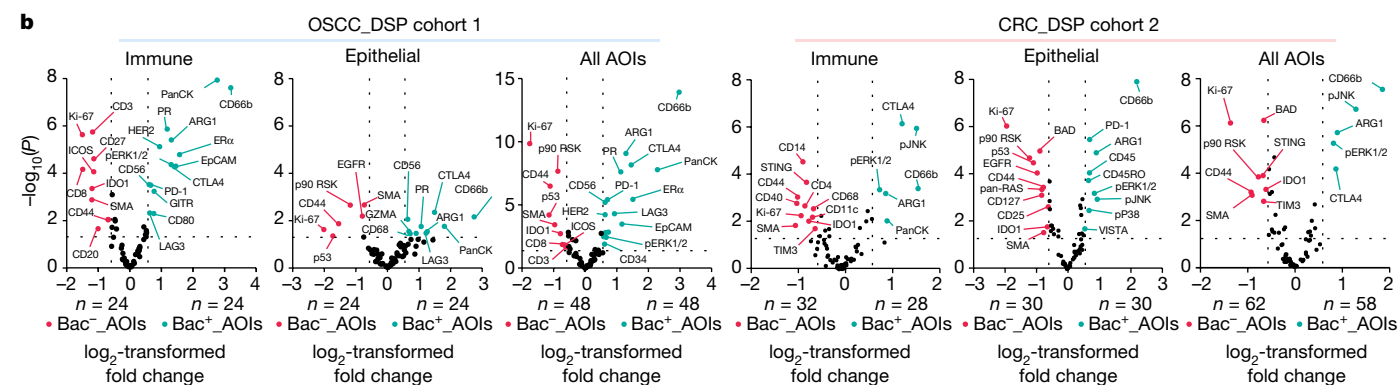
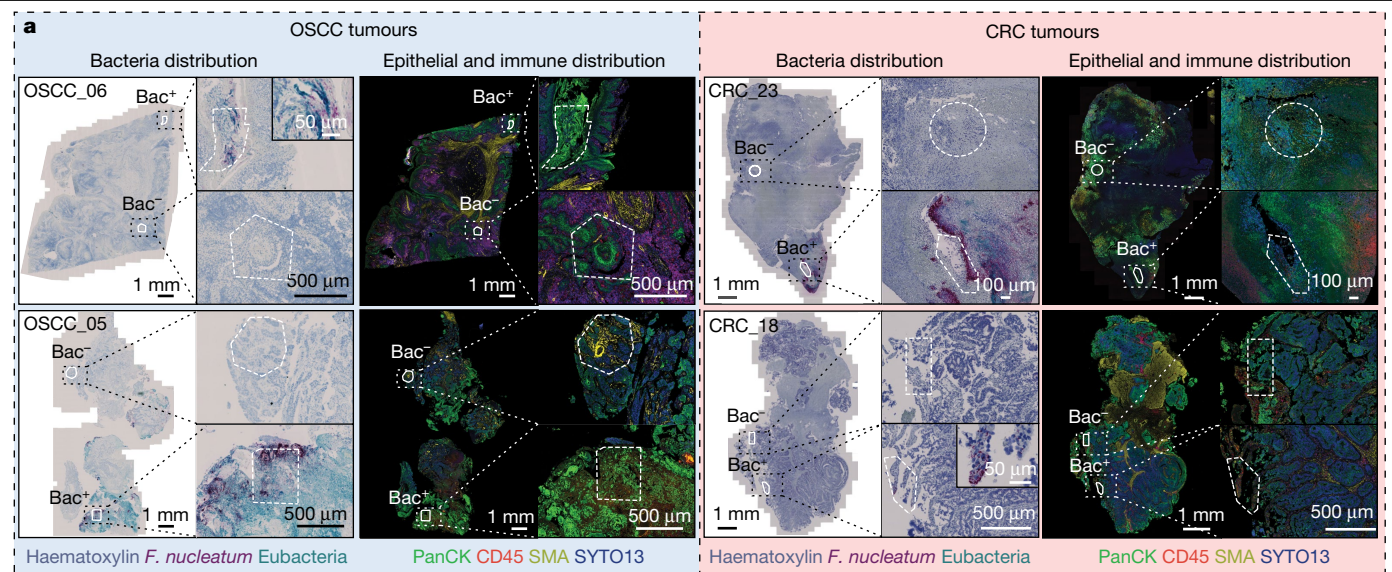
**Fig. 1 | Assessing the spatial distribution of intratumoral bacteria throughout the tumour tissue.** **a**, Haematoxylin and eosin (H&E) staining (left), spatial distribution of total bacterial reads (centre) and total UMI transcripts (right) throughout the tumour tissue in the 10x Visium capture slides from human OSCC and CRC specimens. **b**, Pie chart of the top 10 most dominant bacterial genera detected in the 10x Visium RNA-sequencing data from the OSCC and CRC tumours. **c**, RNAscope-FISH imaging showing the

distribution of bacteria across the tumour tissue in a sequential slide following the 10x Visium section. The *F. nucleatum* probe is red and the eubacterial probe is cyan. Scale bars, 1 mm. **d**, Spatial distribution of *Parvimonas*, *Peptoniphilus* and *Fusobacterium* UMIs detected in the 10x Visium OSCC specimen data. **e**, Spatial distribution of *Fusobacterium*, *Bacteroides* and *Leptotrichia* UMIs detected in the 10x Visium CRC specimen data.

heterogeneous distribution of microorganisms in the tumour tissue in a subset of patients. Through targeted RNAscope-fluorescence in situ hybridization (RNAscope-FISH) imaging we visually confirmed the heterogeneous spatial distribution of these bacterial communities, including *Fusobacterium nucleatum*, for which both densely populated compartments of bacterial cell biomass and bacteria-negative regions are observed within the same tumour specimen (Extended Data Fig. 1e). The RNAscope approach was validated for *F. nucleatum* through quantitative PCR and microbiome analysis (Extended Data Fig. 1f).

To gain further resolution on the spatial distribution and identity of the intratumoral microbiota, we applied an unbiased approach through 10x Visium spatial transcriptomics to a specimen of CRC and a specimen of OSCC (Extended Data Fig. 1g). After tissue processing, each captured microbial transcript, largely consisting of ribosomal RNA, was flanked with a barcode oligo sequence from the 10x Visium capture spot, providing spatial coordinates for the bacterial transcripts across the tumour tissue (Fig. 1a). In addition, the sequencing reads from individual microbial transcripts contained a unique molecular identifier (UMI), which enabled us to quantify the bacterial transcriptional load of viable organisms in these tissue sections (Fig. 1a). The resulting sequencing data were assessed using GATK PathSeq<sup>17</sup> to taxonomically resolve in situ sequencing reads to the

genus level (Fig. 1b and Supplementary Table 3). Within each block, a sequential tissue slide for targeted RNAscope-FISH confirmed the spatial distribution of bacteria within these tumours (Fig. 1c). Overall, bacterial transcripts were identified in 28% and 46% of the capture spots within OSCC and CRC tumours, respectively. When bacterial transcripts were detected, the number of different bacterial genera identified per capture spot ranged from 1 to 42 with a median of 8 in the OSCC tumour, and from 1 to 31 with a median of 2 in the CRC tumour. The UMI metric allowed the tissue transcriptional load of specific organisms to be quantified, and identified *Parvimonas*, *Peptoniphilus* and *Fusobacterium* as the most dominant genera in the OSCC tumour (Fig. 1d), and *Fusobacterium* and *Bacteroides* as the most dominant genera in the CRC tumour (Fig. 1e). Although a greater number of dominant genera were detected in the OSCC specimen (more than 1% relative abundance), the dominant genera in the CRC specimen (*Fusobacterium* and *Bacteroides*) had an order of magnitude more reads and UMIs than those in the OSCC specimen (Extended Data Fig. 1h and Supplementary Table 3). By applying and adapting this spatial transcriptomics approach to the intratumoral microbiota, we are able to directly identify, quantify and spatially map viable bacteria within histologically intact tumour tissues from patients. The detection of co-localized communities of both isolated genera and several



**Fig. 2 | Evaluating the effect of the tumour-associated microbiota in local microniches. a**, RNAscope-CISH images show the distribution of *F. nucleatum* (dark red) and other bacterial communities (eubacteria probe: cyan) in the tumour tissue; a sequential immunohistochemistry image shows the distribution of CD45<sup>+</sup> (red) and PanCK<sup>+</sup> (green) cells to identify the immune and epithelial compartments, respectively, in the tumour tissue. Inset images indicate representative AOs that are positive and negative for bacteria and the corresponding UV exposure regions. **b**, Volcano plots from DSP data comparing the protein expression profiles in bacteria-positive AOs and

bacteria-negative AOs from 8 OSCC (left) and 10 CRC (right) tumour specimens, referred to as microniche-level analysis. AO comparative analysis, based on bacterial status, from immune (CD45<sup>+</sup>), epithelial (PanCK<sup>+</sup>) or combined (all AOs) segmented data is shown. The number of AOs per group is indicated. Dashed lines indicate the threshold of significant gene expression, defined as  $\log_2$ -transformed fold change  $\geq 0.58$  and  $\leq -0.58$  with  $-\log_{10}(P) \geq 1.301$  after linear mixed effect model (LMM) analysis and Benjamini-Hochberg multiple-correction testing. The p prefix indicates phosphorylation; ERK1/2 refers to ERK1 and ERK2; PR, progesterone receptor.

different genera within capture spots highlights the complexity of intratumoral microbiota interactions across these tumour tissues.

### Intratumoral niches colonized by microorganisms

Given that the intratumoral microbiota has a heterogeneous distribution within individual tumour tissues, we sought to determine whether this spatial distribution correlated with distinct functions within the TME. Using a targeted approach through the GeoMx digital spatial profiling (DSP) platform (Extended Data Fig. 2a), we quantified the expression profile of 77 proteins that are associated with anti-tumour immunity and cancer progression. Segmented profiling was implemented to enrich the extracted protein data from either immune or epithelial cancer compartments within tissue areas of interest (AOs), which were annotated by RNAscope-chromogenic in situ hybridization (RNAscope-CISH) as positive or negative for bacteria (Bac<sup>+</sup> or Bac<sup>-</sup>, respectively; Fig. 2a,b).

Within CD45<sup>+</sup> immune compartments of both OSCC (DSP cohort 1;  $n = 8$  patients) and CRC (DSP cohort 2;  $n = 10$  patients) tumours, we independently show that bacteria reside in highly immunosuppressive microniches that are characterized by an enrichment of mature CD66b<sup>+</sup>

myeloid cells along with an upregulation of the immunosuppressive molecule ARG1 (arginase 1) and the immune checkpoint protein CTLA4 (cytotoxic T-lymphocyte-associated protein 4) (Fig. 2b). In addition, in both cancer types, we detected increased levels of phosphorylated ERK1 and ERK2 (Fig. 2b), which suggests that the myeloid response against intratumoral bacteria might occur through activation of the MAPK signalling pathway<sup>18</sup>. In OSCC tumours, the T-cell-inhibitory receptor PD-1 was overexpressed in bacteria-positive microniches as compared to bacteria-negative areas within specimens (Fig. 2b). This corresponded to a relative downregulation of T cell markers such as CD3, CD8, CD4, CD27 and CD44 in both the OSCC and the CRC cancer tissue, along with reduced expression of the proliferation marker Ki-67, and suggests that T cells are excluded in bacteria-colonized regions of these two cancers of the gastrointestinal tract (Fig. 2b and Supplementary Table 4).

In the PanCK<sup>+</sup> epithelial tumour compartment of both cancer types, bacteria-colonized regions were less vascularized than bacteria-negative regions, with reduced expression of smooth muscle actin (SMA) and lower levels of proliferation, as characterized by the downregulation of Ki-67 and p90 RSK (ref.<sup>19</sup>) (Fig. 2b). In bacteria-colonized

microniches of both OSCC and CRC tumour tissue, we detected a significant reduction in the protein expression of the wild-type configuration of the tumour suppressor p53, indicating that bacterial localization correlates with highly transformed cancer cells within the TME (Fig. 2b). Furthermore, bacteria-colonized microniches had significantly increased levels of phosphorylation of JNK, ERK1 and ERK2 and P38 in CRC tumours, thus revealing signalling pathways that are activated in response to bacteria (Fig. 2b). When applied to a single tumour specimen from a patient with CRC (CRC\_23) using all 24 AOs, similar protein expression profiles were obtained (Extended Data Fig. 2b,c and Supplementary Table 4). The combination of RNAscope and immunohistochemistry (IHC) techniques supported the findings from DSP of an overexpression of PD-1 in bacteria-positive microniches in OSCC specimens, in addition to a significant reduction in the levels of Ki-67, suggesting that infected regions of the OSCC and CRC tumour tissue have a lower proliferation potential than uninfected regions (Extended Data Fig. 3a,b).

Further analysis was performed in which we compared CRC tumours that have detectable bacteria (DSP cohort 2;  $n = 10$  patients) with CRC tumours that were negative for bacteria by RNAscope analysis (DSP cohort 3;  $n = 9$  patients), to determine whether bacteria-colonized microniches could have a broader effect at the tumour tissue level. We found that bacteria-positive tumours showed reduced expression levels of CD4 and CD8, along with an increased expression of immunosuppressive molecules such as CTLA4 and ARG1, and an enrichment of CD66b<sup>+</sup> myeloid cells (Extended Data Fig. 2d and Supplementary Table 4), supporting previous bulk tissue analysis<sup>20,21</sup>.

RNAscope and IHC confirmed that bacteria-positive regions of tissue had significant increases in CD11b<sup>+</sup> and CD66b<sup>+</sup> myeloid cells, along with lower densities of CD4<sup>+</sup> and CD8<sup>+</sup> T cells, as compared to immediately adjacent bacteria-negative regions; this indicates that the effect of the tumour-associated microbiota is highly localized (Extended Data Fig. 3c,d).

### Microorganism-driven single-cell heterogeneity

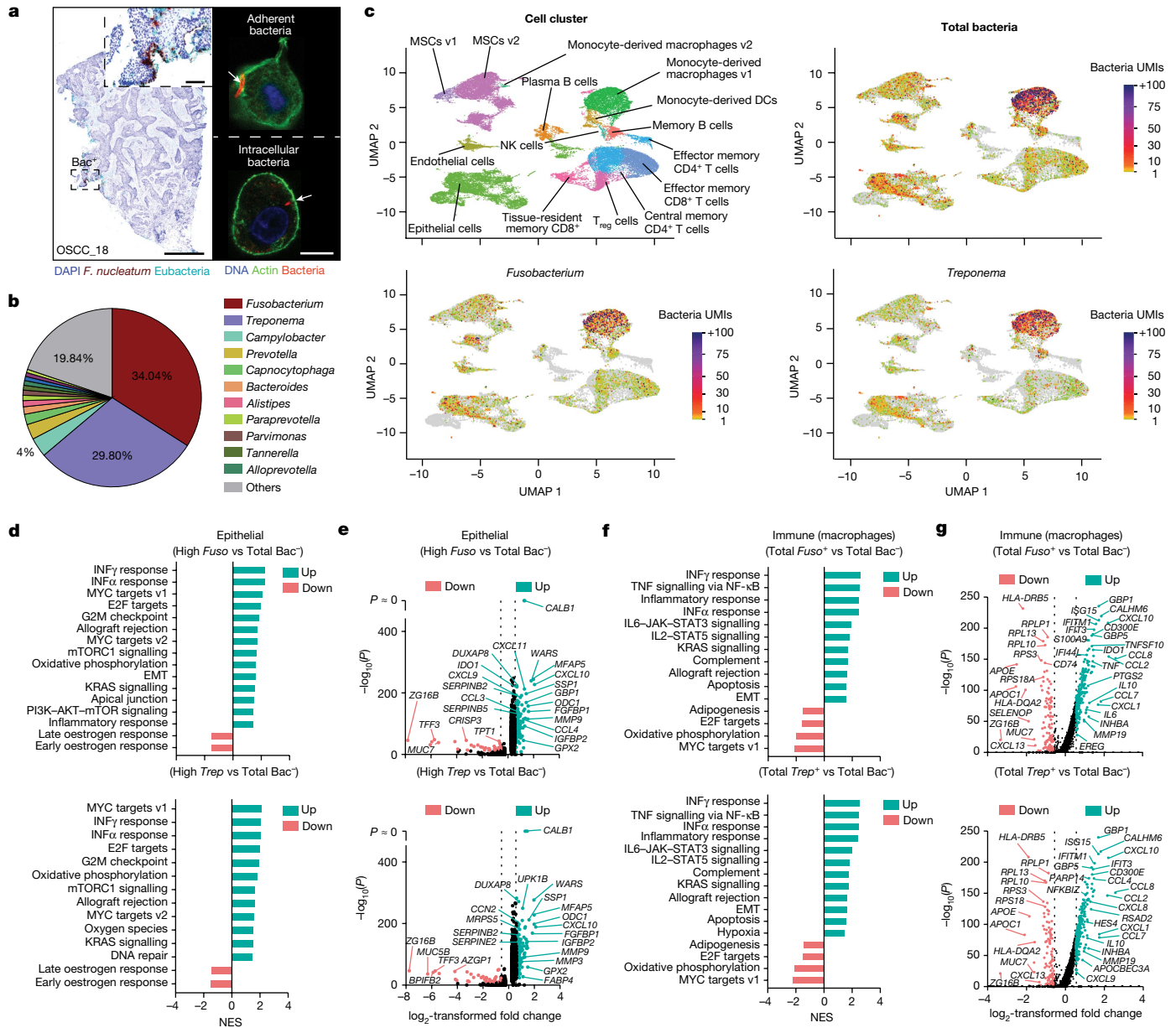
The presence of bacteria within individual host cells of the TME has been reported across a range of human cancer types<sup>2,13</sup>. However, we have little information on the identity of invasive bacteria, the host cells that they interact with and how these host–bacterial associations affect cellular function within the TME. To investigate bacterial–host cell-to-cell interaction within the TME and the effect on host cell transcriptomics, we developed INVADeseq (invasion–adhesion-directed expression sequencing) by introducing a primer that targets a conserved region of bacterial 16S rRNA, facilitating the generation of cDNA libraries with bacterial transcripts from the bacteria-associated human cells (Extended Data Fig. 4a). Addition of this bacteria-targeting primer did not affect the gene-expression profile of human CRC cells (Extended Data Fig. 4b), and validation co-culture experiments with non-adherent and non-invasive *Escherichia coli* DH5 $\alpha$  (Extended Data Fig. 4c) showed specificity for cell-associated bacteria.

To further validate this approach, the human CRC cell line HCT116 was infected with three invasive bacterial species—*F. nucleatum*, *Porphyromonas gingivalis* and *Prevotella intermedia*—at a multiplicity of infection (MOI) of 100:1 and 500:1, and processed for INVADeseq (Extended Data Fig. 4d). Confocal imaging indicated the presence of intracellular bacteria in cancer cells after bacterial co-culture (Extended Data Fig. 4e). Using INVADeseq, we mapped bacterial reads to single human cells (Extended Data Fig. 4f,g). At the cell-cluster level for these epithelial single cells (clusters 1–10), most *F. nucleatum*- and *P. gingivalis*-positive single cells were distributed in cancer cell clusters 5 and 6, respectively (Extended Data Fig. 4g). Both cell clusters (clusters 5 and 6) were very minor cell populations in the uninfected control group (Extended Data Fig. 4f). When compared to uninfected controls (MOI = 0), the appearance of cell clusters 5 (*Fusobacterium*

cluster) and 6 (*Porphyromonas* cluster) coincided with a relative reduction in the percentage of cluster 1 (uninfected control cluster) (Extended Data Fig. 4g). This finding suggests that *F. nucleatum* and *P. gingivalis* affect cancer cell heterogeneity by altering distinct transcriptional programs that contribute to specific cell clusters (Extended Data Fig. 4g).

After integrating data from the three HCT116 samples (Extended Data Fig. 4h,i), we compared the gene expression of *F. nucleatum*- or *P. gingivalis*-associated single epithelial cells to that of the bacteria-negative epithelial single cells (Total Bac<sup>-</sup>). We noted that the number of differentially expressed genes increased when a bacterial UMI cut-off ( $\geq 3$ ), a proxy for bacterial transcriptional load, was applied (Extended Data Fig. 5a–d and Supplementary Table 5). Furthermore, a comparison of cells from cluster 5 (*Fusobacterium* cluster) and cluster 6 (*Porphyromonas* cluster) to bacteria-negative cells from cluster 1 (uninfected control cluster) showed that bacteria-infected cells exhibited a significant upregulation of signalling pathways that are involved in the response to bacterial infection, such as the TNF pathway and pathways related to inflammation and hypoxia, as well as cancer cell progression via the epithelial–mesenchymal transition (EMT) and the p53 signalling pathway<sup>22,23</sup>. Bacteria-infected cells also showed a downregulation of cell-cycle signalling pathways that relate to the formation of the mitotic spindle and the G2–M DNA damage checkpoint, as compared with cells from the uninfected control cluster (Extended Data Fig. 5a–d). At the gene-expression level, bacteria-associated single epithelial cells showed significant increases in the expression of molecules that are positively associated with metastasis, such as *PLAU*, *PLAUR*, *RELB* and *AREG*, along with an upregulation of the chemokines *CXCL1*, *CXCL2*, *CXCL3* and the neutrophil chemoattractant *CXCL8*, along with members of the TNF family (Extended Data Fig. 5a–d). We also noted a significant upregulation of transcription factors including *NFKBIA*, *NFKB2*, *NEAT1*, *SAT1* and members of the JUN and FOS family, with a downregulation of the cyclins *CCNB1* and *CCNA2* (Extended Data Fig. 5a–d). Similar findings were observed when CRC-derived HT-29 cells were treated with *F. nucleatum* at a MOI of 100:1; that is, an increase in the expression of genes that encode molecules related to inflammation through TNF, hypoxia, the EMT and p53 signalling pathways, and a reduction in the expression of genes that are involved in DNA repair (Extended Data Fig. 5e–g and Supplementary Table 5).

The INVADeseq method was subsequently implemented to examine bacteria–host interactions in fresh tumour tissues from seven patients with OSCC. After the tissues were dissociated to single cells, confocal imaging showed that single cells from a tumour from a patient with OSCC contained cell-adherent and intracellular bacteria (Fig. 3a). Integrated scRNA-seq from the seven tumours revealed that the intratumoral microbiota is dominated by bacterial species that belong to the *Fusobacterium* (34%) and *Treponema* (29.8%) genera (Fig. 3b). Mapping bacterial reads from INVADeseq analysis to annotated single cells showed that *Fusobacterium* and *Treponema* were predominantly associated with the epithelial and monocyte-derived macrophage-v1 (referred to as the macrophage cluster) cell clusters in these patient tumours, with a total bacterial infection rate of 25% and 52%, respectively (Fig. 3c and Extended Data Fig. 6a). INVADeseq cannot distinguish whether bacteria are actively invading the macrophage cells or whether the macrophages are phagocytizing the bacteria; however, we refer to these cells as ‘macrophages with bacteria engulfed’. Within the epithelial cell clusters, cells in cluster 3 were identified as aneuploid, confirming that these are tumour cells with severe chromosomal instability (Extended Data Fig. 6b–d). Notably, this aneuploid epithelial cell cluster contained most of the bacterial UMI transcripts, as compared to other euploid epithelial cell clusters (Extended Data Fig. 6d). Gene set enrichment analysis (GSEA) confirmed that the cells from the bacteria-dominant epithelial cell cluster 3 were indeed cancer cells, with gene-expression signatures characterized by an upregulation of signalling pathways involved in cancer progression, including EMT, PI3K–AKT–mTOR,



**Fig. 3 | Effect of cell-associated intratumoral bacteria on transcriptomics in host single cells. a**, RNAscope-FISH (left) shows the distribution of intratumoral bacteria in a tumour from a patient with OSCC. Confocal images (right) show bacteria-associated single cells after tissue dissociation. Scale bars, 1 mm (left); 5  $\mu$ m (right). **b**, Microbiome composition at the genus level after integration of tumour scRNA-seq data from seven patients with OSCC using the INVADeseq method. **c**, UMAP plots indicate host cell annotation and bacteria transcripts (UMI) from total bacteria and *Fusobacterium*- and *Treponema*-associated cells in integrated tumour single-cell data from seven patients with OSCC as indicated. Colour bars indicate the bacterial UMI transcripts for total bacteria and for each bacterial species as indicated. DCs, dendritic cells; MSCs, mesenchymal stem cells; T<sub>reg</sub> cells, regulatory T cells. **d**, GSEA analysis showing the signalling pathways that are differentially regulated in cells that contain  $\geq 3$  *Fusobacterium* UMI (High *Fuso*) or  $\geq 3$  *Treponema* UMI (High *Trep*) transcripts versus (vs) total bacteria-negative cells

(Total Bac<sup>-</sup>) from the epithelial cell cluster. **e**, Volcano plots showing the differentially expressed genes between cell populations described in **d**. Dashed lines indicate the threshold of significant gene expression, defined as  $\log_2$ -transformed fold change  $\leq -0.58$  and  $\geq 0.58$  with  $-\log_{10}(P) \geq 1.301$ . **f**, GSEA analysis showing the signalling pathways that are differentially regulated between total *Fusobacterium* (Total *Fuso*<sup>+</sup>) or total *Treponema* (Total *Trep*<sup>+</sup>) associated cells versus bacteria-negative cells (Total Bac<sup>-</sup>) in the monocyte-derived macrophage-v1 cell cluster. **g**, Volcano plots showing the differentially expressed genes between cell populations described in **f**. Dashed lines indicate the threshold of significant gene expression, defined as  $\log_2$ -transformed fold change  $\leq -0.58$  and  $\geq 0.58$  with  $-\log_{10}(P) \geq 1.301$ . The normalized enrichment scores (NESs) in **d, f** were calculated using the Wilcoxon rank sum test. LMM analysis followed by Benjamini-Hochberg multiple-correction testing was used to calculate the fold change and *P* values for each gene in **e, g**.

hypoxia and the interferon (IFN) response, among others (Extended Data Fig. 6e–g).

To determine whether the dominant cell-associated bacterial genera, *Fusobacterium* and *Treponema*, affected epithelial signalling pathways, *Fusobacterium*- or *Treponema*-associated single cells (UMI  $\geq 3$ ) were

compared to bacteria-negative cells (Total Bac<sup>-</sup>) from the epithelial cell cluster. After GSEA analysis, we observed a significant upregulation of IFN and JAK–STAT signalling, with increased expression of molecules from the SERPIN family; chemokines such as *CXCL10*, *CXCL11*, *CCL4* and *CCL3*; and metalloproteinases, including *MMP9* and *MMP3* (Fig. 3d, e

and Supplementary Table 6). A comparison of general bacteria-positive epithelial cells (Total Bac<sup>+</sup>), independent of a specific genus, and bacteria-negative cells (Total Bac<sup>-</sup>) showed that gene expression and cell signalling pathways related to cancer progression were modestly affected in bacteria-positive epithelial cells, as compared to the effects that were observed in cells infected with specific taxa (Extended Data Fig. 6h–j and Supplementary Table 6). This is likely to be reflective of taxa-specific epithelial cell interactions or capabilities rather than a general bacteria-induced response.

At the specimen level, the total bacterial load from each sample was negatively correlated with the expression of *TP53* and positively correlated with its negatively regulated target molecule, *SAT1* (Extended Data Fig. 7a)—consistent with our findings from DSP (Fig. 2), in which bacteria colonized microniches with reduced levels of wild-type p53. In addition, the total bacterial load negatively correlated with the expression of the proliferation marker *MKI67*, which encodes Ki-67 (Extended Data Fig. 7a), providing support for our spatial microniche data (Fig. 2 and Extended Data Fig. 3a,b).

In the macrophage cell cluster, by comparing Total Bac<sup>+</sup> to Total Bac<sup>-</sup> cells, we found that macrophages with bacteria engulfed had significantly increased expression levels of genes that are involved in the inflammatory response through activation of TNF, INF $\gamma$  and IFN $\alpha$ , and genes that are involved in the production of interleukins through the JAK–STAT signalling pathway, such as *IL1B*, *IL6* and *IL10*. Macrophages with bacteria engulfed also overexpressed the chemokines *CCL2*, *CCL4*, *CCL8*, *CCL7*, *CXCL1* and *CXCL10* (Extended Data Fig. 6k,l and Supplementary Table 7). This gene-expression signature was observed when analysing cells associated with bacteria in general (Extended Data Fig. 6k,l), but also when assessing specific bacterial genera, including *Fusobacterium* and *Treponema* (Fig. 3f,g and Supplementary Table 7). Furthermore, at the specimen level, the bacterial load from each OSCC specimen was positively correlated with the potent neutrophil chemoattractant *CXCL8* and negatively correlated with the expression of *CD3E* (Extended Data Fig. 7a), supporting the DSP findings that intratumoral bacteria-colonized microniches are immunosuppressive by recruiting neutrophils and excluding CD3<sup>+</sup> T cells (Fig. 2 and Extended Data Fig 3c,d).

Unlike our findings in ‘macrophages with bacteria engulfed’ single cells, in which the response appears generalized to the presence of bacterial lipopolysaccharide or other widespread damage-associated molecular patterns, in epithelial single cells, specific dominant taxa such as *Fusobacterium* and *Treponema* enhanced signatures of cancer progression. Overall, this shows that the cell-associated members of the intratumoral microbiota can drive heterogeneity in patient tumours at the single-cell level within immune and epithelial populations.

An independent analysis of tumour single-cell data from the individual patients with OSCC revealed inter-patient heterogeneity in bacterial load, dominant cell-associated bacterial taxa and magnitude of the inflammatory gene-expression response (Extended Data Fig. 8a–d and Supplementary Tables 8–10). Similar to the integrated analysis, the percentage of bacteria-associated single cells is significantly higher in the aneuploid cancer epithelial cell cluster (cluster 3) compared to the euploid epithelial cell clusters (Extended Data Fig. 8e). This single-cell analysis of individual patients shows that specific cell-associated bacteria can significantly affect intratumoral heterogeneity at the single-cell level (Extended Data Fig. 8a–d and Supplementary Tables 8–10).

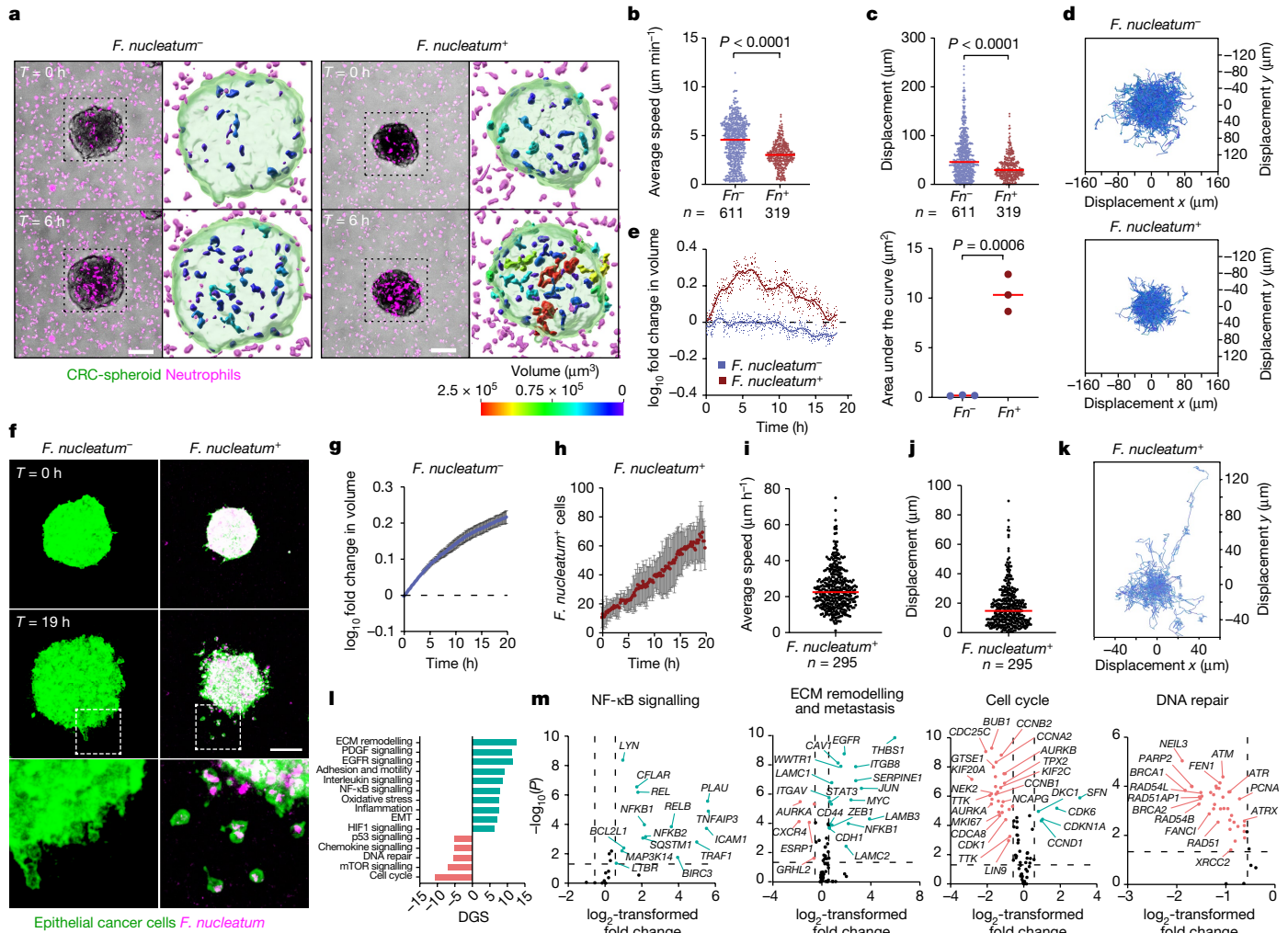
### Bacteria-induced migration of cancer cells

To evaluate the direct interactions of a dominant member of the intratumoral microbiota with immune or epithelial cancer cells, we used a reductionist in vitro co-culture approach. We co-cultured CRC epithelial spheroids with an *F. nucleatum* CRC isolate, followed by embedding in collagen matrices that contained neutrophils distributed uniformly throughout the gel (Supplementary Methods). By using live-cell

confocal microscopy, the embedded neutrophils are tracked inside *F. nucleatum*-infected spheroids and could be compared to control uninfected spheroids (Fig. 4a and Supplementary Video 1). In the absence of *F. nucleatum*, neutrophils migrated freely inside the spheroids with an average speed of  $4.329 \mu\text{m min}^{-1} \pm 0.08766$  (s.e.m.) (Fig. 4b) and average cell displacement of  $57.21 \mu\text{m}$  (Fig. 4c,d). In the presence of *F. nucleatum*, neutrophils responded to bacterial infection by reducing their migration capabilities with an average speed of  $3.593 \mu\text{m min}^{-1} \pm 0.08561$  (s.e.m.) (Fig. 4b) and a mean cell displacement of  $34.53 \mu\text{m}$  (Fig. 4c,d) as they formed cell clusters inside the spheroids, reaching a maximum size at around 6 h, after which the clusters gradually started to disassemble<sup>24</sup> (Fig. 4e). The recruitment and retention of neutrophils to the cancer cell spheroids that were infected with *F. nucleatum* indicates that the intratumoral microbiota has an active role in the enrichment of neutrophils within bacteria-colonized microniches of patient tumours, as was observed through spatial profiling (Fig. 2b). Neutrophil cluster formation (Extended Data Fig. 9a,b) was accompanied by significantly increased levels of phosphorylation of ERK and p38 MAPK in response to *F. nucleatum* (Extended Data Fig. 9c). This suggests that the observed upregulation of phosphorylation of ERK and p38 MAPK in bacteria-colonized microniches within patient tumours is in part driven by a myeloid response to intratumoral bacteria (Fig. 2b).

We also show that the CRC epithelial cells infected with *F. nucleatum* detached from the spheroid mass and migrated into the surrounding collagen gel as single epithelial cells (Fig. 4f and Supplementary Video 2). By contrast, uninfected cancer epithelial cells invaded as a collective, as the entire spheroid mass spread through the gel at an average expansion rate of  $1.34 \times 10^5 \mu\text{m}^3 \text{h}^{-1}$  (Fig. 4g). Notably, intracellular bacteria were detected in migratory cancer cells as they invaded the collagen gel (Fig. 4h). Cell-tracking analysis showed that invading cancer cells migrated with an average speed of  $19.99 \mu\text{m h}^{-1} \pm 0.827$  (s.e.m.) (Fig. 4i) and a mean displacement of  $20.54 \mu\text{m}$  (Fig. 4j,k). Similar findings were observed when assessing a mouse CRC cell line, in which *F. nucleatum*-infected cancer cells invaded the surrounding collagen gel as individual cells (Extended Data Fig. 9d–i and Supplementary Video 3). Invasive bacteria not only promote the invasion of cancer cells in collagen matrices<sup>25</sup>, but also change the motility patterns of infected cancer cells and thereby promote cell heterogeneity at a functional level, as shown by the broad range of cell-displacement and velocity values measured in cells as they migrate through the gel (Fig. 4i–k).

To identify the altered signalling pathways involved, spheroids containing *F. nucleatum* were dissociated for transcriptomic analysis using the nCounter platform. Differential expression at the pathway level indicated that exposure to *F. nucleatum* led to a significant upregulation of signalling pathways that are involved in cancer progression, including remodelling of the extracellular matrix, metastasis and cell adhesion and migration, as well as an upregulation of signalling through growth factors such as EGFR and PDGF, and signalling through the EMT and NF- $\kappa$ B pathways (Fig. 4l,m). Furthermore, *F. nucleatum*-treated spheroids exhibited a downregulation of signalling pathways related to the cell cycle, DNA damage repair and p53 signalling (Fig. 4l,m and Supplementary Table 11). Transcriptional modifications induced by *F. nucleatum* were also found in the mouse CRC cell line, with an upregulation of metastasis- and EMT-related genes (Extended Data Fig. 9j,k and Supplementary Table 11). Moreover, we observed a significant downregulation in *MKI67* expression in *F. nucleatum*-exposed cancer spheroids (Supplementary Table 11), consistent with the downregulation of Ki-67 in bacteria-colonized microniches that was observed in the DSP analysis from patient tumours (Fig. 2) and the negative correlation with bacterial load that was observed in the whole-sample analysis from the scRNA-seq data (Extended Data Fig. 7a). Of note, despite lower proliferation levels, the spheroids infected with *F. nucleatum* had increased single-cell migration capabilities (Fig. 4f and Supplementary Video 2). This is of particular interest given our previous finding that the dominant intratumoral microbiota—including *F. nucleatum*—persists with



**Fig. 4 | *F. nucleatum* induces neutrophil swarming and the migration of cancer epithelial cells.** **a**, Live-cell confocal imaging showing neutrophil movements relating to CRC spheroids without (left) or with (right) *F. nucleatum*. Colour bars represent neutrophil cluster volume ( $\mu\text{m}^3$ ). Scale bars, 100  $\mu\text{m}$ . **b, c**, Average speed (**b**) and cell displacement (**c**) of neutrophils migrating inside untreated control (lilac) and *F. nucleatum*-treated (red) spheroids. Red bars indicate mean. Data points represent individual tracks; *n* indicates the number of tracks per condition; three independent experiments. *P* values calculated by Mann–Whitney test. **d**, Neutrophil cell trajectory plots. **e**, Left, the  $\log_{10}$ -transformed fold change in volume over time of neutrophil clusters relative to the initial volume ( $T = 0$  h). Data points represent average volume per time point, per condition. Right, quantification of the area under the curve for the fold change in volume. **f**, Confocal microscopy of HCT116 spheroid invasion capabilities without (left) or with (right) *F. nucleatum* over 19 h. Inset images represent differences in migration modes. Scale bar, 100  $\mu\text{m}$ . **g**, The  $\log_{10}$ -transformed fold change in volume over time, representing the

expansion rate of uninfected CRC spheroids. Error bars, s.d. **h**, Number of *F. nucleatum*-positive single cancer cells detaching from the spheroid over time. Error bars, s.d. **i, j**, Average speed (**i**) and cell displacement (**j**) of single cells escaping the *F. nucleatum*-infected spheroid. Red bars indicate mean. Data points represent individual tracks; *n* indicates the number of tracks per condition; three independent experiments. **k**, Cell trajectories of invading cancer cells escaping the *F. nucleatum*-infected spheroids. **l**, Signalling pathway analysis of CRC spheroids infected with *F. nucleatum* compared to uninfected control. The directed global significance (DGS) score was calculated as the square root of the mean squared *t*-statistic for genes in a gene set. ECM, extracellular matrix. **m**, Volcano plots of differential gene expression for selected pathways in *F. nucleatum*-infected spheroids compared to uninfected controls. Dashed lines indicate the threshold of significance, defined as  $\log_2$ -transformed fold change  $\leq -0.58$  and  $\geq 0.58$  and  $-\log_{10}(P) \geq 1.301$  after LMM analysis and Benjamini–Hochberg multiple-correction testing.

CRC in distant-site metastases<sup>8</sup>, and warrants further investigation into the effect of the intratumoral microbiota on the cell migration–proliferation dichotomy in cancer<sup>26,27</sup>. Finally, flow cytometry analysis showed that cancer epithelial cells also upregulated the levels of phosphorylation of ERK1 and ERK2 in the presence of *F. nucleatum* during the formation of cancer cell clusters<sup>28</sup> (Extended Data Fig. 9l–n). Together, these data show that *F. nucleatum* derived from human CRC actively induces the recruitment of myeloid cells at the sites of bacterial infection and promotes transcriptional changes in CRC epithelial cells that facilitate invasion to the surrounding environment and may confer quiescent properties.

## Discussion

Historically, tumour heterogeneity was attributed solely to intrinsic genetic alterations in cancer cells during clonal expansion<sup>29</sup>. Studies in the 1990s<sup>30,31</sup> revealed that extrinsic factors derived from the TME<sup>32,33</sup> have an important role in tumorigenesis. The intercellular interactions between cancer cells and other non-malignant cell populations such as fibroblasts, endothelial and immune cells in the TME are known to contribute to tumour heterogeneity by promoting transcriptomic changes in transformed cells as the cancer evolves<sup>34–36</sup>. As our understanding of the TME advances, so too does our understanding of

what affects tumour heterogeneity. Genomics-based studies have shown that most major types of human cancer contain an intratumoral microbiota<sup>2,16</sup>. These microbial communities vary by cancer type, and specific bacteria can contribute to the initiation and progression of cancer, affect the response of patients to treatment and thus affect survival<sup>2,8,12,14–16,21,37</sup>. Nevertheless, the intrinsic heterogeneity present has made it difficult to understand the interplay between different components of the TME, including bacteria–host interactions within the native tissue context. The development of spatial transcriptomics<sup>38</sup> and scRNA-seq technologies<sup>39,40</sup> has enabled eukaryotic components of the TME to be studied, but the effect of the intratumoral microbiota in the TME has so far been overlooked. In this study, by adapting and applying these technologies, we conclude that the intratumoral microbiota is heterogeneously distributed across human tumours. Further, we show that it is a fundamental component of the TME that can alter the biology of distinct cellular compartments, affecting anti-tumour immunity and the migration of cancer epithelial cells. By activating transcriptional factors from the JUN and FOS family, intracellular bacteria can generate gene signatures that are consistent with cancer cell invasion, metastasis, DNA damage repair and cell dormancy. Likewise, invasive bacteria are responsible for recruiting myeloid cells to induce an inflammatory response through JAK–STAT signalling, promoting T cell exclusion and tumour growth by secreting specific interleukins and chemokines into the surrounding environment. Although we focused here on two cancer types at the extremes of the gastrointestinal tract, the tools and technologies that we describe could be applied to analyse the 33 major cancer types that have so far been shown to contain an intratumoral microbiota. Analyses that move beyond correlative associations of the microbiota with human cancers, towards those that assess the functional effect of the intratumoral microbiota, will identify molecular and cellular targets for the prevention and treatment of such cancers. Collectively, this work shows that the distribution of the intratumoral microbiota within patient tumours is not random, but rather, that the microbiota is highly organized in microniches with immune and epithelial cell functions that support cancer progression.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-022-05435-0>.

1. Sepich-Poore, G. D. et al. The microbiome and human cancer. *Science* **371**, eabc4552 (2021).
2. Nejman, D. et al. The human tumor microbiome is composed of tumor type-specific intracellular bacteria. *Science* **368**, 973–980 (2020).
3. Kostic, A. D. et al. Genomic analysis identifies association of *Fusobacterium* with colorectal carcinoma. *Genome Res.* **22**, 292–298 (2012).
4. Rao, A., Barkley, D., Franca, G. S. & Yanai, I. Exploring tissue architecture using spatial transcriptomics. *Nature* **596**, 211–220 (2021).
5. Tang, F. et al. mRNA-seq whole-transcriptome analysis of a single cell. *Nat. Methods* **6**, 377–382 (2009).
6. Merritt, C. R. et al. Multiplex digital spatial profiling of proteins and RNA in fixed tissue. *Nat. Biotechnol.* **38**, 586–599 (2020).
7. LaCourse, K. D., Johnston, C. D. & Bullman, S. The relationship between gastrointestinal cancers and the microbiota. *Lancet Gastroenterol. Hepatol.* **6**, 498–509 (2021).
8. Bullman, S. et al. Analysis of *Fusobacterium* persistence and antibiotic response in colorectal cancer. *Science* **358**, 1443–1448 (2017).
9. Parhi, L. et al. Breast cancer colonization by *Fusobacterium nucleatum* accelerates tumor growth and metastatic progression. *Nat. Commun.* **11**, 3259 (2020).
10. Fu, A. et al. Tumor-resident intracellular microbiota promotes metastatic colonization in breast cancer. *Cell* **185**, 1356–1372 (2022).

11. Jin, C. et al. Commensal microbiota promote lung cancer development via  $\gamma\delta$  T cells. *Cell* **176**, 998–1013 (2019).
12. Riquelme, E. et al. Tumor microbiome diversity and composition influence pancreatic cancer outcomes. *Cell* **178**, 795–806 (2019).
13. Kalaora, S. et al. Identification of bacteria-derived HLA-bound peptides in melanoma. *Nature* **592**, 138–143 (2021).
14. Yu, T. et al. *Fusobacterium nucleatum* promotes chemoresistance to colorectal cancer by modulating autophagy. *Cell* **170**, 548–563 (2017).
15. Geller, L. T. et al. Potential role of intratumor bacteria in mediating tumor resistance to the chemotherapeutic drug gemcitabine. *Science* **357**, 1156–1160 (2017).
16. Poore, G. D. et al. Microbiome analyses of blood and tissues suggest cancer diagnostic approach. *Nature* **579**, 567–574 (2020).
17. Walker, M. A. et al. GATK PathSeq: a customizable computational tool for the discovery and identification of microbial sequences in libraries from eukaryotic hosts. *Bioinformatics* **34**, 4287–4289 (2018).
18. Van Ziffle, J. A. & Lowell, C. A. Neutrophil-specific deletion of Syk kinase results in reduced host defense to bacterial infection. *Blood* **114**, 4871–4882 (2009).
19. Lara, R., Seckl, M. J. & Pardo, O. E. The p90 RSK family members: common functions and isoform specificity. *Cancer Res.* **73**, 5301–5308 (2013).
20. Mima, K. et al. *Fusobacterium nucleatum* and T cells in colorectal carcinoma. *JAMA Oncol.* **1**, 653–661 (2015).
21. Serna, G. et al. *Fusobacterium nucleatum* persistence and risk of recurrence after preoperative treatment in locally advanced rectal cancer. *Ann. Oncol.* **31**, 1366–1375 (2020).
22. Pastushenko, I. & Blanpain, C. EMT transition states during tumor progression and metastasis. *Trends Cell Biol.* **29**, 212–226 (2019).
23. Müller, P. A. & Vousden, K. H. Mutant p53 in cancer: new functions and therapeutic opportunities. *Cancer Cell* **25**, 304–317 (2014).
24. Kienle, K. et al. Neutrophils self-limit swarming to contain bacterial growth in vivo. *Science* **372**, eabe7729 (2021).
25. Oliveira, M. J. et al.  $\beta$ -casein-derived peptides, produced by bacteria, stimulate cancer cell invasion and motility. *EMBO J.* **22**, 6161–6173 (2003).
26. Hatzikirou, H., Basanta, D., Simon, M., Schaller, K. & Deutsch, A. ‘Go or grow’: the key to the emergence of invasion in tumour progression? *Math. Med. Biol.* **29**, 49–65 (2012).
27. Gallaher, J. A., Brown, J. S. & Anderson, A. R. A. The impact of proliferation–migration tradeoffs on phenotypic evolution in cancer. *Sci. Rep.* **9**, 2425 (2019).
28. Wrenn, E. D. et al. Regulation of collective metastasis by nanoluminal signaling. *Cell* **183**, 395–410 (2020).
29. Dagogo-Jack, I. & Shaw, A. T. Tumour heterogeneity and resistance to cancer therapies. *Nat. Rev. Clin. Oncol.* **15**, 81–94 (2018).
30. van der Bruggen, P. et al. A gene encoding an antigen recognized by cytolytic T lymphocytes on a human melanoma. *Science* **254**, 1643–1647 (1991).
31. Dvorak, H. F. Tumors: wounds that do not heal. Similarities between tumor stroma generation and wound healing. *N. Engl. J. Med.* **315**, 1650–1659 (1986).
32. Anderson, N. M. & Simon, M. C. The tumor microenvironment. *Curr. Biol.* **30**, R921–R925 (2020).
33. Joyce, J. A. & Fearon, D. T. T cell exclusion, immune privilege, and the tumor microenvironment. *Science* **348**, 74–80 (2015).
34. Kieffer, Y. et al. Single-cell analysis reveals fibroblast clusters linked to immunotherapy resistance in cancer. *Cancer Discov.* **10**, 1330–1351 (2020).
35. Qian, J. et al. A pan-cancer blueprint of the heterogeneous tumor microenvironment revealed by single-cell profiling. *Cell Res.* **30**, 745–762 (2020).
36. Wu, F. et al. Single-cell profiling of tumor heterogeneity and the microenvironment in advanced non-small cell lung cancer. *Nat. Commun.* **12**, 2540 (2021).
37. Pernigoni, N. et al. Commensal bacteria promote endocrine resistance in prostate cancer through androgen biosynthesis. *Science* **374**, 216–224 (2021).
38. Vickovic, S. et al. High-definition spatial transcriptomics for in situ tissue profiling. *Nat. Methods* **16**, 987–990 (2019).
39. Macosko, E. Z. et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202–1214 (2015).
40. Hunter, M. V., Moncada, R., Weiss, J. M., Yanai, I. & White, R. M. Spatially resolved transcriptomics reveals the architecture of the tumor-microenvironment interface. *Nat. Commun.* **12**, 6278 (2021).

**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



## Reporting summary

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

## Data availability

Raw sequencing data from bulk 16S ribosomal RNA gene sequencing, 10x Visium spatial transcriptomics and INVADeSeq bacterial 16S rRNA and human (GEX) gene libraries are available in the NCBI Sequence Read Archive (SRA) repository under the Bioproject accession number PRJNA811533. PathSeq, Cell Ranger and Space Ranger analyses used GRCh38 as the human genome reference.

## Code availability

Custom code for data processing and analysis of 10x Visium spatial transcriptomics and scRNA-seq data is available at [https://github.com/FredHutch/Galeano-Nino-Bullman-Intratatumoral-Microbiota\\_2022](https://github.com/FredHutch/Galeano-Nino-Bullman-Intratatumoral-Microbiota_2022).

**Acknowledgements** This research was supported by the Experimental Histopathology Shared Resource of the Fred Hutch/University of Washington Cancer Consortium (P30 CA015704), the Genomics and Bioinformatics Shared Resource of the Fred Hutch/University of Washington Cancer Consortium (P30 CA015704) and the Cellular Imaging Shared Resource (CISR) of the Fred Hutch/University of Washington Cancer Consortium (P30 CA015704). Scientific Computing Infrastructure at Fred Hutch was funded by ORIP grant S10OD028685. Research reported in this publication was supported by the National Institute of Dental and Craniofacial Research of the National Institutes of Health under award number R01 DE027850 (to C.D.J.), the National Cancer Institute under award number R00 CA229984-03 (to S.B.), the Inter-disciplinary Training Grant in Cancer T32 CA080416 (to J.L.G.N.) and the Cancer Research Institute Irvington Postdoctoral Fellowship (CRI award CRI4208 to J.L.G.N.). We thank G. Bacon

for philanthropic funding to S.B.; H. Johnston for guidance and discussions; H. Srinivasan and M. Fitzgibbon of the Fred Hutch Bioinformatics shared resource for advice on single-cell analysis; B. Reid and C. Sanchez for access to specimens; G. Javier for help with neutrophil assays; B. Hardin from Genomics and M. Haraguchi from Experimental Histopathology for processing specimens; L. Schroeder from Cellular Imaging for technical support in confocal microscopy; and H. Bouzek for discussions on statistical approaches. The illustrations of humans in Extended Data Figs. 1a and 2a were created with Biorender.com.

**Author contributions** J.L.G.N., C.D.J. and S.B. designed the study and wrote the paper. H.W. performed computational analysis of 10x Visium and INVADeSeq data. S.S.M. provided computational guidance. A.B., K.D.L., J.L.G.N. and H.W. performed microbiome analysis. J.L.G.N. and A.G.K. performed RNAscope and IHC assays. B.B., N.F. and J.H. provided patient specimens. J.L.G.N. analysed GeoMx DSP data. K.D.L., A.B. and C.S. performed INVADeSeq experiments and library preparation. K.D.L. and A.B. carried out the bacteria and cell line co-cultures. K.D.L. did the confocal microscopy for invasion. K.D.L. and A.B. processed patient specimens for microbial culture, single-cell analysis and OCT embedding. J.L.G.N. and H.W. performed the gene-expression pathway analysis of INVADeSeq data. J.L.G.N. performed co-culture, live-cell imaging and RNA-sequencing data analysis of epithelial cell spheroids and neutrophils with and without *F. nucleatum*. E.S. performed RNAscope validation and optimization. J.L.G.N. analysed flow cytometry data. J.L.G.N. and H.W. performed statical analysis. A.T., S.S.M, B.B., K.D.L. and H.W. provided edits to the paper. All authors contributed to the final version of the paper.

**Competing interests** S.B. has consulted for GlaxoSmithKline and BiomX. C.D.J. has consulted for Series Therapeutics and Azitra. S.B. is an inventor on US patent application no. PCT/US2018/042966, submitted by the Broad Institute and Dana-Farber Cancer Institute, which covers the targeting of *Fusobacterium* for the treatment of colorectal cancer. K.D.L. is currently employed by NanoString Technologies. The remaining authors declare no competing interests.

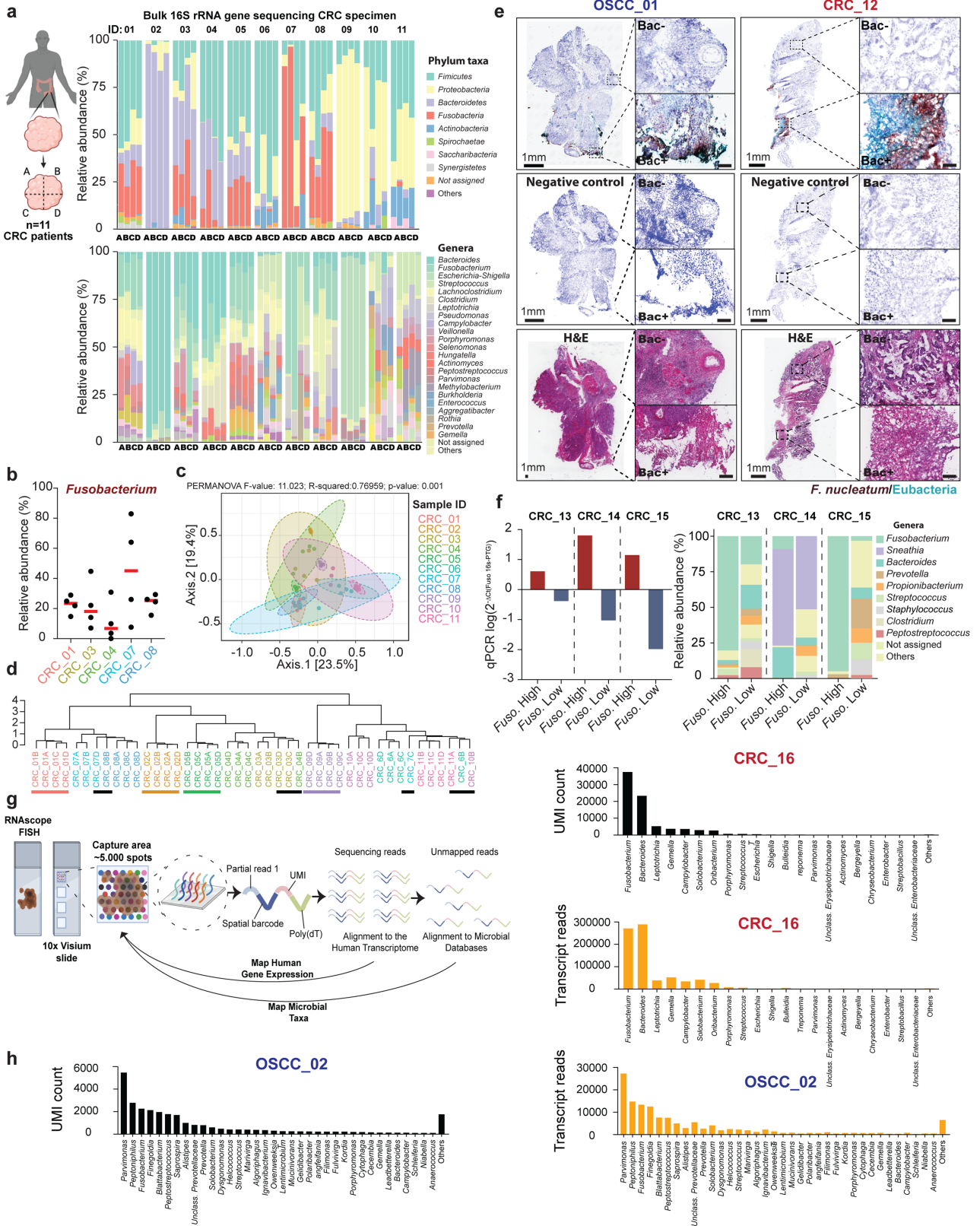
## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41586-022-05435-0>.

**Correspondence and requests for materials** should be addressed to Christopher D. Johnston or Susan Bullman.

**Peer review information** Nature thanks Raz Ben-Uri, Fuchou Tang and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

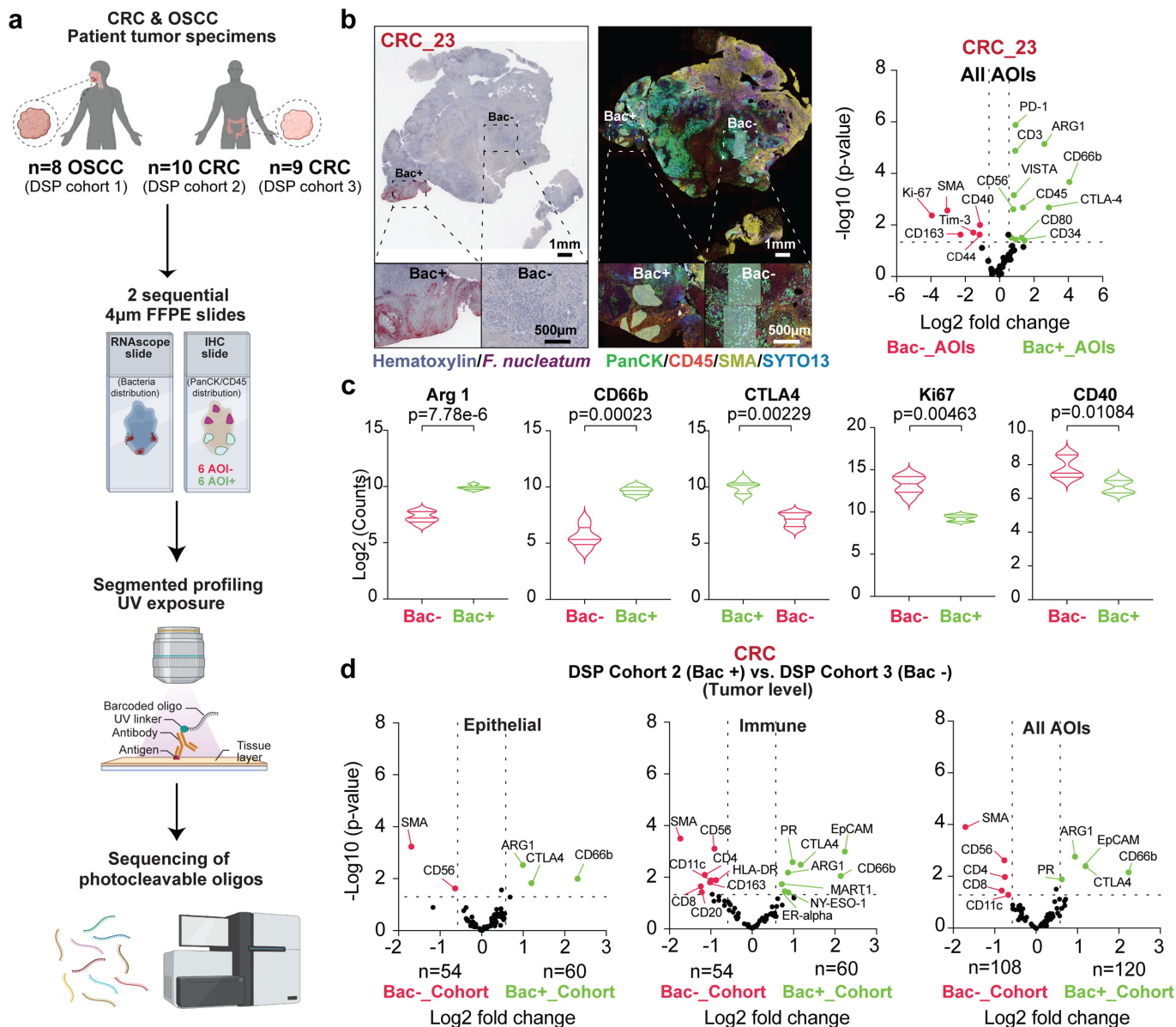
**Reprints and permissions information** is available at <http://www.nature.com/reprints>.



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

**Extended Data Fig. 1 | Heterogeneous distribution of the intratumoral microbiota throughout the tumour tissue.** **a**, Relative abundance of bacterial communities at the phylum and genus level for each tumour piece (n = 4 per patient) from 11 human CRC tumour specimens, identified via bulk 16S rRNA gene amplicon sequencing. Tumour tissue pieces (n = 4) are denoted as A, B, C and D from the 11 patients with CRC as indicated. **b**, Relative abundance of *Fusobacterium* genera in each tumour piece from 5 positive CRC specimens described in (a). **c**, Principal component analysis (PCoA) plot representing beta diversity clustering (Bray-Curtis Index) of bacterial communities at the genus level from each piece of CRC tumour tissue and PERMANOVA analysis. **d**, Dendrogram representing clustering of the microbiome composition at genus level in the tumour pieces as described in (a). The index of dissimilarity between samples was calculated using the Bray-Curtis test. Hierarchical cluster analysis was performed to detect clustering of patient specimens (colour bars) by using the Ward clustering algorithm. **e**, Top: RNAscope-FISH images indicating the spatial distribution of *F. nucleatum* and other bacterial communities (eubacteria) across the tumour tissue from a OSCC and CRC specimen. *F. nucleatum* probe is red and eubacterial probe is cyan. Middle: RNAscope-FISH images showing the negative control staining for the images. Bottom: Haematoxylin and eosin staining (H&E) of the RNAscope images. **f**, Validation of RNAscope probes. Left: PCR quantifying *Fusobacterium nucleatum* 16S rRNA

gene in macro-dissected tissue based on RNAscope probe binding containing relatively high (*Fuso. High*) or low (*Fuso. Low*) *F. nucleatum* positivity as it is indicated. Right: Microbiome analysis using 16S rRNA gene sequencing in tumour areas that are “*Fuso. High*” or “*Fuso. Low*” as it is indicated. **g**, Schematic showing the experimental approach: RNAscope imaging was implemented to identify tumour areas positive for bacteria or *F. nucleatum* from OSCC and CRC tumours embedded in OCT blocks. Tumour tissues were trimmed to fit the capture area (6.5 mm x 6.5 mm) on the 10x Visium slide. Following tissue permeabilization, RNA is released from cells and bind to an array of probes that are attached to the surface of the slide within capture spots. Each probe has a unique molecular identifier (UMI) and a barcode sequence providing the spatial coordinates for each transcript. cDNA is generated from the captured RNA through a reverse transcription reaction. The barcoded cDNA is denatured and pooled and then further processed to generate cDNA libraries. All transcripts are aligned against the human transcriptome to map the human gene-expression profile across the sample. The unmapped reads are then aligned against microbial databases through GATK PathSeq to identify the microbiome composition. **h**, Distribution of the bacterial UMI count and bacterial reads for top bacterial genera detected in 10x Visium data from the OSCC and CRC cases as it is indicated.



**Extended Data Fig. 2 | The tumour-associated microbiota resides in highly immunosuppressive microniche with a low proliferation rate.**

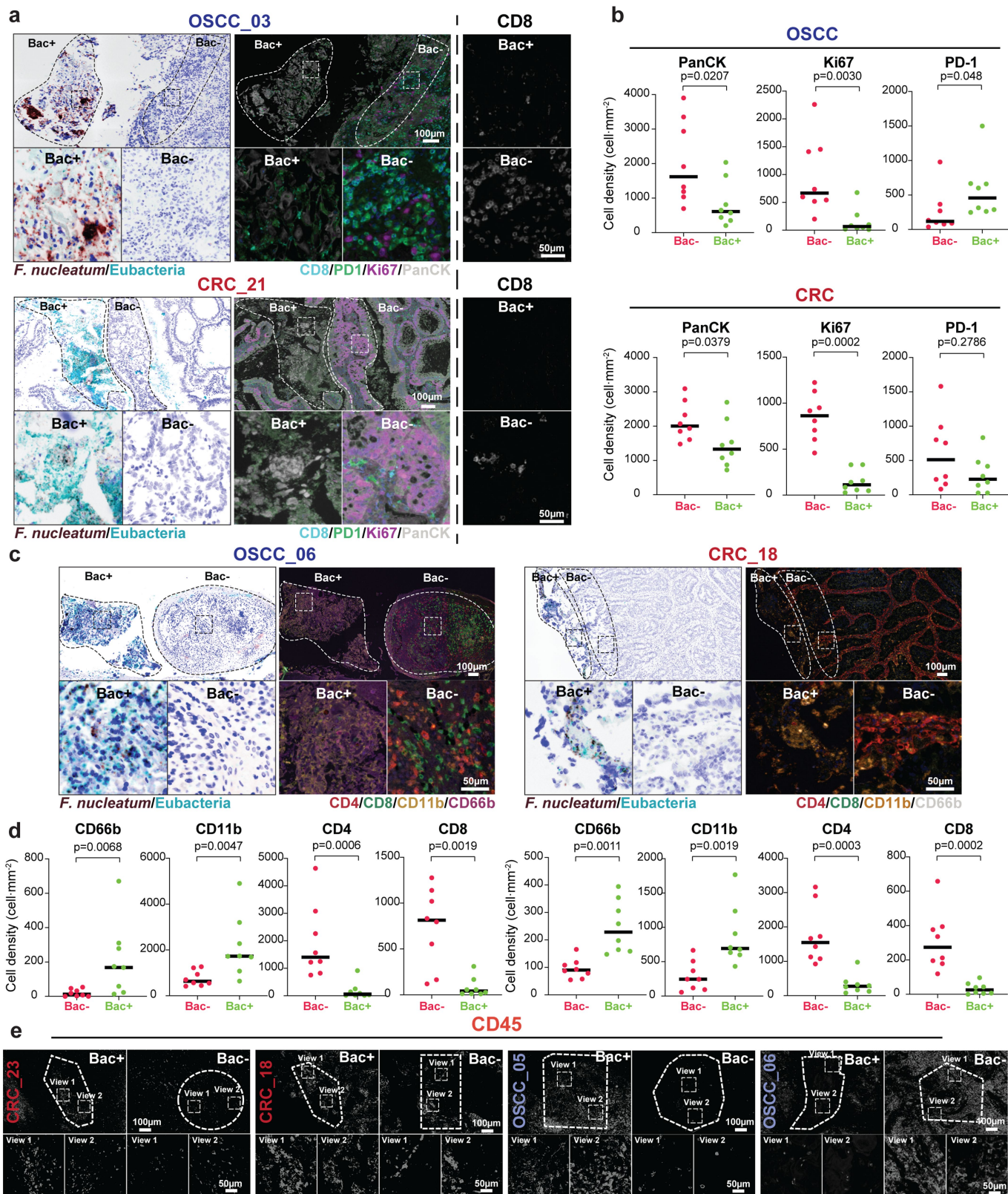
**a**, Experimental approach: GeoMx DSP was implemented to assess bacteria-associated microniche in one OSCC DSP cohort (n = 8) and two CRC DSP cohorts (RNAscope bacteria-positive cohort n = 10 and RNAscope bacteria-negative cohort n = 9). Sequential 4 µm-FFPE slides were prepared to identify spatial bacterial tumour distribution (RNAscope-CISH using *F. nucleatum* and eubacteria probes) and immunohistochemistry for immune (CD45+) and epithelial (PanCK+) compartments on the DSP slide treated with the 77-antibody panel. Segmented profiling for CD45+ and PanCK+ was performed on bacteria-positive AOIs (AOI<sub>bac+</sub>) and bacteria-negative AOIs (AOI<sub>bac-</sub>) per specimen, releasing photocleavable barcoded oligos for sequencing. Sequenced oligos provided the spatial information of the respective protein target in the bacteria positive or negative regions.

**b**, RNAscope-CISH (left) showing the distribution of *F. nucleatum* (dark red), throughout the tumour tissue from a CRC specimen. A sequential slide (right) showed the distribution of the immune (CD45+; red) and epithelial (PanCK+; green) compartments by IHC staining. Inset images indicated the AOIs that were selected for DSP analysis from a bacteria-positive (Bac+) and bacteria-negative (Bac-) regions as it is indicate. Volcano plot showed the differential

expression of genes from a single CRC sample comparing Bac+ with Bac- regions from the same tissue sample. Dashed lines indicate the threshold of significant gene expression defined as the Log2 fold change  $\geq 0.58$  and  $\leq -0.58$  with a  $-\text{Log}_{10}$  p value  $\geq 1.301$  following LMM analysis and Benjamini-Hochberg multiple-correction testing. A 52-antibody panel were included here (this did not include the Cell Death and MAPK modules applied to DSP cohorts 1–3).

**c**, Violin plots demonstrate the immuno-suppressive microenvironment in bacteria-positive regions (Bac+) from the sample described in (b), highlighting the upregulation of ARG1 and CTLA4 and the enrichment of myeloid CD66b+ cells with lower expression of Ki67 and the T cell co-stimulatory molecule CD40 compared to bacteria-negative regions. p values calculated by t-test.

**d**, Volcano plots indicate the differential gene-expression profile using the GeoMx DSP platform comparing AOIs from tumours (DSP cohort 2) that were RNAscope bacteria positive (Bac+; n = 120) against AOIs from tumours (DSP cohort 3) that were RNAscope bacteria negative (Bac-; n = 108). Using segmented analysis, the barcode oligos were collected either from the immune (CD45+) segment, epithelial (PanCK+) segment or both (All AOIs). Dashed lines indicate the threshold of significant gene expression defined as the Log2 fold change  $\geq 0.58$  and  $\leq -0.58$  with a  $-\text{Log}_{10}$  p value  $\geq 1.301$  following LMM analysis and Benjamini-Hochberg multiple-correction testing.

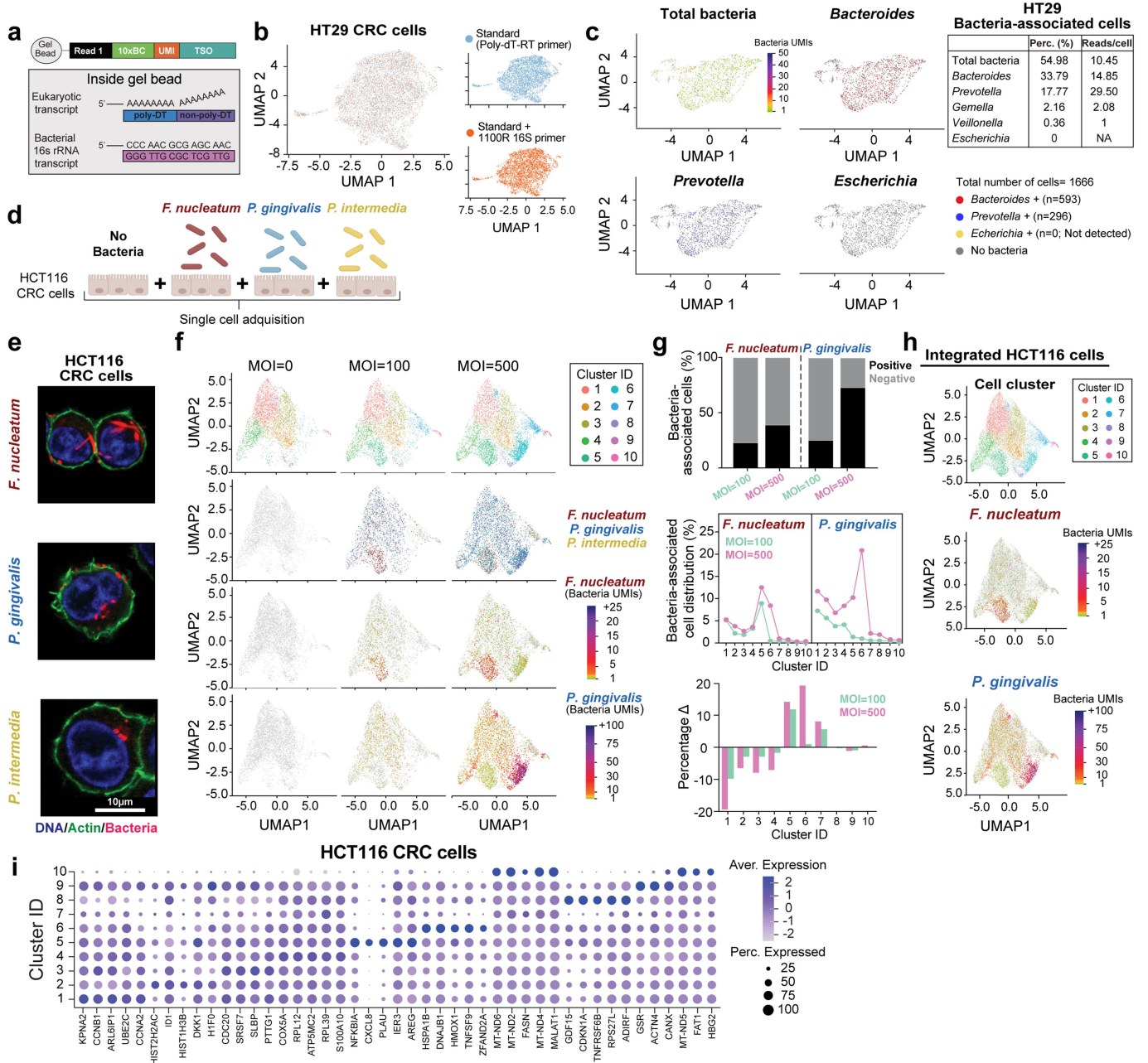


Extended Data Fig. 3 | See next page for caption.

# Article

**Extended Data Fig. 3 | The tumour-associated microbiota resides in T-cell-excluded areas, with lower proliferation capabilities. a,** RNAscope-FISH images showing bacteria-positive regions (Bac+) with the corresponding adjacent bacteria-negative region (Bac-) from a OSCC and CRC sample. A sequential IHC slide indicates the staining of CD8, PD-1, PanCK and Ki-67 in Bac+ and Bac- regions from the same tumour samples as it is indicated. Right panels indicate the CD8 fluorescent signal from a OSCC and CRC case comparing Bac+ vs Bac- regions from the same tumour tissue. **b,** Quantification of cell densities for PanCK, Ki-67 and PD-1, expressing cells in bacteria-positive regions in comparison to the contiguous bacteria-negative regions. p-values were calculated by Mann-Whitney test **c,** Representative RNAscope-FISH images showing bacteria-positive regions (Bac+) with the corresponding

adjacent bacteria-negative region (Bac-) from a OSCC and CRC sample (dashed areas within images). A sequential IHC slide indicates the distribution of immune cell populations including myeloid CD66b+ or CD11b+ cells and CD4+ or CD8+ T cells. Magnified/inset images show the immune cell population that is more abundant in bacteria-positive (Bac+) and bacteria-negative (Bac-) regions for each tumour sample as it is indicated. **d,** Quantification of cell densities of CD66b, CD11b, CD4 and CD8 expressing cells in Bac+ regions compared to the contiguous Bac- region from the same (n = 4) OSCC (left) and (n = 4) CRC (right) tumour samples in two separate field of views. p-values were calculated by Mann-Whitney test. **e,** IHC images showing the CD45 fluorescent signal from the tumour samples described in Fig. 2a.



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

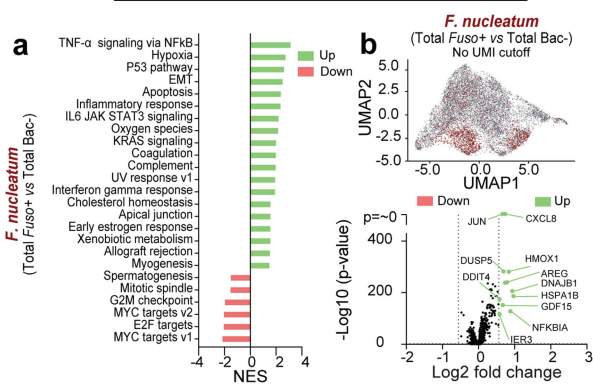
# Article

**Extended Data Fig. 4 | Detection of bacteria-associated single cells using the INVADeseq technique.** **a**, Schematic showing a modified gel bead emulsion (GEMs) by introducing a primer (1100R16S) that targets a conserved region of the bacterial 16S rRNA, thus allowing cDNA generation of bacterial transcripts from the associated human single cells. In addition to the standard 10x genomics 5' library preparation, bacterial cDNA was amplified with a nested conserved 16S rRNA gene primer and size selected libraries were sequenced and assessed through GATK PathSeq to identify bacterial taxa. Sequencing reads from the 16S rRNA amplified libraries retain the 10x genomics barcode sequence which facilitated mapping of annotated bacterial reads directly to the host single cells they are associated with. **b**, UMAP plots showing single-cell transcriptome of HT-29 cells with (orange dots) and without (blue dots) the 1100R16S primer in the amplification mix before single-cell cDNA generation. UMAP plot inserts show the transcriptome for each condition, indicating no differences in the human gene-expression profile when the 1100R16S primer was added. **c**, UMAP plots indicating single-cell transcriptome of HT-29 cells co-incubated with *Bacteroides fragilis*, *Prevotella intermedia*, *Gemella haemolysans*, *Veillonella parvula* and *Escherichia coli DH5a* for 3 h at MOI = 100. Insert table indicates the percentage of bacteria-associated single cells and total bacterial reads per cell per bacterial taxa. Note: *Escherichia coli DH5a* reads were not detected in human single cells. **d**, Experimental approach: HCT116 cancer cells were co-culture with either *Fusobacterium nucleatum*, *Porphyromonas gingivalis* or *Prevotella intermedia* at total multiplicity of infection (MOI) of 0, 100 and 500 for 3 hrs and processed

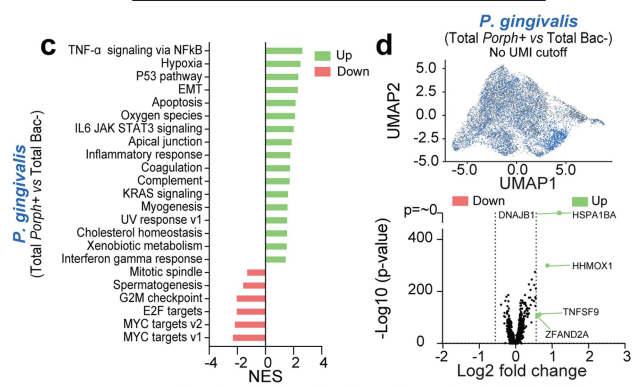
for INVADeseq. **e**, Confocal images showing intracellular bacteria in HCT116 cancer cells after 3 h of incubation with *F. nucleatum*, *P. gingivalis* and *P. intermedia* as it is indicated. **f**, From top to bottom, UMAP plots from scRNA-seq data showing: Cell cluster distribution based on epithelial cell transcription, infected HCT116 cancer cells with *F. nucleatum*, *P. gingivalis* and *P. intermedia*, and expression level (UMI, bacterial load) of *F. nucleatum* and *P. gingivalis* transcripts in cancer cells following bacteria treatment for 3 h at multiplicity of infection (MOI) 0, 100 and 500 as it is indicated. Cluster ID indicates a unique transcriptional cellular group predicted by Seurat package (See methods). Colour bars indicate the expression level (UMI counts) of *F. nucleatum* and *P. gingivalis* transcripts as it is indicated. **g**, Top: Percentage of bacteria-associated cells positive for either *F. nucleatum* or *P. gingivalis* at MOI = 100 and 500 as it is indicated. Middle: Distribution of *F. nucleatum* and *P. gingivalis*-associated cells across all cell cluster annotated in (f) whereby all cell clusters combined, bacteria positive and negative, equal 100%. Bottom: Relative change in the percentages ( $\Delta\%$ ) of cancer epithelial cell clusters between bacteria-associated cancer cells at MOI = 100 or 500 compared to the untreated control cell population (MOI = 0) for each cancer cell cluster annotated in (f). **h**, UMAP plots showing cancer epithelial cell clusters and detection of *F. nucleatum* and *P. gingivalis* transcripts following data integration from the experimental conditions described in (d). **i**, Dot plot showing the relative expression of gene markers for each ID cluster from CRC epithelial cells derived from the HCT116 cell line. Colour bars indicate the average expression level, and the dots represent the percentage expression level for each gene marker.



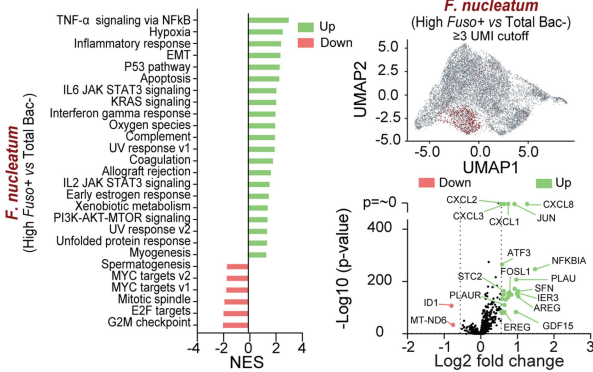
### All clusters single cell analysis HCT116



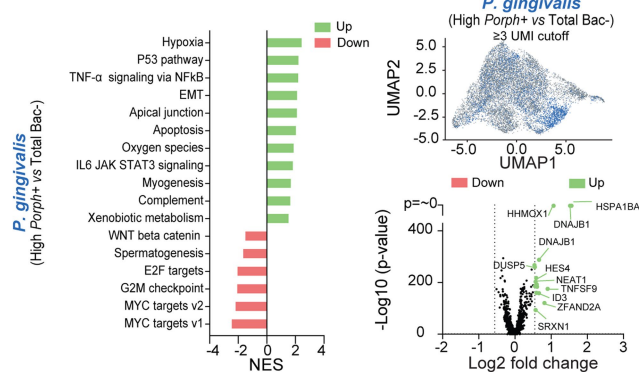
### All clusters single cell analysis HCT116



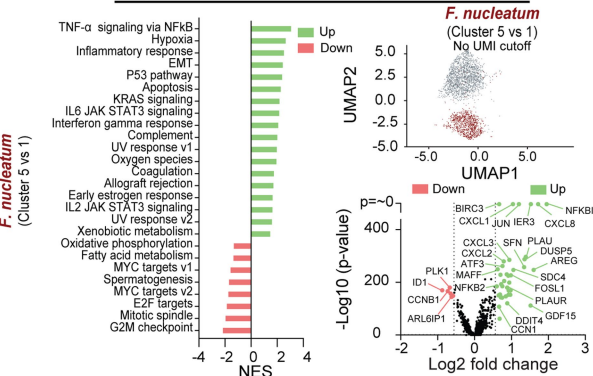
### *F. nucleatum* single cell analysis HCT116



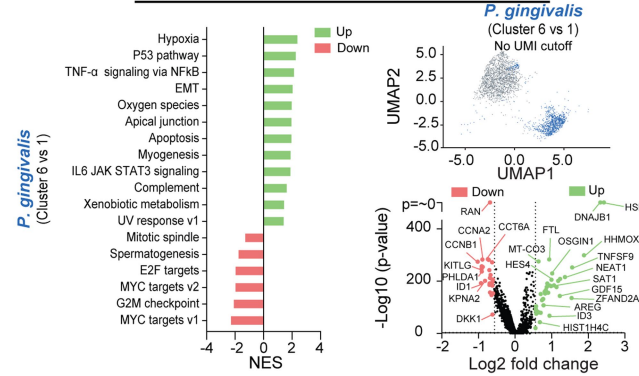
### *P. gingivalis* single cell analysis HCT116



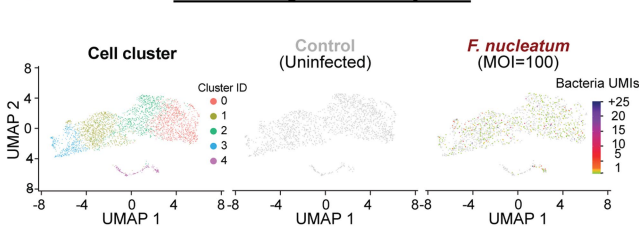
### Specific cell cluster analysis HCT116



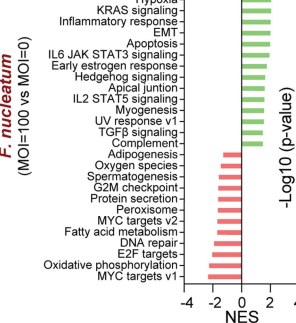
### Specific cell cluster analysis HCT116



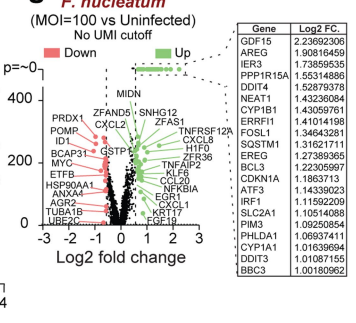
### e HT-29 Single Cell Analysis



### f



### g



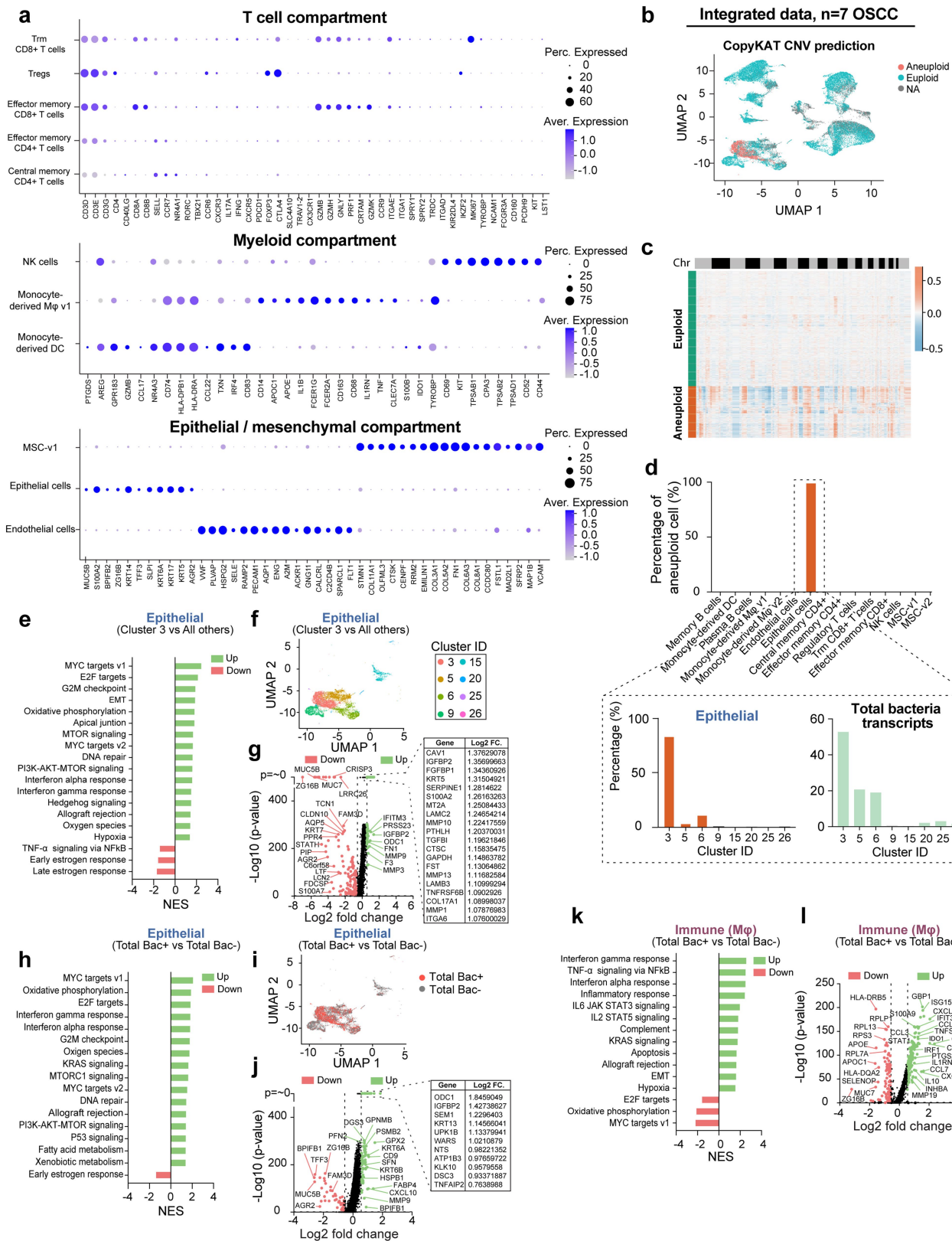
Extended Data Fig. 5 | See next page for caption.

# Article

## Extended Data Fig. 5 | Differential gene expression and GSEA analysis comparing distinct CRC single-cell groups on the basis of bacteria association.

**a.** GSEA analysis indicating the signalling pathways that are differentially regulated in HCT116 cells co-incubated with *Fusobacterium nucleatum* at MOI = 500 for 3 h between different single-cell groups as follow: Top: Total *F. nucleatum*-associated cells (Total *Fuso*+) compared to total bacteria-negative cells (Total Bac-) in the entire sample “All clusters single cell analysis”. Middle: *F. nucleatum*-associated cells that contain  $\geq 3$  *F. nucleatum* UMIs (High *Fuso*) compared to total bacteria-negative cells (Total Bac-) in the entire sample “*F. nucleatum* single cell analysis”. Bottom: *F. nucleatum*-associated cells from cell cluster 5 compared to bacteria-negative cells from cluster 1 “Specific cell cluster analysis”. **b.** UMAP plots showing the cell population that are highlighted for each analysis described in (a), in which the *F. nucleatum*-associated cells are coloured in red and the bacteria-negative cells are coloured in grey as it is indicated. Volcano plots indicate the differential gene-expression profile between the cell populations described in (a). Dashed lines indicate the threshold of significant gene expression defined as the  $\text{Log}_2$  fold change  $\leq -0.58$  and  $\geq 0.58$  with a  $-\text{Log}_{10}$  p value  $\geq 1.301$ . **c.** GSEA analysis indicates the signalling pathways that are differentially regulated in HCT116 cells co-incubated with *Porphyromonas gingivalis* at MOI = 500 for 3 h between different cellular groups as follow: Top: Total *P. gingivalis*-associated cells (Total *Porph*+) compared to total bacteria-negative cells (Total Bac-) in the entire sample “All clusters single cell analysis”. Middle: *P. gingivalis*-associated cells that contain  $\geq 3$  *P. gingivalis* UMIs (High *Porph*) compared to

total bacteria-negative cells (Total Bac-) in the entire sample “*P. gingivalis* single cell analysis”. Bottom: *P. gingivalis*-associated cells from cell cluster 6 compared to bacteria-negative cells from cluster 1 “Specific cell cluster analysis”. **d.** UMAP plots show the cell populations that are highlighted for each analysis described in (c), in which the *P. gingivalis*-associated cells are coloured in blue, and the bacteria-negative cell population are coloured in grey as it is indicated. Volcano plots indicate the differential gene-expression profile between the cell populations described in (c). Dashed lines indicate the threshold of significant gene expression defined as the  $\text{Log}_2$  fold change  $\leq -0.58$  and  $\geq 0.58$  with a  $-\text{Log}_{10}$  p value  $\geq 1.301$ . **e.** UMAP plots showing cell cluster distribution and the detection of bacterial transcripts in HT-29 cancer cells treated with or without *F. nucleatum* at MOI of 100 for 2 h. Colour bar indicates expression level (UMI counts) of *F. nucleatum*. **f.** GSEA analysis indicating the signalling pathways that are differentially regulated in HT-29 cancer cells treated with *F. nucleatum* compared to an uninfected control cancer cell population. **g.** Volcano plot showing the gene-expression profile in HT-29 cells treated with *F. nucleatum* relative to bacteria-negative cells. Dashed lines indicate the threshold of significant gene expression defined as the  $\text{Log}_2$  fold change  $\geq 0.58$  and  $\leq -0.58$  with a  $-\text{Log}_{10}$  p value  $\geq 1.301$ . Wilcoxon Rank Sum test was implemented to calculate the normalized enrichment score (NES) in panels (a), (c) and (f). A LMM analysis followed by Benjamini–Hochberg multiple-correction test was used to calculate the fold change and p-values for each gene in panels (b), (d) and (g).

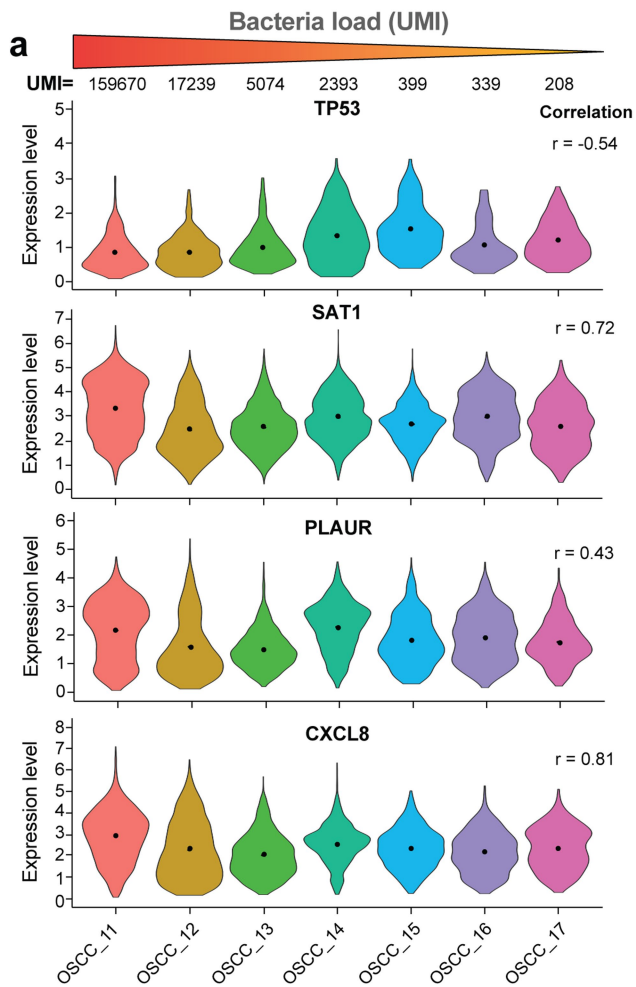


Extended Data Fig. 6 | See next page for caption.

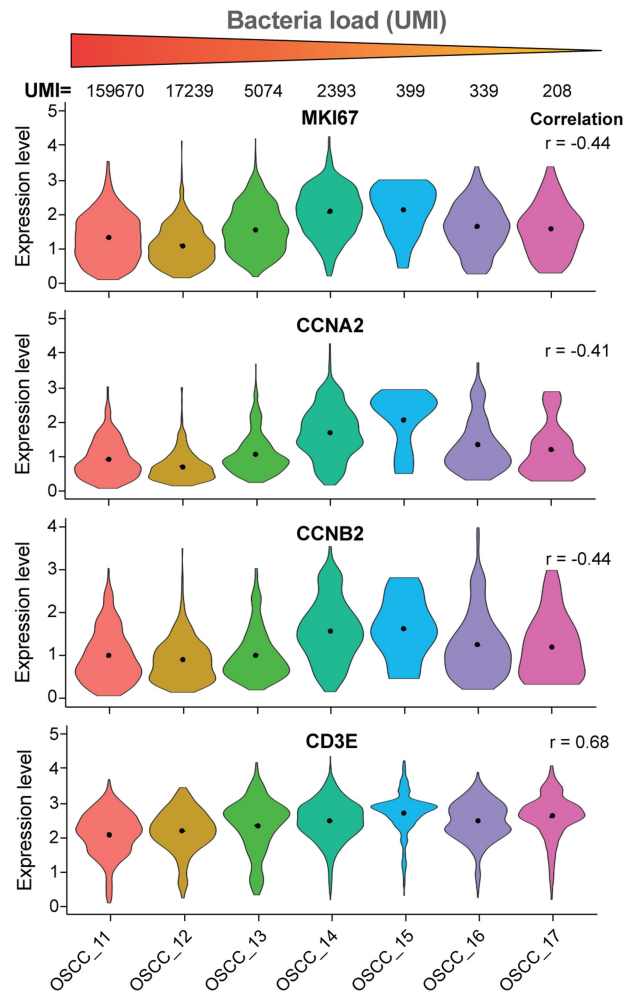
# Article

**Extended Data Fig. 6 | Bacteria-associated single cells correlate with upregulation of cell inflammation and cancer progression pathways in patients with OSCC.** **a**, Dot plot showing the relative expression of gene markers for the T cell, myeloid and epithelial/ mesenchymal cell compartment from single-cell data from patients with OSCC (n = 7 patients) as it is indicated in the cell clusters annotated in the UMAP plot from Fig. 3c. Colour bars indicate the average expression level, and the dots represent the percentage of expression for each gene marker. **b**, UMAP plot indicates the distribution of aneuploid and euploid cells in samples from patients with OSCC. **c**, Heat map shows the copy number variations (CNV) across 41,723 cells from the epithelial cell cluster as it is indicated. **d**, Percentage of aneuploid cells for each cell type annotated in Fig. 3c. Insert tables indicate the percentage of aneuploid cells and bacterial transcripts (UMIs) for each sub-cell clusters in the epithelial cell cluster as it is indicated. **e**, GSEA analysis showing the signalling pathways that are differentially regulated in cells from the sub-cell cluster 3, which contains the majority of aneuploid cancer cells, relative to other sub-cell clusters (All others) from the epithelial cell cluster. **f**, UMAP plot highlighting the epithelial sub-clusters that are contained in the epithelial cell cluster detected in Fig. 3c. **g**, Volcano plot showing the gene-expression profile in cancer cells from sub-cell cluster 3 in comparison to other sub-cell clusters from the epithelial cell cluster. Dashed lines indicate the threshold of significant gene expression

defined as the Log2 fold change  $\geq 0.58$  and  $\leq -0.58$  with a  $-\text{Log}_{10}$  p value  $\geq 1.301$ . **h**, GSEA analysis showing the signalling pathways that are differentially regulated comparing total bacteria-associated cells (Total Bac+) vs bacteria-negative cells (Total Bac-) in the epithelial single cells as described in Fig. 3c. **i**, UMAP plot highlighting the transcriptional cellular group that is associated with bacterial invasion (Total Bac+; red cells) along with the bacteria-negative cell population (Total Bac-; grey cells). **j**, Volcano plot indicates the gene-expression profile in total bacteria-associated cell (Total Bac+) relative to total bacteria-negative cells (Total Bac-) in the epithelial single-cell cluster described in Fig. 3c. **k**, GSEA analysis showing the signalling pathways that are differentially regulated comparing total bacteria-associated cells (Total Bac+) vs bacteria-negative cells (Total Bac-) in the macrophage single cells as described in Fig. 3c. **l**, Volcano plot indicates the gene-expression profile in total bacteria-associated cell (Total Bac+) relative to total bacteria-negative cells (Total Bac-) in the macrophage cell cluster described in Fig. 3c. Dashed lines indicate the threshold of significant gene expression defined as the Log2 fold change  $\geq 0.58$  and  $\leq -0.58$  with a  $-\text{Log}_{10}$  p value  $\geq 1.301$ . Wilcoxon Rank Sum test was implemented to calculate the normalized enrichment score (NES) in panels (e), (h) and (k). ALMM analysis followed by Benjamini-Hochberg multiple-correction test was used to calculate the fold change and p-values for each gene in panels (g), (j) and (l).

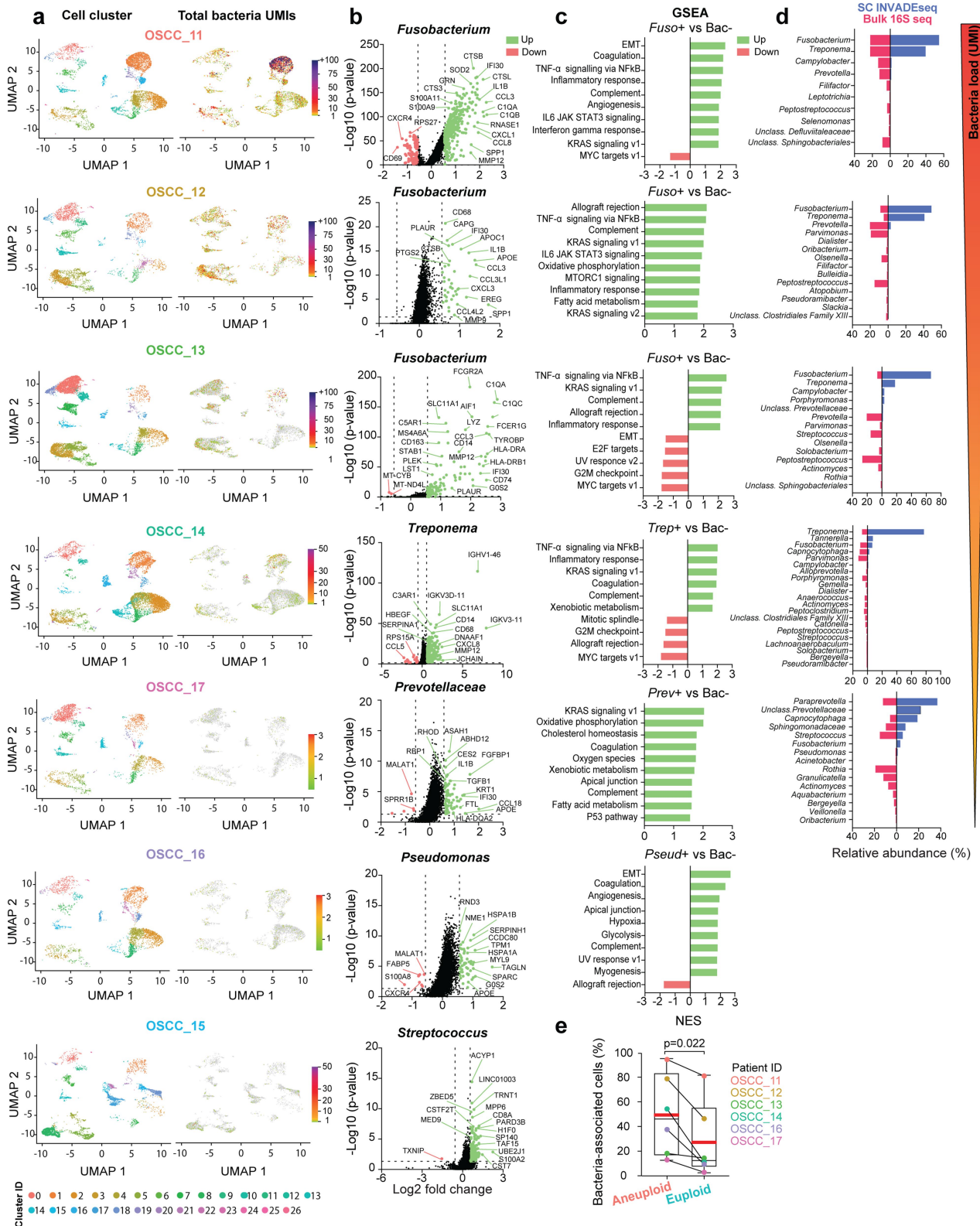


**Extended Data Fig. 7 | Expression levels of genes that are correlated or anticorrelated with bacterial load (UMI) in patients with OSCC.**  
**a**, Expression level of genes as a function of single-cell-associated bacterial load (UMI) in seven patients with OSCC. The colour bars (top) indicate a



gradient of bacterial load across the samples; the bacterial UMI values are designated for each patient. The correlation values were measured for each gene based on mean gene expression and bacterial UMI counts as it is indicated.

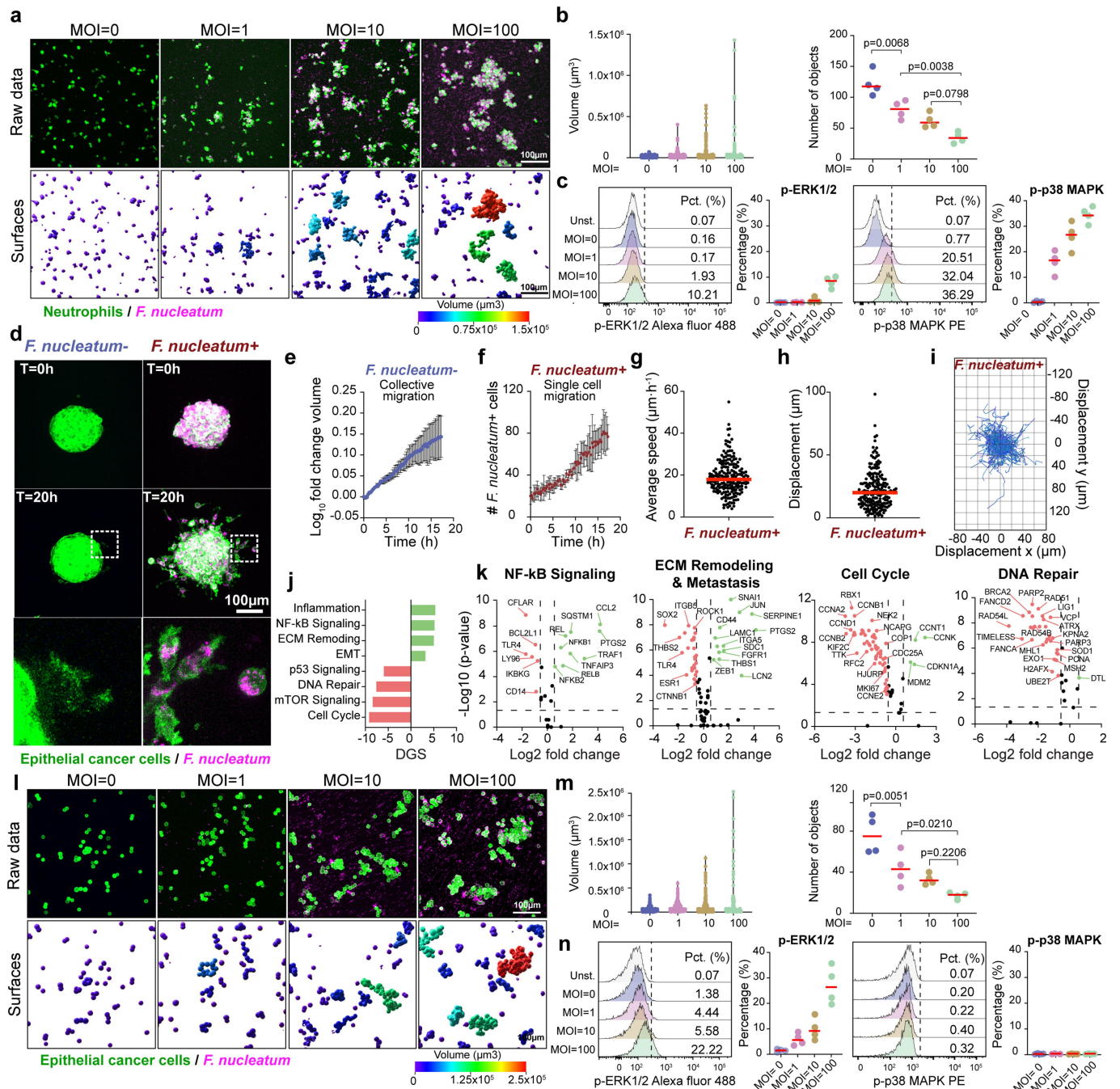
All Cluster Analysis: Individual Patient OSCC



Extended Data Fig. 8 | See next page for caption.

**Extended Data Fig. 8 | Differential gene expression, GSEA and microbiome analysis for each individual patient with OSCC.** **a**, UMAP plots showing host cell annotation and detected bacterial transcripts (UMI) from each patient with OSCC as indicated. **b**, Volcano plots indicate the differential gene-expression profile comparing total bacteria-associated cells against total bacteria-negative cells in the entire sample for the dominant bacterial genera for each patient with OSCC; a  $\geq 3$  bacteria UMI cut-off was applied for samples OSCC\_11, OSCC\_12, OSCC\_13 and OSCC\_14. Dashed lines indicate the threshold of significant gene expression defined as the Log<sub>2</sub> fold change  $\geq 0.58$  and  $\leq -0.58$  with a  $-\text{Log}_{10}$  p value  $\geq 1.301$  following LMM analysis and Benjamini–Hochberg multiple-correction testing. **c**, GSEA analysis showing the top 10 signalling pathways that are differentially regulated comparing total bacteria-associated cells against total bacteria-negative cells in the entire sample for the respective dominant bacterial species for each patient with OSCC as indicated. A  $\geq 3$  bacteria UMI cut-off was applied for samples OSCC\_11, OSCC\_12, OSCC\_13 and

OSCC\_14. No differentially regulated pathways were detected in OSCC\_15 based on *Streptococcus* positive and negative cells. **d**, Microbiome analysis using the INVADeSeq technique (SC INVADeSeq) on single cells and bulk 16S rRNA gene sequencing (Bulk 16S seq) on tissue homogenate from the same dissociated tissue piece for 5 of the patients with OSCC. Genera with  $\geq 1\%$  relative abundance in either SC INVADeSeq or Bulk 16S seq data are shown. The coloured bar (right) indicates the bacterial load (UMI count) in which the OSCC samples were arranged from the highest to lowest based on the bacterial UMI count **e**, Percentages of bacteria-associated cells in the aneuploid and euploid enriched epithelial cell clusters for each patient with OSCC (data points), 6 patients that contained both aneuploid and euploid epithelial cells are included. Box-whiskers indicate medians and the interquartile range. The box represents the middle 50% of scores for each group. Red bars indicate the mean of the combined data. p-values indicates statistical significance calculated by two-tailed paired t test.



Extended Data Fig. 9 | See next page for caption.



**Extended Data Fig. 9 | *F. nucleatum* induces the formation of cell clusters in immune and cancer cells.** **a**, Confocal images showing cluster formation of differentiated neutrophils derived from human HL-60 cells (green) co-cultured with *F. nucleatum* (pink) at different multiplicity of infection (MOI). Top micrographs display the raw imaging data. Bottom micrographs display the corresponding mask surfaces for each experimental condition by using Imaris software. Colour bar indicates the size (volume  $\mu\text{m}^3$ ) of the objects. **b**, Left: Violin plot indicates the quantification of volume of individual neutrophil clusters (data points) in the present of *F. nucleatum* as shown in (a); combined data from 4 independent experiments. Right: Dot plot shows the number of neutrophil clusters (objects) per field of view as indicated in (a). Data points represent the number of cell objects for each independent experiment ( $n = 4$ ). p-values were calculated by one-way ANOVA followed by Bonferroni multiple comparison test. **c**, Flow cytometry plots show the levels of phosphorylation of ERK and p38 MAPK in neutrophils treated with *F. nucleatum*. Corresponding dot plots indicate the level of phosphorylation for each independent experiment (data points;  $n = 4$ ). **d**, Spheroids derived from a mouse CRC cell line CT26WT were treated with or without *F. nucleatum* for 12 h and then embedded in collagen matrices. The cell invasion capabilities for both conditions were evaluated using live-cell confocal imaging over a period of 19 h. Amplified images show the difference in the migration mode from both conditions. **e**,  $\text{Log}_{10}$  fold change volume over time of uninfected CRC spheroids revealing the expansion rate of uninfected control spheroids. Data points indicate the average values from three independent experiments. Error bars indicate the standard deviation (SD). **f**, Number of *F. nucleatum*-positive cancer cells that

detached from the spheroid mass as single motile cells from three independent experiments. Errors bars indicate the SD. **g-h**, Distribution of the average speed and cell displacements of single cells that escape the spheroid mass infected with *F. nucleatum*. Combined data from three independent experiments. Red bars indicate the mean. Data points represents individual tracks; "n" indicates number of tracks per condition. **i**, Cell trajectories from an origin point of invading cancer cells that escape the spheroids infected with *F. nucleatum*. **j**, Signalling pathway analysis of CRC spheroids infected with *F. nucleatum* in comparison to uninfected control spheroids. **k**, Volcano plots showing the regulation of genes in selected signalling pathways in spheroids infected with *F. nucleatum* in comparison to uninfected spheroids. Dashed lines indicate the threshold of significant gene expression defined as the  $\text{Log}_2$  fold change  $\leq -0.58$  and  $\geq 0.58$  with a  $-\text{Log}_{10}$  p value  $\geq 1.301$  following LMM analysis and Benjamini-Hochberg multiple-correction testing. **l**, Confocal images showing cell cluster formation of cancer cells derived from the human cell line HCT116 treated with *F. nucleatum* at different MOI as it is indicated. **m**, Left: Violin plot indicates the quantification of volume of individual cancer cell clusters (data points) in the present of *F. nucleatum* as shown in (l); combined data from 4 independent experiments. Right: Dot plot shows the number of cancer cell objects per field of view as indicated in (l). Data points represent the number of clusters for each independent experiment ( $n = 4$ ). p-values were calculated by one-way ANOVA followed by Bonferroni multiple comparison test. **n**, Flow cytometry plots show the levels of phosphorylation of ERK and p38 MAPK in cancer cells treated with *F. nucleatum*. Corresponding dot plots indicate the level of phosphorylation for each independent experiment (data points;  $n = 4$ ).

## 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

BD FACSDiva™ Software  
GeoMx Digital Spatial Profiler v2.1 Instrument Software  
Leica Application Suite X (LAS X)  
NextSeq 1000/2000 Control Software Suite v1.2.0

## Data analysis

FlowJo v9  
 Imaris Cell Imaging Software (Bitplane) v9.2.0  
 nSolver™ Analysis Software v.4 (NanoString)  
 GraphPad Prism 7.0 Software  
 Halo Image Analysis Software v3.4 (Indica Labs)  
 R packages: ClusterProfiler v3.18.1, CopyKAT v1.0.5, Harmony v1.0, Seurat v4.0.4, SingleR v1.4.1  
 Python package: Pysam 0.16.0.1  
 CellRanger v6.1.1 (10x Genomics)  
 GATK PathSeq v4.1.3.0 Pathogen discovery pipeline (Broad institute)  
 SpaceRanger v1.3.0 (10x Genomics)  
 Trimmomatic v0.39  
 MicrobiomeAnalyst

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](#)

Codes for data processing and analysis is available at <https://github.com/FredHutch/Galeano-Nino-Bullman-Intratumoral-Microbiota-2022>. Raw sequence data from microbiome analysis using bacterial 16S ribosomal RNA gene sequencing, 10x Visium spatial transcriptomics and single-cell RNA sequences were deposited in the NCBI Sequence Read Archive (SRA) under the Bioproject accession number PRJNA811533. PathSeq, Cell Ranger and Space Ranger analyses used GRCh38 as the human genome reference.

## 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://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

Patient specimen sample size was determined according to biological material available. For the GeoMx analysis, 12 areas of interest were selected per patient specimen and patient cohorts were analyzed independently as stated in the figure legends and methods. For the INVADeseq approach validation with the HCT116 cell line (MOI=0,100 and 500) we captured a total of 12,412 single cells and all cells that passed quality filtering (as detailed in the data exclusions) were included in the analysis. For the INVADeseq approach with the patient OSCC specimens, 42,810 single cells were captured from seven patients, all cells that passed quality filtering (as detailed in the data exclusions) were included in the analysis. For the 10x Visium OSCC and CRC cases, data was obtained from 3404 and 3106 capture spots within the tissue area. Statistical tests were performed for all the analyses as indicated in the figure legends and methods.

### Data exclusions

For single-cell RNA sequencing, low quality events were removed by filtering transcripts that were expressed in less than three cells (min.cells feature from Seurat R package) and cells that contain less than 200 unique transcripts (min.feature option from Seurat R package). For the INVADeseq approach validation with the HCT116 cell line (MOI=0,100 and 500), of the total 12,412 single cells captured, 39 single cells were excluded as they did not pass the quality filter as detailed in the methods. For the INVADeseq approach with the seven patient OSCC specimens, of 42,810 total single cells captured, 1087 single cells did not pass the quality filters (as detailed in the methods) and were excluded from analyses. For 10x visium spatial transcriptomics, capture spots that contain less than 3 unique transcripts were removed from the analysis by min.feature option from Seurat R package.

### Replication

A total of 11 fresh-frozen colorectal patient cancers (CRC) samples were selected for microbiome bulk analysis using 16S ribosomal RNA gene sequencing. Microbiome composition and bacteria distribution using 10x Visium spatial transcriptomics were measured in a CRC and oral squamous cell carcinoma (OSCC) case embedded in OCT medium. An additional 18 CRC and 8 OSCC FFPE embedded samples were collected for GeoMx-DSP analysis. A total of 7 fresh OSCC samples were processed for single-cell RNA sequencing. For in-vitro functional assays the experiments were conducted at least three times for data reproducibility. All replicates successfully show consistent results.

### Randomization

Sample randomization into experimental groups are not relevant, as the design of the study aims to quantify discriminative features between already established groups.

### Blinding

Blinding is not relevant for this study since the aim of this work is to quantify discriminative features between already established experimental

## 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 checked="" type="checkbox"/>	<input 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

anti-EpCAM, Mouse mAb, BioLegend, Cat: 32402, Clone: 9C4  
 anti-CD66b, Mouse mAb, BD Biosciences, Cat: 555723, Clone: G10F5  
 anti-CD11b, Rabbit mAb, BioSB, Cat: BSB6440, Clone: EP45  
 anti-CD4, Rabbit mAb, CellMarque, Cat: 104R-26 (AC-0173), Clone: EP204  
 anti-CD8, Mouse mAb, Dako, Cat: M7103, Clone: 144B  
 Anti-Phospho-p44/42 MAPK (Erk1/2) (Thr202/Tyr204) Alexa Fluor® 488 Conjugate Mouse mAb, Cell signaling, Cat: #4374, Clone: (E10)  
 Anti-Phospho-p38 MAPK (Thr180/Tyr182) PE Conjugate, Rabbit mAb, Cell signaling, Cat: #6908, Clone: (E10)  
 Anti-PD1, Rabbit mAb, Abcam, Cat: ab137132, Clone: EPR4877(2)  
 Anti-Ki67, Mouse, mAb, Dako, Cat: M7240, Clone: MIB-1  
 Anti-Panck, Mouse, mAb, Dako, Cat: M35153, Clone: AE1/AE3

### Validation

All antibodies used in the study were validated by the manufacture. Data is available at the manufacture's website as it is indicated below:

anti-EpCAM, Mouse mAb, BioLegend, Cat: 32402, Clone: 9C4: <https://www.biolegend.com/en-gb/products/purified-anti-human-cd326-epcam-antibody-3755?GroupID=BLG5134>  
 anti-CD66b, Mouse mAb, BD Biosciences, Cat: 555723, Clone: G10F5: <https://www.bdbiosciences.com/en-us/products/reagents/flow-cytometry-reagents/research-reagents/single-color-antibodies-ruo/purified-mouse-anti-human-cd66b.555723>  
 anti-CD11b, Rabbit mAb, BioSB, Cat: BSB6440, Clone: EP45. <https://www.biosb.com/biosb-products/cd11b-antibody-rmab-ep45/>  
 anti-CD4, Rabbit mAb, CellMarque, Cat: 104R-26 (AC-0173), Clone: EP204. [https://www.cellmarque.com/antibodies/CM/2186/CD4\\_EP204](https://www.cellmarque.com/antibodies/CM/2186/CD4_EP204)  
 anti-CD8, Mouse mAb, Dako, Cat: M7103, Clone: 144B. [https://www.agilent.com/en/product/immunohistochemistry/antibodies-controls/primary-antibodies/cd8-\(concentrate\)-76631](https://www.agilent.com/en/product/immunohistochemistry/antibodies-controls/primary-antibodies/cd8-(concentrate)-76631)  
 anti-Phospho-p44/42 MAPK (Erk1/2) (Thr202/Tyr204) Alexa Fluor® 488 Conjugate Mouse mAb, Cell signaling, Cat: #4374, Clone: (E10) <https://www.cellsignal.com/products/antibody-conjugates/phospho-p44-42-mapk-erk1-2-thr202-tyr204-e10-mouse-mab-alexa-fluor-488-conjugate/4374>  
 Anti-Phospho-p38 MAPK (Thr180/Tyr182) PE Conjugate, Rabbit mAb, Cell signaling, Cat: #6908, Clone: (E10) <https://www.cellsignal.com/products/antibody-conjugates/phospho-p38-mapk-thr180-tyr182-3d7-rabbit-mab-pe-conjugate/6908>  
 Anti-PD1 antibody, Abcam, Cat: (ab137132) Clone: [EPR4877(2)] <https://www.abcam.com/pd1-antibody-epr48772-ab137132.html>  
 Anti-Ki67, Mouse, mAb, Dako, Cat: M7240, Clone: MIB-1 [https://www.agilent.com/en/product/immunohistochemistry/antibodies-controls/primary-antibodies/ki-67-antigen-\(dako-omnis\)-76239](https://www.agilent.com/en/product/immunohistochemistry/antibodies-controls/primary-antibodies/ki-67-antigen-(dako-omnis)-76239)  
 Anti-Panck, Mouse, mAb, Dako, Cat: M35153, Clone: AE1/AE3, [https://www.agilent.com/en/product/immunohistochemistry/antibodies-controls/primary-antibodies/cytokeratin-\(concentrate\)-76562](https://www.agilent.com/en/product/immunohistochemistry/antibodies-controls/primary-antibodies/cytokeratin-(concentrate)-76562)

## Eukaryotic cell lines

Policy information about [cell lines](#)

### Cell line source(s)

HT-29, HCT-116 and CT26WT were purchased from the American Type Culture Collection (ATCC)  
 HL60 cells expressing GFP-firefly luciferase construct were kindly given by Prof. Cameron Turtle (originally from the ATCC)

### Authentication

None of the cell lines used were authenticated.

### Mycoplasma contamination

Mycoplasma testing was performed independently by the Research cell bank facility at the Fred Hutch using the MycoProbe

Mycoplasma contamination	Mycoplasma Detection Kit (R&D systems) that can detect the 16S ribosomal RNA of the most common strains of mycoplasma. All cell lines used in this study tested negative for Mycoplasma.
Commonly misidentified lines (See <a href="#">ICLAC</a> register)	No commonly misidentified cell lines were used in the study

## Human research participants

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

Population characteristics	All patients included in this analysis were adults diagnosed with either colorectal adenocarcinoma or oral cavity squamous cell carcinoma. A major criteria for inclusion of specimens in this study was that patients were treatment naive at the time of tumor surgical resection. Gender, age or ethnicity of patients were not an exclusion factor as we did not perform analyses based on patient metadata.
Recruitment	All CRC specimens in this study were obtained from bio-repositories and no patients were recruited for this study. The fresh OSCC specimens were obtained from patients with oral cavity squamous cell carcinoma at the University of Washington medical center. These patients had consented to an excess specimen protocol for research, whereby tissue that was obtained under this protocol did not impact patient treatment or outcome. A criteria for patient selection was that they were treatment naive at the time of surgical resection and HPV negative. Age, gender or ethnicity did not impact selection criteria as these factors did not impact our analysis.
Ethics oversight	The use of patient specimens for this work was approved by the Fred Hutchinson Cancer Center IRB under the following protocol numbers RG #: 1006552, 1006974

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

## Flow Cytometry

### 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	For intracellular staining, a total of 10000 cells were incubated LIVE/DEAD fixable Aqua Dead cell staining for 30min on ice. Following incubation, the cells were fixed for 30min and permeabilize using the BD Cytofix/Cytoperm plus kit. The permeabilization solution contained a cocktail of antibodies against phosphoproteins including 1.25µg/ml Phospho-p44/42 MAPK (Erk1/2) (Thr202/Tyr204) (E10) Mouse mAb (Alexa Fluor 488 conjugate) and 1.25µg/ml Phospho-p38 MAPK (Thr180/Tyr182) (3D7) Rabbit mAb (Phycoerythrin conjugate). Following 45min of antibody incubation the cells were washed twice with the permeabilization solution. Cells were resuspended in 200µl FACS wash buffer and 1x10 <sup>4</sup> viable cells (LIVE/DEAD aqua negative cells) were acquired on the BD Fortessa x20 flow cytometer. A tube only containing F. nucleatum was also acquired to gate-out undesired bacteria cell events from the analysis. The geometric mean fluorescence intensity (MFI) was calculated for each phosphoprotein and for each experimental condition relative to an unstained control condition. Flow cytometry data were analyzed with FlowJo software.
Instrument	BD Fortessa x20 flow cytometer
Software	Collection software: BD FACSDiva™ Software Analysis software: FlowJo™ v10.8.1
Cell population abundance	A total of 10000 viable cells events were acquired for each experimental condition
Gating strategy	In all assays, cells were gate for live, single cells and then gated according to the specific staining. A tube only containing F. nucleatum was also acquired to gate-out undesired bacteria cell events from the analysis.

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