npj | biofilms and microbiomes

Article

Published in partnership with Nanyang Technological University

https://doi.org/10.1038/s41522-025-00717-7

Fusobacterium nucleatum is enriched in invasive biofilms in colorectal cancer

Jessica Queen[1], Zam Cing[2], Hana Minsky[1], Asmita Nandi[1], Taylor Southward[1], Jacqueline Ferri[1], Madison McMann[1], Thevambiga Iyadorai[3], Jamuna Vadivelu[3], April Roslani[3], Mun Fai Loke[3], Jane Wanyiri[1], James R. White[4], Julia L. Drewes[1] ✉ & Cynthia L. Sears[1] ✉

Fusobacterium nucleatum is an oral bacterium known to colonize colorectal tumors, where it is thought to play an important role in cancer progression. Recent advances in sequencing and phenotyping of *F. nucleatum* have revealed important differences at the subspecies level, but whether these differences impact the overall tumor ecology, and tumorigenesis itself, remain poorly understood. In this study, we sought to characterize *Fusobacteria* in the tumor microbiome of a cohort of individuals with CRC through a combination of molecular, spatial, and microbiologic analyses. We assessed for relative abundance of *F. nucleatum* in tumors compared to paired normal tissue, and correlated abundance with clinical and pathological features. We demonstrate striking enrichment of *F. nucleatum* and the recently discovered subspecies *animalis* clade 2 (Fna C2) specifically in colon tumors that have biofilms, highlighting the importance of complex community partnerships in the pathogenesis of this important organism.

Colorectal cancer (CRC) is a leading cause of cancer morbidity and mortality, with 1.8 million new cases and >800,000 deaths annually[1]. Microbial dysbiosis is increasingly recognized as a factor in colon tumor initiation and progression. This dysbiosis can manifest as enrichment or depletion of specific bacterial taxa or in alterations of bacterial communities, such as the development of mucus-invasive biofilms[2]. *Fusobacterium nucleatum*, a Gram-negative anaerobe and common member of the human oral microbiota, is known to be associated with CRC[3–6]. Clinical studies from North American, European, and Asian cohorts have demonstrated enrichment of *F. nucleatum* in a subset of CRC compared to healthy colon tissues[3–6]. Data from animal models and translational studies have suggested potential roles for this organism in colon tumor development, progression, metastasis and/or treatment response[7]. In the mouth, *F. nucleatum* plays an important role as a bridging species linking early and late colonizers of polymicrobial biofilms[8]. Biofilms have also been recognized as common features of colon tumors[9,10], where it is posited that *F. nucleatum* and other organisms invade into the mucosa to contact host cells and promote inflammation and tumorigenesis. *F. nucleatum* is a heterogenous species consisting of four subspecies anticipated to become distinct species: *animalis*, *nucleatum*, *polymorphum*, and *vincentii*[11]. Recently, subspecies *animalis* has been further classified into two clades, with identification of clade Fna C2 as prominently featured in the colon tumor microenvironment[12]. In this study, we sought to characterize the tumor microbiome of a large cohort of individuals with CRC through a combination of sequencing and spatial analysis, and to identify features associated with enrichment of *F. nucleatum* and its subspecies and clades.

Results

Colon tumors frequently have polymicrobial biofilms containing Fusobacteria

The study population consisted of 116 individuals diagnosed with colorectal cancer in Malaysia, from whom paired tumor and distant normal colon tissues were collected at the time of CRC tumor resection surgery (Fig. 1a). Four of these individuals underwent surgery for resection of masses later identified as adenomas (polyps). This 116 person cohort is inclusive of two previously published smaller cohorts (MAL1 and MAL2)[10], to which we added an additional 72 individuals with CRC, yielding an assessment of 143 additional tumor and/or paired normal samples (MAL3). Demographic and clinical data collected included age, sex, ethnicity, tumor location, and tumor stage (Fig. 1b). The proximal colon through the hepatic flexure was defined as right colon, and distal to the hepatic flexure as left colon, as previously described[10]. To assess for the presence of biofilms, we performed fluorescence in situ hybridization (FISH) on methacarn-fixed tissues of colon tumor and paired normal samples (Fig. 2a). We found that biofilms were prevalent on colon tumors, visualized in 68.8% (77/111) (Fig. 2b). Biofilms were more frequently observed on right-sided tumors (33/36; 91.7%) compared to left (44/75; 58.7%; $p = 0.0003$) (Fig. 2c). Of 70 biofilm-positive tumors available for multi-probe FISH, we identified broad patterns of

[1]Johns Hopkins University School of Medicine, Baltimore, MD, USA. [2]University of Maryland Baltimore County, Baltimore, MD, USA. [3]Universiti Malaya, Kuala Lumpur, Malaysia. [4]Resphera Biosciences, Baltimore, MD, USA. ✉e-mail: julia.drewes@gmail.com; csears@jhmi.edu

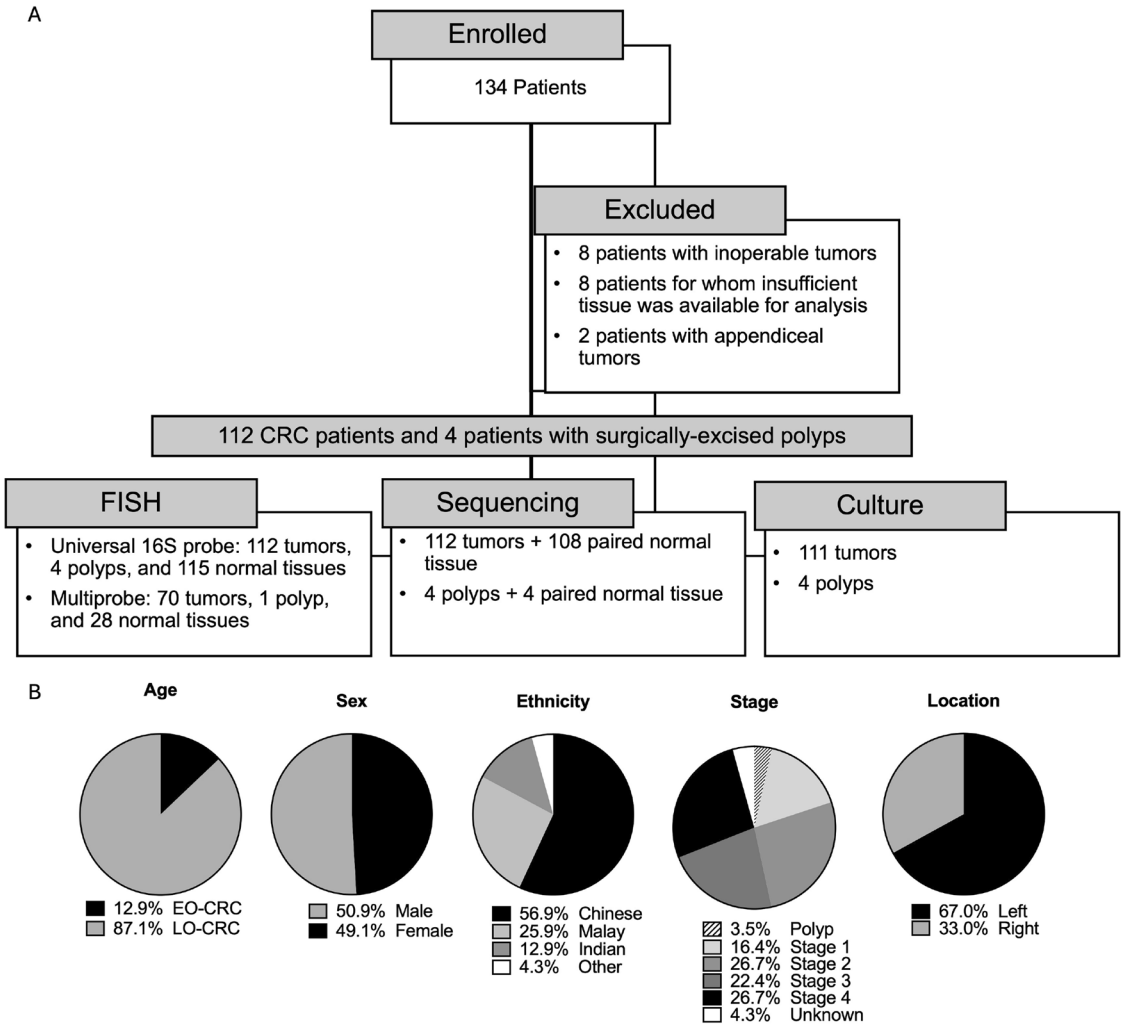

Fig. 1 | The CRC cohort. A Flow diagram of study participants and downstream analysis of samples. **B** Demographic characteristics of the cohort ($n = 116$), including age, sex, ethnicity, stage, and tumor location. EO-CRC early-onset colorectal cancer, defined as age under 50. LO-CRC late-onset colorectal cancer, defined as age 50 or above.

microbial biofilm community membership, consisting of either polymicrobial biofilms (94.3%; 66/70) with (Fig. 2d) or without (Fig. 2e) *Fusobacterium* (genus), or, rarely, biofilms predominated by proteobacteria (5.7%; 4/70) (Fig. 2f). Among the 66 polymicrobial biofilms, we visualized *Fusobacterium spp*. in 68.2% (45/66) of tumors, ranging from sparse populations of individual rods to dense blooms where *Fusobacterium* was the predominant organism. Although biofilms were not associated with tumor stage (Fig. 2g), the presence of *Fusobacterium* in tumor biofilms was associated with later tumor stage compared to polymicrobial biofilms where *Fusobacterium* was absent ($p = 0.0212$) (Fig. 2h). There was high concordance in biofilm status between tumors and paired normal tissue; in individuals who had biofilms observed on their tumors, 79.0% (60/76) also had biofilms on distant normal tissue (Fig. S1a). In those with no biofilms visualized on their tumors ($n = 34$), there were no biofilms observed in paired normal tissue. Similar to tumor biofilms, biofilm positive normal tissues screened by multiprobe FISH were predominantly polymicrobial (89.3%; 25/28) (Fig. S1b). In contrast to tumors, polymicrobial biofilms on normal tissue typically had sparse *Fusobacterium*, with no consistent observations of dense blooms or epithelial invasion of *Fusobacterium* in normal tissue (Fig. S1c, d).

Fusobacteria are enriched in the colon tumor microbiome

We performed 16S rRNA amplicon sequencing on tumor biopsies and paired normal tissue to identify bacterial taxa enriched or depleted in the tumor microenvironment. Two subsets of this cohort were previously analyzed by 16S rRNA amplicon sequencing and published[10]: MAL1 (consisting of 22 tumors and 1 adenoma with 20 paired normal samples) and MAL2 (consisting of 24 tumors with 22 paired normal samples, 3 of which were replicate pairs from MAL1). We performed 16S rRNA amplicon sequencing on an additional 79 tumors, 3 adenomas, and 71 paired normal samples; this subset, termed MAL3, included 10 replicate tumors previously sequenced in the MAL1 and/or 2 cohorts. Although all three cohorts were sequenced independently, each was sequenced using the V3–V4 hypervariable region as the amplification target, and data were analyzed using the same computational pipeline to assign species-level taxonomic identification (Resphera Insight, Baltimore, MD). Analysis of these ten replicates demonstrated no significant variation by sequencing cohort (Fig. S1a), as measured by Bray–Curtis dissimilarity, with significance estimated by Permutational multivariate analysis of variance (PERMANOVA). Linear discriminant Effect Size (LEfSe) analysis identified few taxa that were significantly enriched in one cohort relative to replicates in other cohorts (Fig. S1b), and notably did not include *Fusobacterium* or other taxa specifically investigated herein. Each set of replicate samples had consistent taxonomic assignments across independent sequencing runs for replicate samples at the phylum through species levels (Table S1a–g; Figs. S2–4). This included *F. nucleatum* and other *Fusobacterium* species, which are the particular focus of this study (Fig. S5). Given the consistent findings of concordant taxonomic assignments between replicates by multiple analytical approaches, we pooled MAL1, MAL2, and MAL3 data for all subsequent analyses (see Table S2 for all taxonomic assignments). Altogether, the combined MAL cohort

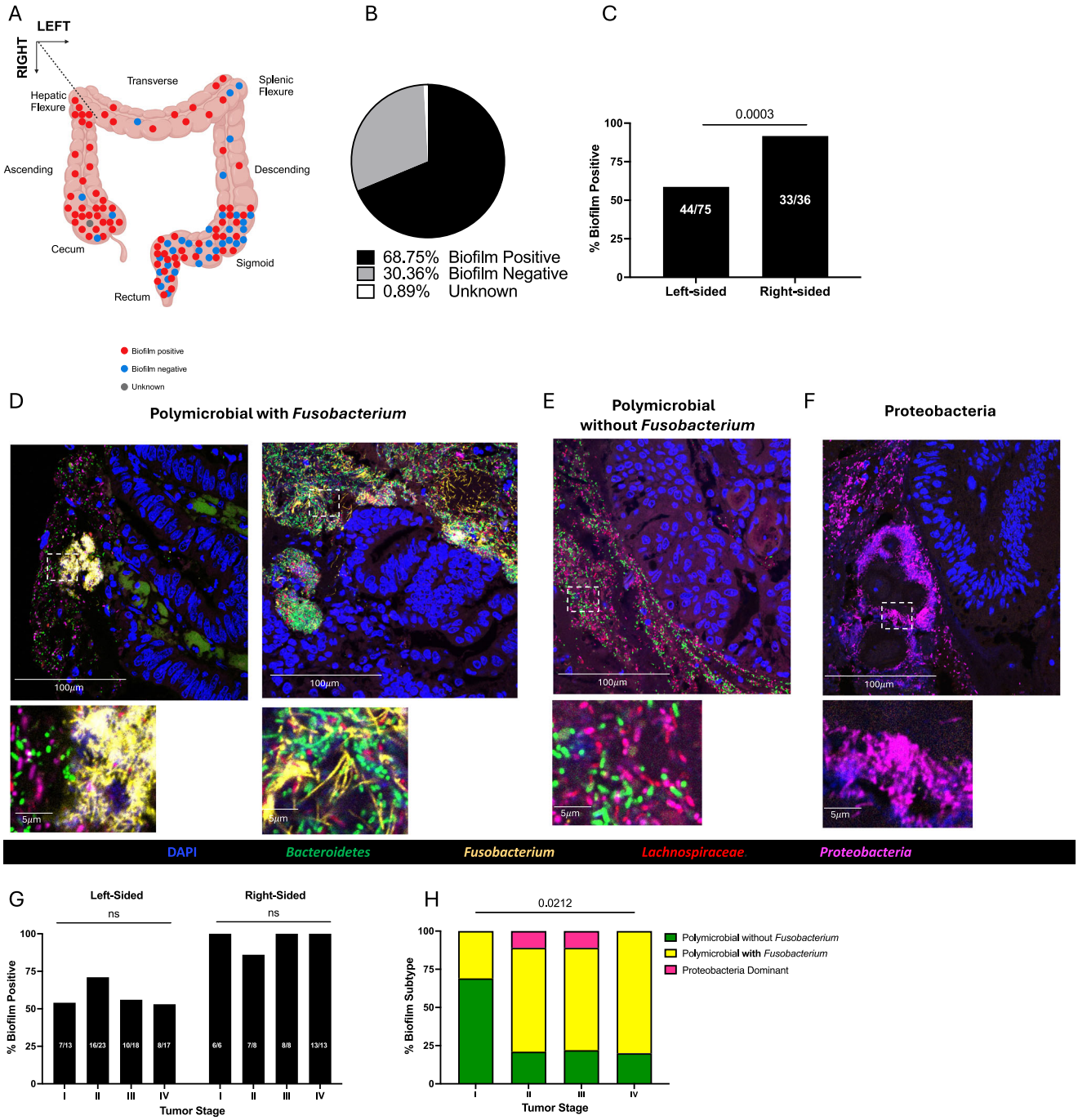

Fig. 2 | Biofilm analysis of the CRC cohort. A The location and biofilm status of CRC tumors overlaid on a diagram of a colon. One tumor marked as unknown had no fixed tissue suitable for screening. **B** Biofilm status of CRC tumors classified as either positive or negative by FISH. **C** Percent of left and right-sided tumors positive for biofilms. **D**–**F** Representative 40X images of multi-probe FISH of each biofilm subtype stained for DAPI (blue), Bacteroidetes (green), Fusobacterium (yellow), Lachnospiraceae (red), and Proteobacteria (pink). **G** Percent of tumors positive for biofilms stratified by tumor location and stage. **C**, **G** analyzed by Fisher's exact test, with $p < 0.05$ considered significant. **H** Stacked bar graph depicting percent of tumor biofilms of each subtype screened by multi-probe FISH ($n = 70$) and stratified by tumor stage. Analyzed by nonparametric Kruskal-Wallis one-way ANOVA, with $p < 0.05$ considered significant.

presented herein consists of 108 unique tumor and normal tissue pairs, 4 adenoma and normal tissue pairs, and an additional 4 tumors for which paired normal sequencing data is unavailable.

Consistent with prior studies[13,14], tumors had lower alpha diversity than normal tissue as measured by the number of observed species ($p = 0.0042$). Biofilm-negative tumors had lower alpha diversity than biofilm-negative normal tissue, as measured by number of observed species ($p = 0.019$) or Chao Index ($p = 0.0346$) (Fig. S6a). Although biofilm-positive tumors were more diverse than biofilm-negative tumors by Chao Index ($p = 0.0096$), there was no significant difference in alpha diversity between biofilm-positive tumor and normal samples, by Chao Index, Shannon Index, Simpson Index, or number of observed species. Neither sample type (tumor vs normal) nor biofilm status (positive vs negative) was associated with differences in beta diversity, as measured by Bray–Curtis dissimilarity, with significance estimated by PERMANOVA (Fig. S6b).

F. nucleatum was present in 86.6% of CRC tumors in this cohort (97/112), with relative abundance ranging from 0.01% to 40% of total microbial reads (Table S2). Of 112 tumors in the study, 108 had paired

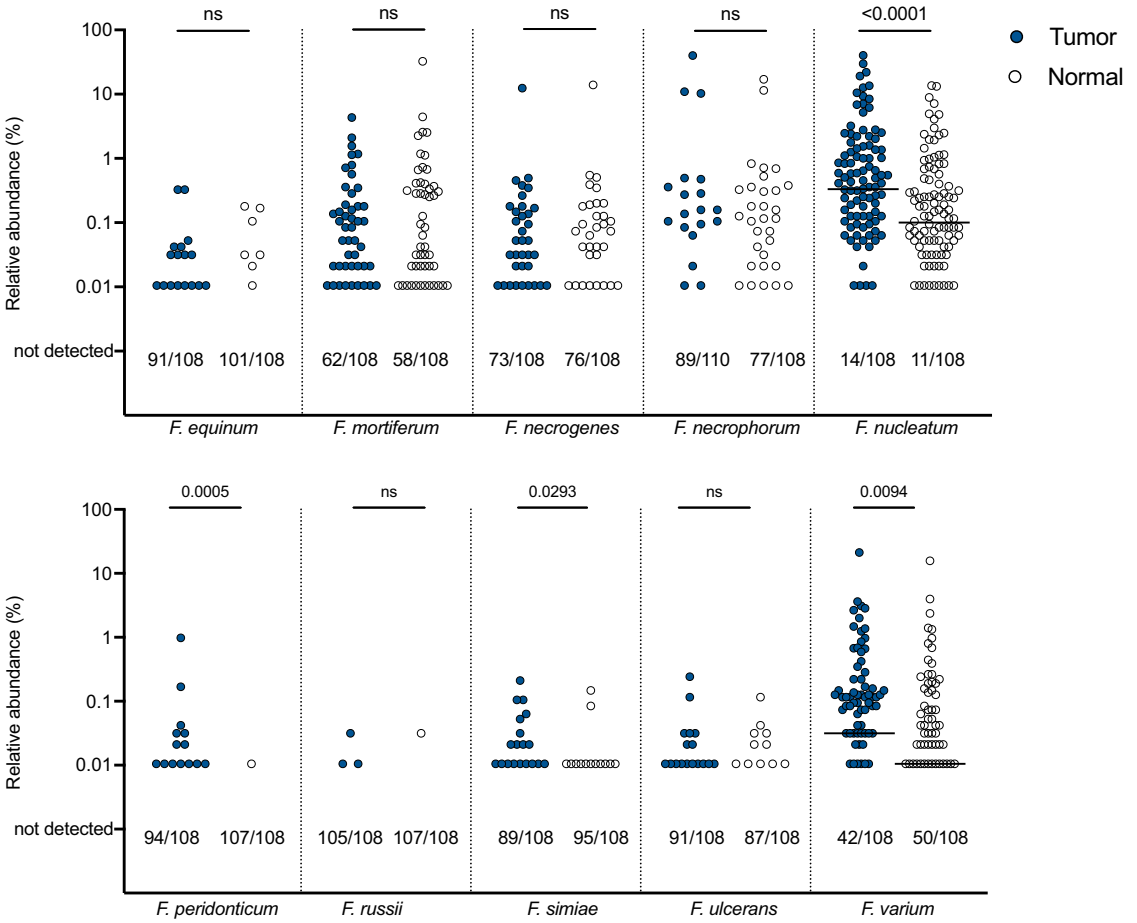

Fig. 3 | Abundance of *Fusobacterium* species in tumor/normal pairs. Relative abundance on a log scale by 16S rRNA amplicon sequencing of each depicted *Fusobacterium* species in pairs of tumor (blue circle) and normal (clear circle) tissues. Bars indicate median. Of 112 individuals with CRC in the study, 108 tumor and normal pairs were available for 16S rRNA sequencing analysis. Numbers below each graph depict the number of samples out of the total in which each species was not detected by sequencing (i.e. 0% abundance). Analyzed by two-tailed Wilcoxon matched-pairs signed rank test, with $p < 0.05$ considered significant. See Fig. S7 for abundance of *F. nucleatum* as a sum of confident and ambiguous species assignments, and for *F. varium* abundance stratified by ethnicity.

normal tissue available for 16S rRNA sequencing. Analysis of the 108 tumor/normal pairs revealed significant enrichment of *F. nucleatum* in the tumor microbiome compared to paired normal samples ($p < 0.0001$) (Fig. 3). Resphera Insight attempts to achieve species-level resolution; however, when a confident single species assignment is not feasible, the method minimizes false positives by providing ambiguous assignments that may include multiple closely related species. This significant enrichment in the tumor microenvironment was observed whether *F. nucleatum* abundance was calculated using confident (Fig. 3) or confident + ambiguous (Fig. S7a) assignments. Other *Fusobacterium* species enriched in tumors included *F. peridonticum* ($p = 0.0005$), *F. simiae* ($p = 0.0293$), and *F. varium* ($p = 0.0094$) (Fig. 3). The latter has previously been found to be enriched only in Southern Chinese populations[15]. In our cohort, *F. varium* was enriched in tumors vs paired normal tissue only for individuals of Malay descent (Fig. S7b). Also of note, *Bacteroides fragilis* was significantly enriched in the tumor microbiome relative to paired normal tissue ($p = 0.0002$), whereas other taxa previously linked to the CRC microbiome were not enriched in our cohort, including *Akkermansia muciniphila, Enterococcus faecalis*, *Escherichia coli*, and *Streptococcus gallolyticus* (Fig. S8). We also observed relative depletion of *Alistipes onderdonkii* ($p = 0.0261$), *Bifidobacterium longum* ($p = 0.0029$), *Parabacteroides distasonis* ($p = 0.0149$), and *Ruminococcus bromii* ($p < 0.0001$), with a trend toward depletion of *Alistipes senegalensis* ($p = 0.0620$) in tumor samples.

Fusobacterium nucleatum and other oral organisms are associated with colon tumor biofilms

We next investigated individual and tumor characteristics associated with *F. nucleatum* abundance in the tumor microbiome. Enrichment of *F. nucleatum* in this cohort was independent of sex (Fig. 4a). Enrichment was seen among individuals with late-onset (LO) CRC, but not those with early-onset (EO) CRC (Fig. 4b), although the latter constituted a small proportion of the study population ($n = 14$). The three major ethnic groups in Malaysia are Malay, Chinese, and Indian, with data suggesting disparities in CRC incidence and mortality between these groups[16,17]. *F. nucleatum* was enriched in tumors of individuals of Chinese and Malay descent, but not statistically significant for the participants of Indian heritage ($n = 14$) (Fig. 4c). *F. nucleatum* was enriched in both left and right-sided tumors (Fig. 4). Notably, *F. nucleatum* enrichment was observed only in later cancer stages ($p = 0.0041$, 0.1037, and 0.0046, respectively, for stages II–IV) (Fig. 4e) and in biofilm-positive tumors ($p < 0.0001$) (Fig. 4f; Fig. S7c). This enrichment in later cancer stages is, in fact, driven by higher abundance of *F. nucleatum* in tumors that have biofilms ($p = 0.0012$, 0.0084, and $p < 0.0001$, respectively, for biofilm-positive tumors at stages II-IV, whereas biofilm-negative tumors had no significant enrichment at any cancer stage (Fig. 4g). This finding supports our observations from multiprobe FISH (Fig. 2h), highlighting that although *F. nucleatum* is present in a majority of tumors in this cohort, it is specifically enriched in association with biofilms in advanced disease.

Because of the association of *F. nucleatum* with oral biofilms, we assessed for enrichment of other oral microbiome community members,

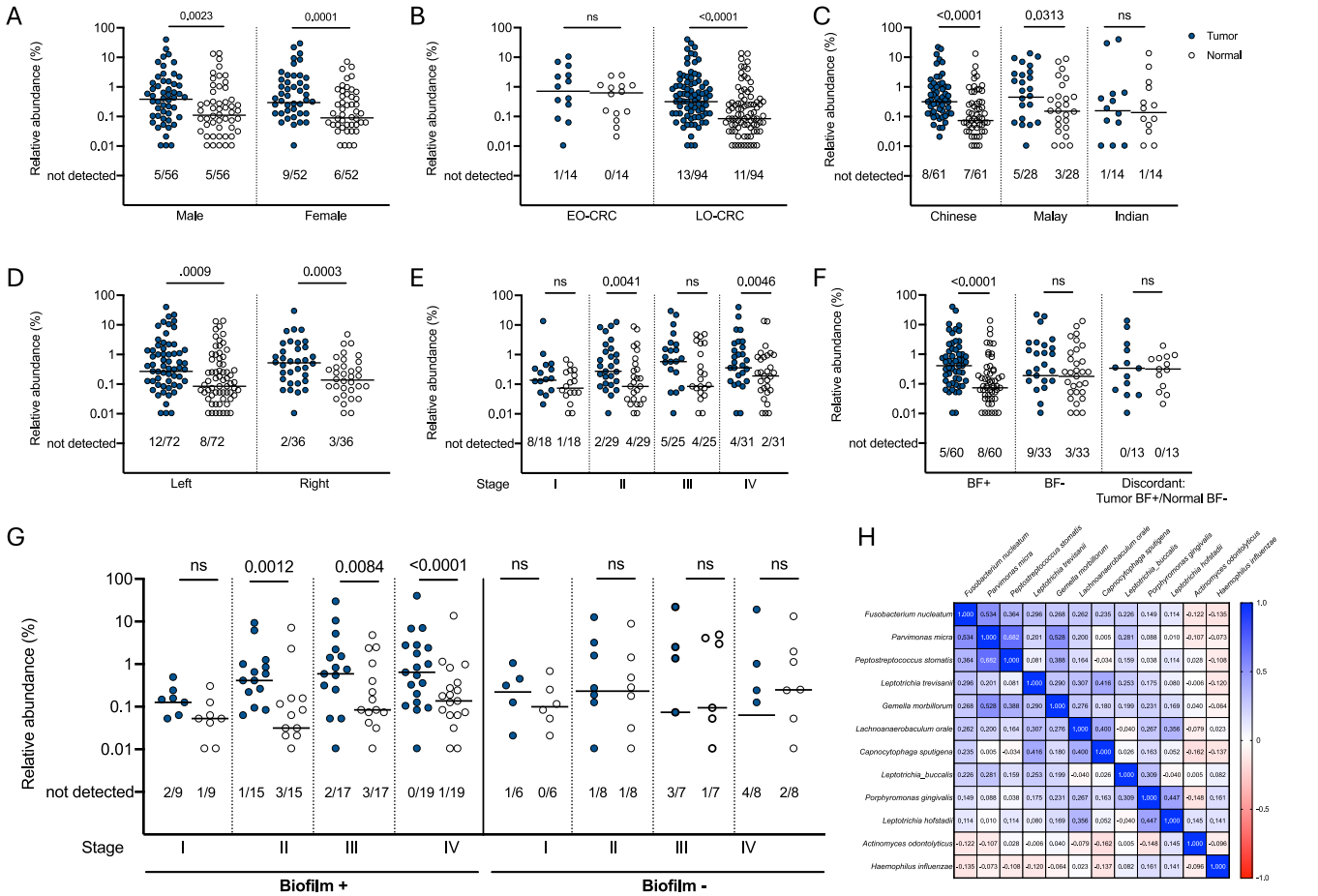

Fig. 4 | *Fusobacterium nucleatum* abundance by individual and tumor characteristics. Relative abundance by 16S rRNA amplicon sequencing of *Fusobacterium nucleatum* in pairs of tumor (blue circle) and normal (clear circle), stratified by **A** sex, **B** age, **C** ethnicity, **D** tumor location, **E** tumor stage, **F** biofilm (BF) status, and **G** both biofilm status and tumor stage. Bars indicated median. Of 112 individuals with CRC in the study, 108 tumor and normal pairs were available for 16S rRNA sequencing analysis. Numbers below each graph depict the number of samples out of the total in which *F. nucleatum* was not detected by sequencing. Analyzed by two-tailed Wilcoxon matched-pairs signed rank test, with $p < 0.05$ considered significant. EO-CRC early-onset colorectal cancer. LO-CRC late-onset colorectal cancer. See Fig. S7 for abundance of *F. nucleatum* as a sum of confident and ambiguous species assignments, stratified by biofilm status. **H** Matrix of *F. nucleatum* and other oral biofilm organism abundances in tumor samples, analyzed by nonparametric Spearman correlation coefficient. See Fig. S8 for relative abundance of each organism in the matrix in tumor/normal pairs.

including taxa known to associate with *Fusobacterium* in dental plaque biofilms[18,19]. 16S rRNA amplicon sequencing analysis revealed enrichment of *Actinomyces odontolyticus* ($p = 0.0127$), *Capnocytophaga sputigena* ($p = 0.0004$), *Gemella morbillorum* ($p < 0.0001$), *Haemophilus influenzae* ($p = 0.0164$), *Lachnoanaerobaculum orale* ($p < 0.0001$), *Leptotrichia hofstadii* ($p = 0.0019$), *Leptotrichia trevisanii* ($p < 0.0001$), *Parvimonas micra* ($p = 0.0255$), and *Porphyromonas gingivalis* ($p = 0.0057$) in the tumor microbiome (Fig. S9). Of these organisms, the relative abundance of *C. sputigena* ($p = 0.015$), *G. morbillorum* ($p = 0.005$), *L. orale* ($p = 0.006$), *L. trevisanii* ($p = 0.002$), and *P. micra* ($p < 0.0001$) were positively associated with abundance of *F. nucleatum* in the tumor microbiome (Fig. 4h). Relative abundances of two additional oral organisms, *Leptotrichia buccalis* ($p = 0.019$) and *Peptostreptococcus stomatis* ($p < 0.0001$) were associated with that of *F. nucleatum*, despite not being enriched in tumors relative to paired normal tissues (Fig. S9).

F. nucleatum subspecies animalis clade 2 is the predominant *Fusobacterium* enriched in colon tumors with biofilms

We next assessed the full cohort for the presence of the recently delineated clades within subspecies *animalis* using clade-specific amplicon sequence variants, as previously described[12]. We found that both subspecies *animalis* clade 1 (Fna C1) and clade 2 (Fna C2) were significantly enriched in the tumor microenvironment compared to normal tissues ($p = 0.0046$ and $p < 0.0001$, respectively) (Fig. 5a). However, Fna C2 was strikingly enriched in both tumors and normal tissues compared to Fna C1 ($p < 0.0001$). Similar to our findings at the species level (Fig. 4f), when compared to paired normal tissues, Fna C1 and C2 were only enriched in tumors that have biofilms ($p = 0.0027$ and $p < 0.0001$, respectively) (Fig. 5b). Abundance of Fna C2 was modestly enriched in biofilm-positive tumors (median 0.78%) relative to biofilm-negative tumors (median 0.19%) ($p = 0.0140$). When biofilm-positive tumors were stratified by stage, Fna C1 only reached statistically significant enrichment in stage 3 tumors (Fig. 5c), whereas Fna C2 was significantly enriched at all tumor stages (Fig. 5d). In our small cohort of four individuals with surgical adenomatous polyps, two had biofilms and two did not (Fig. S10A). *F. nucleatum*, including Fna C1 and Fna C2, were detected in a majority of these adenomas (Fig. S10B), although the small sample size precluded comparative analysis. Altogether, we find striking enrichment of Fna C2 in tumors that span the adenoma to advanced carcinoma sequence, and particularly in tumors displaying biofilms.

We assessed whether Fna C2 abundance by 16S rRNA amplicon sequencing correlated with the density of *Fusobacterium* visualized by FISH (Fig. 5e). Of note, the *Fusobacterium* FISH probe is genus-specific. In tumor biofilms with dense blooms of Fusobacteria ($n = 29$), a small number of tumors had higher abundance of non-*nucleatum* Fusobacteria by

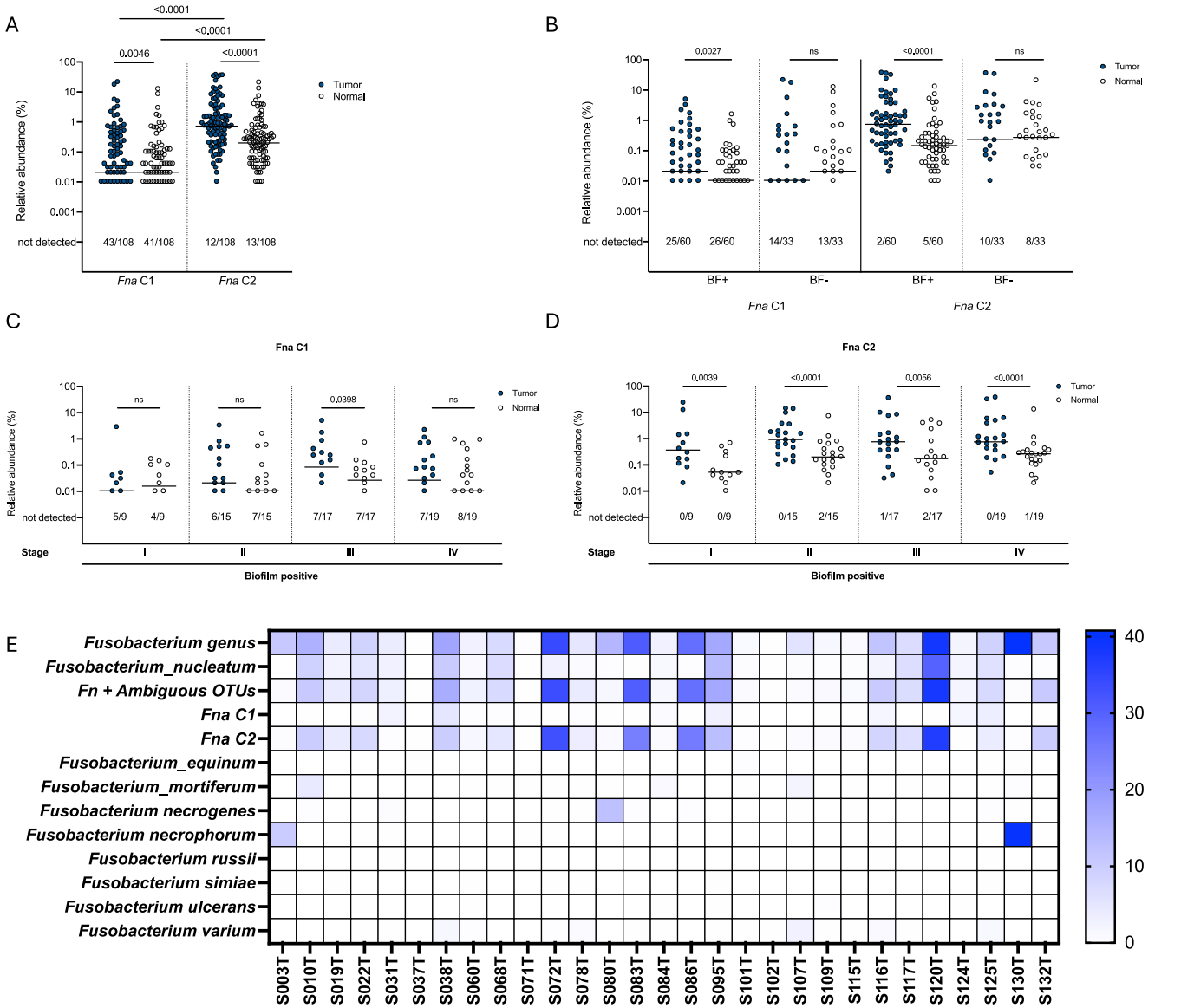

Fig. 5 | *Fusobacterium nucleatum subspecies animalis* abundance by clade and correlated with FISH. A Relative abundance by 16S rRNA amplicon sequencing of *Fusobacterium nucleatum subspecies animalis* clade 1 (Fna C1) and clade 2 (Fna C2) in pairs of tumor (blue circle) and normal (clear circle). Tumor/normal pairs analyzed by two-tailed Wilcoxon matched-pairs signed rank test, and tumor/tumor or normal/normal comparisons analyzed by two-tailed Mann Whitney test, with $p < 0.05$ considered significant. **B** Relative abundance of Fna C1 and C2 stratified by biofilm (BF) status and **C**, **D** by both biofilm status and tumor stage. Bars indicated median. Of 112 individuals with CRC in the study, 108 tumor and normal pairs were available for 16S rRNA sequencing analysis. Numbers below each graph depict the number of samples out of the total in which each clade was not detected by sequencing. **E** Heatmap depicting relative abundance by 16S rRNA amplicon sequencing for the *Fusobacterium* genus and each labeled *Fusobacterium* species or clade for $n = 29$ tumors where dense blooms of Fusobacteria were visualized by FISH.

sequencing analysis. However, in general, the abundance of Fna C2 was similar to the total estimated abundance of *F. nucleatum* (confident + ambiguous assignments) (Table S2), consistent with Fna C2 being dominant in the tumor microenvironment.

To identify whether other *F. nucleatum* subspecies were detectable in CRC tissues and to cross compare methodologies, we complemented our 16S rRNA amplicon sequencing analysis with a PCR-based method for distinguishing the *F. nucleatum* subspecies[20]. Of 34 tumors tested, *F. nucleatum* was detectable by PCR in 25, and among these, subspecies *animalis* was most frequently detected (Fig. S11a), consistent with previous reports[20,21]. When stratified by biofilm status, subspecies *animalis* was most frequently detected in biofilm positive tumors, whereas subspecies *vincentii* was more frequently detected in biofilm negative tumors (Fig. S11b). Notably, multiple subspecies were detected in 20% of tumor biopsies tested (5/25).

In parallel to sequencing and PCR analysis of *Fusobacterium* in the tumor microbiome of our cohort, we also used selective culture media to isolate unique *Fusobacterium* strains. Of the 13 *Fusobacterium* strains isolated (from 111 tumors), 6 were *F. nucleatum*, 50% of which were subspecies *animalis* (Table S3).

Biofilms and Fna C2 abundance correlate with predicted metabolic shifts

To begin to investigate functional consequences of Fna C2 enrichment and biofilm formation in the tumor microenvironment, we used the PICRUSt2 computational analysis pipeline to predict differences in metabolic pathways inferred from our 16S rRNA amplicon data (Fig. 6; Figs. S12, S13, S14; Table S2I). We also identified taxa that contributed to enrichment of these pathways (Fig. 6b; Table S4). Analysis was performed on 112 tumor samples

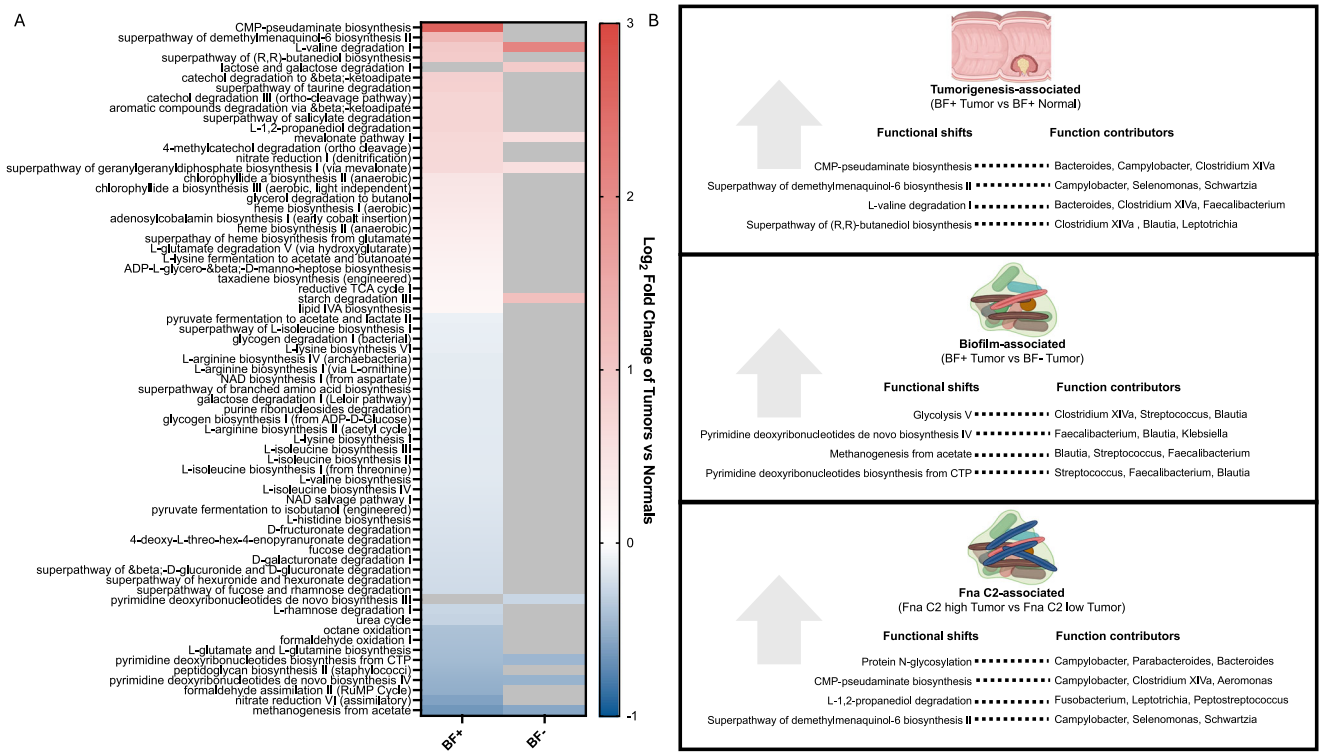

Fig. 6 | PICRUSt2 pathways associated with tumors, biofilms, and Fna C2 abundance. A Heatmap of all statistically significant PICRUSt2 pathways with differences in gene content between tumor ($n = 112$) and normal ($n = 112$) samples for either biofilm-positive tissues (left) or biofilm-negative tissues (right), as analyzed by Mann Whitney test, with $p < 0.05$ considered significant. Data depicted as Log_2 Fold Change between tumor and normal samples (>0 indicates higher in tumor). **B** For each indicated comparison, the listed pathways are the 4 PICRUSt2 pathways with the greatest Log_2 Fold Change between groups. For each pathway, the genera listed are the top 3 taxa contributing to enrichment of that pathway. See Figs. S12, S13, and S14A for the full display of significantly enriched or depleted pathways for each comparison. See Table S4 for all taxa significantly contributing to each enriched pathway.

and 112 normal samples (inclusive of 4 normal samples from patients with surgical polyps). We identified 67 PICRUSt2 pathways that predicted differential metabolic function between biofilm-positive tumor vs. biofilm-positive normal samples, and only 9 that predicted differences in metagenome content between biofilm-negative tumor and biofilm-negative normal samples (Fig. 6a). These data suggest that the presence of biofilms strongly impacts metabolic shifts in the microenvironment during tumor development. For example, in biofilm-positive tumors, there was a significant increase in gene content related to CMP-pseudaminate biosynthesis. Pseudaminic acids are sialic-acid-like sugars on the surface of many gram-negative bacteria, including *Bacteroides* and *Campylobacter*[22,23], which were the top taxa contributing to enrichment of this pathway in tumors compared to normal tissues (Fig. 6b). Data suggests that microbial-derived sialic-acid like sugars may facilitate immune evasion and biofilm formation[24].

To further investigate the impact of biofilms on predicted tumor microbiome metabolism, we compared biofilm-positive and biofilm-negative tumors, identifying 106 PICRUSt2 pathways that significantly differed (Fig. S12a). Among the PICRUSt2 pathways with the highest fold change between biofilm-positive and negative tumors (Fig. 6b) were pathways related to glucose metabolism and nucleotide biosynthesis, both of which are known to be dysregulated in cancer[25,26]. Interestingly, enrichment of these pathways in biofilm-positive tissues relative to biofilm-negative tissues was driven both by oral bacteria such as *Streptococcus*, as well as several commensal taxa typically thought of as beneficial to the host (e.g., *Clostridium* XIVa, *Blautia*, and *Faecalibacterium*).

Comparison of tumor samples with low Fna C2 abundance (<1%) and high abundance (≥1%) yielded 83 differentially predicted PICRUSt2 pathways. 78.3% (65/83) of the pathways associated with Fna C2 abundance in tumors were unique from those associated with biofilm status (Fig. S14a). Many of these pathways are involved in amino acid and cofactor biosynthesis, carbohydrate degradation, energy metabolism, or represent metabolic superpathways. Among the PICRUSt2 pathways with the highest fold-change in high Fna C2 abundance relative to low-abundance tumors (Fig. 6b) were protein N-glycosylation, which is known to be aberrant in tumors and is potentially a critical factor in tumor immune evasion[27]. Also associated with Fna C2 abundance was L-1,2-propanediol degradation, with *Fusobacterium* itself as the top contributing taxon, altogether consistent with the previous observation that Fna C2 has conservation of genetic content associated with 1,2-propanediol (1,2-PD) metabolism[12]. Several of the genera associated with these Fna C2-associated PICRUSt2 pathways (Fig. 6b) contain oral microbiome members (e.g., *Campylobacter*, *Leptotrichia*, and *Selenomonas*), further linking the oral microbiota with putative functional changes in the tumor microenvironment.

Lastly, we sought to investigate whether Fna C1 abundance was associated with a metabolic signature similar to Fna C2. When Fna C1 is present in the tumor microenvironment, it is typically at low abundance (median 0.02%) and absent in nearly 40% of tumors (Fig. 5a). We therefore sought to identify PICRUSt2 pathways significantly associated with presence or absence of Fna C1. We identified 34 pathways differentially represented in tumors with or without Fna C1, with many pathways associated with aromatic compound degradation and carbohydrate metabolism (Fig. S14b). Of these 34 pathways, only 5 pathways were significantly associated with both Fna C1 and Fna C2 abundance: L-glutamate degradation V, L-1,2-propanediol degradation, protein N-glycosylation, tRNA processing, and teichoic acid biosynthesis.

Discussion

Fusobacterium nucleatum, which typically resides in the oropharynx, has been consistently linked to the colon cancer microbiome in numerous studies across diverse cohorts[3,4,28–31]. *F. nucleatum* enrichment has been associated with specific pathological and clinical features such as high levels

of microsatellite instability (MSI-high), the CpG island methylator phenotype (CIMP), and poor prognosis[32–35]. Controversy remains as to the precise role(s) *F. nucleatum* plays in tumor initiation, progression, and immune evasion, although a large body of data argues against it merely existing as a passive colonizer of colon tumors. One key barrier to uncovering critical host-bacterial interactions in the tumor microenvironment is the diversity of *F. nucleatum* strains employed to date in mechanistic studies that may not accurately reflect the genomic content or adhesion factor repertoire of CRC-associated strains. Therefore, a key question has been which subspecies of this diverse organism are enriched in colon tumors. Inconsistent phenotypes in mouse models have also raised the question as to the critical community partners with which *F. nucleatum* may collaborate to colonize the lower gastrointestinal tract and promote carcinogenesis[36–38].

In this large cohort of individuals with CRC, we report that *F. nucleatum* is significantly enriched in the colon tumor microbiome when compared to distant normal tissues. A combination of sequencing, culture, and spatial evaluation has, in particular, allowed us to link *F. nucleatum* abundance in the tumor microenvironment with the presence of tumor biofilms as well as with enrichment of other oral microbiome members, particularly in advanced cancer stages. These data build on previous analyses linking *F. nucleatum* to CRC biofilms[10], and confirm the finding that biofilms on colon tumors frequently contain *Fusobacteria*. Although *F. nucleatum* has previously been linked to CRC biofilms, we report the surprising finding that *F. nucleatum* is only enriched in tumor tissue relative to normal tissue in individuals who have biofilms, a phenomenon largely driven by subspecies *animalis* clade 2 (Fna C2).

We find that subspecies *animalis* is most frequently detected by PCR and isolated by culture, consistent with other reports[20,21]. One limitation of our study was inability to perform strain-level comparisons of *Fusobacterium* isolates because of low culture yield, generating a library of only 13 Fusobacteria strains from the cohort of 112 individuals. Culture yield was likely impacted by long-term storage of tumor biopsies for 5–10 years prior to attempts at isolating *Fusobacterium* strains. Recent work has identified two clades (Fna C1 and C2) within subspecies *animalis* that were frequently detected in the mouths of healthy individuals, whereas only clade Fna C2 was enriched in colon tumors[12]. We confirm this finding in our cohort, with Fna C2 representing the most abundant *Fusobacterium* in the CRC microbiome, although our data show that other *Fusobacterium* species are also significantly enriched in the tumor microenvironment. Although we consider Fna C2 a high priority group for future investigation, fusobacterial communities in CRCs are complex, strongly suggesting that investigation of community features beyond Fna C2 is required to understand the mechanistic determinants of biofilms, *Fusobacterium*, and CRC.

By combining microbial sequencing and spatial analysis, we find that *F. nucleatum*, dominated by Fna C2, is frequently present in CRC biofilms and can bloom into dense populations, the latter rarely identified in biofilms on normal tissue[10]. This observation supports the theory that *F. nucleatum* is itself not a tumor initiator, but blooms in the tumor environment under specific conditions that allow it to flourish and subsequently promote tumor progression. A recent study in a screening colonoscopy cohort found that biofilms co-evolve with adenomas, increasing in density as polyps increase in size[39]. Likewise, our data suggest that biofilms likely precede or at least coincide with tumor initiation, as they are consistently concordant between tumor and distant normal tissue. However, while biofilms are observed in normal tissue in our cohort, *Fusobacterium* appears to colonize and bloom within polymicrobial biofilms only as tumors progress.

Notably, among the subset of tumors with *Fusobacterium* blooms, the overall abundance of *F. nucleatum* and other *Fusobacterium* species ranged in the tumor microenvironment from 0.06 to 40.8% of total microbial reads, suggesting that there are concentrations of Fusobacteria in specific niches in the tumor microenvironment that are underappreciated by nonspatial sequencing techniques. One factor to consider in this analysis is that tumor samples processed for 16S rRNA sequencing and FISH were from different biopsy regions. Although microbiome alpha-diversity is greater between individuals than between multiple biopsies from within a single tumor[40], there is a great deal of spatial heterogeneity of tumor microbiome composition[41]. Future investigations into the role of *F. nucleatum* in CRC pathogenesis will need to harness emerging spatial omics analyses to better understand the specific interactions of *F. nucleatum* and biofilm community members with host cells.

Although we saw only a modest impact of biofilms on compositional diversity, PICRUSt2 analysis predicted a marked impact on metabolic diversity, perhaps owing in part to the spatial organization of the microbial community. One key open question is whether *F. nucleatum* serves a similar role in colonic biofilms as in oral biofilms—that of a bridging species orchestrating the composition, structure, and potentially the function of polymicrobial biofilm communities. PICRUSt2 analysis suggests that both biofilms and Fna C2 abundance in the tumor microenvironment may be associated with altered, but differential, metabolic functions, a key hypothesis for future study. Whether the combined metabolic impact of biofilms and Fna C2 accelerates tumor development and/or metastatic disease is an open question. Consistent with previous studies, we also find enrichment of other oral organisms in colon tumors and contributions of oral organisms to enriched PICRUSt2 pathways, suggesting that *F. nucleatum* expansion occurs in the context of an oral microbiome community[10,42,43].

These analyses build on existing data demonstrating an association of *F. nucleatum* with the colon tumor microenvironment, and add critical resolution that highlights the abundance of *F. nucleatum*, and specifically subspecies *animalis* clade 2, in mucus-invasive biofilms observed particularly in late-stage disease. Our findings suggest that *F. nucleatum*:bacterial interactions within mucus-invasive biofilms are critical, yet understudied, features of CRC carcinogenesis.

Methods

Study participants

This study was approved by the Johns Hopkins Institutional Review Board and the University of Malaya Medical Centre (UMMC, Kuala Lumpur, Malaysia) Medical Ethics Committee. Written informed consents (provided in Malay and English) were obtained from all participants. All samples were obtained in accordance with the Health Insurance Portability and Accountability Act. Individuals who had received pre-operative radiation, chemotherapy or had a prior history of CRC were excluded. All individuals in the study underwent a standard mechanical bowel preparation. Standard pre-operative intravenous, but not oral, antibiotics were administered in all surgical cases. We collected data on the following variables: age, sex, ethnicity, BMI, smoking history, alcohol consumption, tumor location, tumor pathology, tumor grade, and tumor stage. Individuals under age 50 were defined as having early-onset CRC (EO-CRC), whereas individuals aged 50 and over were defined as having late-onset CRC (LO-CRC). Study data were collected and managed using REDCap electronic data capture tools hosted at Johns Hopkins University.

Sample collection

Excess colon tumor and paired normal tissues were collected for analysis from individuals undergoing surgery at UMMC. Surgical tissues were fixed in methacarn or flash frozen for later analysis.

Fluorescence in situ hybridization (FISH) analysis of biofilms

FISH was performed on methacarn-fixed, paraffin-embedded tissue sections as previously described[10]. Briefly, sequential sections were stained with the Eub338 universal bacterial probe for the presence of biofilms. Serial sections of samples confirmed to have biofilms were stained with specific bacterial probes for *Fusobacterium*, *Bacteroidetes*, *Betaproteobacteria*, *Gammaproteobacteria*, and *Lachnospiraceae* (see Oligonucleotide Table). Quantification of bacteria, lambda scanning, and linear unmixing were performed on a Zeiss 780 laser scanning confocal microscope as previously described[10]. A biofilm was defined as a dense aggregation of bacteria, with at least 20 bacteria within 1 μm of the epithelium (equivalent to 2×10^9 bacteria/ml)[9], covering an expanse of at least 200 μm of the CEC layer, in at least one of three screened areas.

16S rRNA gene Illumina library generation and sequencing

16S rRNA amplicon sequencing of cohorts MAL1, MAL2, and MAL3 were performed at different facilities with different methodologies for the three sets of samples. For each cohort, the V3-V4 hypervariable region of the 16S rRNA gene was amplified and sequenced.

For the MAL1 cohort, samples were lysed by suspending in 700 µL PBS and incubating with lysozyme (5 µL of 10 mg/mL stock), mutanolysin (15 µL of 1 mg/mL stock), and lysostaphin (5 µL of 1 mg/mL stock) in lysing matrix tubes at 37 °C. After 30 min, 10 µL Proteinase K and 50 µL 10% SDS was added, and samples were vortexed and then incubated at 55 °C for 45 min. Mechanical lysis was performed in a FastPrep-24 5G instrument at 6.0 m/s for 40 sec, then centrifuged at 10,000 *g* for 3 min. The ZR Fecal DNA MiniPrep kit (Zymo Research) was then used for DNA extraction. The V3-V4 region was amplified using primers 319 F (5′-ACTCCTACGGGAGGCAGCAG-3′) and 806R (5′-GGACTACHVGGGTWTCTAAT-3′) containing a linker sequence required for Illumina MiSeq 300 bp paired-end sequencing and a 12-bp heterogeneity-spacer index sequence. Sequencing was performed on the Illumina MiSeq (Illumina, San Diego, CA) according to the manufacturer's protocol.

For the MAL2 cohort, samples were lysed and DNA extracted using the MasterPure DNA Purification Kit (Epicentre/Illumina) according to the manufacturer's instruction. Primers used for amplification were S-D-Bact-0341-b-S-17 forward (5′-NNNNCCTACGGGNGGCWGCAG-3′) and S-D-Bact-0785-a-A-21 reverse (5'-GACTACHVGGGTATCTAATCC-3'), designed to include Illumina-compatible adaptors. Sequencing was performed on the MiSeq System (Illumina, San Diego, CA, USA) at the Monash University Malaysia Genomics Facility using the MiSeq 500-cycle reagent kit V2 on standard flow cell.

For the MAL3 cohort, DNA was extracted using the ZR Quick-DNA Fecal/Soil Microbe 96 Kit or Miniprep Kit (Zymo Research) with the following modification: 1.0 mm Zirconia-Silicate beads (BioSpec) were added to the 0.1/0.5 mm beads provided in the kit for improved tissue homogenization. Samples were homogenized and lysed using the Mini-beadbeater 96 (BioSpec) at maximum speed for 60 s x 3. Lysate was then processed according to manufacturer instructions. Sequencing was performed at Novogene with DNA amplified with primers 341 F (5'-CCTAYGGGRBGCASCAG-3') and 806R (5'-GGACTACNNGGGTATCTAAT-3'), and sequencing performed on the Illumina NovaSeq 6000 platform.

Analysis of all 16S rRNA amplicon sequence data sets

Paired-end 16S rRNA amplicon sequences were trimmed for quality and length using Trimmomatic, merged using FLASH (Fast Length Adjustment of Short reads)[44], and quality screened using QIIME[45]. Spurious hits to the PhiX control genome were identified using BLASTN (Basic Local Alignment Search Tool) and removed. Passing sequences were trimmed of 16S rRNA primers, evaluated for PCR chimeras and filtered for host-associated contaminant using Bowtie2[46]. Chloroplast contaminants were identified and filtered using the RDP (Ribosomal Database Project) classifier[47] with a confidence threshold of 50%. High-quality 16S rRNA amplicon sequences were assigned to a high-resolution taxonomic lineage using Resphera Insight, which uses a hybrid global-local alignment strategy to a manually curated 16S rRNA database with 11,000 unique species[10,48,49]. This approach attempts to achieve species-level resolution when possible; however, when a confident single species assignment is not feasible, the method minimizes false positives by providing "ambiguous assignments" i.e., a list of candidate species reflecting the ambiguity. Sequences that cannot be confidently assigned to any known species are clustered de novo (at 97% identity) and resulting clusters are given an OTU # identifier and annotated with a closest relative (or set of relatives). Taxonomic assignments are presented as unambiguous assignments by Resphera Insight unless otherwise indicated. Normalization of observations were performed by rarefaction followed by alpha and beta diversity characterization.

Taxonomic abundances were converted from counts to percentages. LEfSe was run to identify strong effect differentiating taxa among the three replicate cohorts of interest, with visualization of LDA scores by barplot and taxonomic cladogram. The method was run requiring a nonparametric *P*-value threshold of 0.01, and LDA score threshold of 2.5. Taxonomic profiles for samples reflecting replicates were compared pairwise for quantitative concordance of detected genera and species and using Pearson and Spearman and correlation. Statistical associations with beta-diversity utilized PERMANOVA. Functional inference of microbial gene content was evaluated using PICRUSt2[50]. Predicted PICRUSt2 metagenomic contributions per taxon to each pathway were calculated per sample and converted to relative abundance values, followed by comparisons between groups of interest using the Mann-Whitney U Test with P-value threshold of 0.05. Classification of *Fusobacterium nucleatum* C1 or C2 subgroups was performed for individual 16S rRNA amplicon sequences, using BLASTN (ungapped alignment; word size 124; minimum 99% identity) against a reference database of C1 and C2 amplicon sequence variant representatives[12]. Candidate sequences with alignments exclusively to C1 or C2 were labeled accordingly.

Subspecies PCRs

DNA was extracted from 3 mm tumor punch biopsies using the ZR Quick-DNA Fecal/Soil Miniprep Kit (Zymo Research) with the following modification: 1.0 mm Zirconia-Silicate beads (BioSpec) were added to the 0.1/0.5 mm beads provided in the kit for improved tissue homogenization. Samples were homogenized and lysed using the Mini-beadbeater 96 (BioSpec) at maximum speed for 60 s × 3. Lysate was then processed according to manufacturer instructions. DNA concentration was measured on a nanodrop, and ranged from 30.4 to 260.5 ng/µl. DNA was screened for *F. nucleatum* using primers specific for *nusG* (forward, 5'-CAACCATTACTTTAACTCTACCATGTTCA-3'; reverse, 5'- GTTGACTTTACAGAAGGAGATTATGTAAAAATC-3')[3] at a primer concentration of 0.9 µM using a PCR program of 98 °C × 60 s; 35 cycles of 98 °C × 30 s, 58 °C × 60 s, 72 °C × 20 s; and 72 °C × 7 min. DNA was then screened using primers specific for each *F. nucleatum* subspecies (See Oligonucleotide Table)[20] at a primer concentration of 0.5 µM, using a PCR program of 98 °C × 60 s; 30 cycles of 98 °C × 30 s, 56 °C × 15 s, 72 °C × 20 s; and 72 °C × 7 min.

Selective culture

Three mm punches were taken from flash-frozen tumor biopsy tissues and homogenized with a sterile pestle in 50 µl brain heart infusion (BHI) media supplemented with hemin (10 µg/mL) and vitamin K (5 µg/mL). Total volume was brought up to 500 µl of BHI and vortexed. 10 µl of the homogenate was spread onto *Fusobacterium* selective agar (FSA; Anaerobe Systems) and 10 µl were subcultured in 3 mL of BHI. Each was incubated for 48–72 h under anaerobic conditions (75% N_2, 5% H_2, 20% CO_2) in an anaerobic chamber. From the homogenate culture in BHI media, 10 µl was then streaked onto FSA plates and incubated as above. From these two FSA plates (1 directly plated, 1 subcultured first in BHI), unique colonies were picked and spread onto new FSA plates for isolation (minimum 6 colonies/tumor) and incubated. After 48–72 h, colonies were scooped up and swirled into 50 µl sterile water, which was boiled to release bacterial DNA. This crude DNA prep was used as a template for screening by quantitative real-time PCR (qRT-PCR) with *Fusobacterium* genus-specific 16S primers (forward, GGATTTATTGGGCGTAAAG; reverse, GGCATTCCT ACAAATATCTACGAA)[51] and probe (5'VIC- TGC AGG GCT CAA CTC TGT ATT GCG - 3'BHQ1), at a concentration of 0.2 µM, and with the following program: 50 °C × 2 min; 95 °C × 10 min; 40 cycles of: 95 °C × 15 s, 58 °C × 60 s. Colonies were considered positive if the CT value was <30. Each positive colony was subcultured into BHI media × 48 h, and genomic DNA was extracted from pelleted bacteria using the ZR Quick-DNA Fecal/Soil Miniprep Kit (Zymo Research). Species and subspecies were confirmed by Sanger sequencing of the 16S rRNA gene (Azenta Life Sciences).

Quantification and statistical analysis

To compare categorical data between defined groups (e.g., biofilm status by tumor location), data were analyzed using a two-tailed Mann Whitney

nonparametric t-test. Comparison of biofilm subtype across tumor stages was analyzed by a nonparametric Kruskal-Wallis one-way ANOVA. Prior to downstream statistical comparisons, 16S rRNA amplicon profiles within each dataset were subsampled to an even level of coverage (9500 reads). For comparisons of relative taxonomic abundances between matched tumor-normal pairs, data were analyzed by a two-tailed non-parametric Wilcoxon signed-rank sum test. To identify associations of oral taxa abundances with *F. nucleatum* abundance, we performed a non-parametric Spearman correlation. Differences in PICRUSt2 pathway content was analyzed by Mann Whitney Test or by Kruskal-Wallis test with Dunn's multiple comparisons, as indicated in the figure legend. For all analyses, differences with a P value of <0.05 were considered significant.

Oligonucleotide Table

Oligonucleotide	Source
Eub338 universal bacterial FISH probe (5′ - Cy3 -GCTGCCTCCCGTAGGAGT - 3′)	MBD0033 Sigma-Aldrich, USA
Fusobacterium FISH probe (Fus714; 5′ - Cy3 – GGCTTCCCCATCGGCATT - 3′)	Valm et al.[52]
Bacteroidetes FISH probe (CFB286; 5′-Alexafluor488-TCCTCTCAGAACCCCTAC - 3′)	Weller et al.[53]
Betaproteobacteria FISH probe (Bet42a; 5′ - Cy5 - GCCTTCCCACTTCGTTT - 3′)	Manz et al.[54]
Gammaproteobacteria FISH probe (Gam42a; 5′ - Cy5 - GCCTTCCCACATCGTTT - 3′)	Manz et al.[54]
Lachnospiraceae FISH probe (Lac435; 5′ - TXR - TCTTCCCTGCTGATAGA - 3′)	Kong et al.[55]
Animalis primers: SR-a2-F, ACTCAAATTATTATGAATGTGATGAAAGA SR-a2-R, GCTACTGAAGGATGAAATGCTGG	Bi et al.[20]
Nucleatum primers: SR-n3-F, CAAGCAACTGAAAATGCTTTAAAAG SR-n3-R, TCCAGGTAAGGAAATTACACCTACTG	Bi et al.[20]
Polymorphum primers: SR-p1-F, TTAGGAAATCTTTTAGAAGCAAAAACA SR-p1-R, TCTACTGTAATAGTTACAAACTCTGCWCC SR-p3-F, CYTATGGYTTTGATTTTGACTTATTTG SR-p3-R, CCAAAGTAATTAAAGCCTCTTGAGC	Bi et al.[20]
Vincentii primers: SR-v1-F, GAGGCTATTGCAAATTAAACTGTTAAA SR-v1-R, CTTTACCACTATTATAAACTAAATAAAT	Bi et al.[20]

Data availability

Raw sequences from MAL1 and MAL2 were previously deposited in the NCBI SRA repository under BioProject accession nos. PRJNA325649 and PRJNA325650, respectively. Raw sequences from MAL3 were deposited in the NCBI SRA repository under BioProject accession no. PRJNA1195962.

Code availability

Primary sequencing data and open source R code used for meta-analyses are available from the authors upon request.

Received: 16 December 2024; Accepted: 1 May 2025;
Published online: 20 May 2025

References

1. Bray, F. et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **68**, 394–424 (2018).
2. Knippel, R. J. & Sears, C. L. The microbiome colorectal cancer puzzle: initiator, propagator, and avenue for treatment and research. *J. Natl. Compr. Cancer Netw.* **19**, 986–992 (2021).
3. Castellarin, M. et al. Fusobacterium nucleatum infection is prevalent in human colorectal carcinoma. *Genome Res.* **22**, 299–306 (2012).
4. Kostic, A. D. et al. Genomic analysis identifies association of Fusobacterium with colorectal carcinoma. *Genome Res.* **22**, 292–298 (2012).
5. Mima, K. et al. Fusobacterium nucleatum in colorectal carcinoma tissue according to tumor location. *Clin. Transl. Gastroenterol.* **7**, e200 (2016).
6. Yachida, S. et al. Metagenomic and metabolomic analyses reveal distinct stage-specific phenotypes of the gut microbiota in colorectal cancer. *Nat. Med.* **25**, 968–976 (2019).
7. Wang, N. & Fang, J. Y. Fusobacterium nucleatum, a key pathogenic factor and microbial biomarker for colorectal cancer. *Trends Microbiol.* **31**, 159–172 (2023).
8. Valadbeigi, H. et al. Mixed oral biofilms are controlled by the interspecies interactions of Fusobacterium nucleatum. *Oral Dis.* https://doi.org/10.1111/odi.14822 (2023).
9. Dejea, C. M. et al. Microbiota organization is a distinct feature of proximal colorectal cancers. *Proc. Natl. Acad. Sci. USA* **111**, 18321–18326 (2014).
10. Drewes, J. L. et al. High-resolution bacterial 16S rRNA gene profile meta-analysis and biofilm status reveal common colorectal cancer consortia. *NPJ Biofilms Microbiomes* **3**, 34 (2017).
11. Kook, J. K. et al. Genome-based reclassification of fusobacterium nucleatum subspecies at the species level. *Curr. Microbiol* **74**, 1137–1147 (2017).
12. Zepeda-Rivera, M. et al. A distinct Fusobacterium nucleatum clade dominates the colorectal cancer niche. *Nature* **628**, 424–432 (2024).
13. Barot, S. V. et al. Distinct intratumoral microbiome of young-onset and average-onset colorectal cancer. *EBioMedicine* **100**, 104980 (2024).
14. Rodriguez, R. M., Hernandez, B. Y., Menor, M., Deng, Y. & Khadka, V. S. The landscape of bacterial presence in tumor and adjacent normal tissue across 9 major cancer types using TCGA exome sequencing. *Comput. Struct. Biotechnol. J.* **18**, 631–641 (2020).
15. Yeoh, Y. K. et al. Southern Chinese populations harbour non-nucleatum Fusobacteria possessing homologues of the colorectal cancer-associated FadA virulence factor. *Gut* **69**, 1998–2007 (2020).
16. Abu Hassan, M. R. et al. Incidence and mortality rates of colorectal cancer in Malaysia. *Epidemiol. Health* **38**, e2016007 (2016).
17. Ramli, S. R., Azhar, Z. I., Raman, S., Yusof, S. N. & Mohamad, M. Survival rate of colorectal cancer and its relation to the individual and geographical variations in Malaysia, 2013–2018. *Cancer Epidemiol.* **95**, 102756 (2025).
18. Valm, A. M. The structure of dental plaque microbial communities in the transition from health to dental caries and periodontal disease. *J. Mol. Biol.* **431**, 2957–2969 (2019).
19. Schoilew, K. et al. Bacterial biofilm composition in healthy subjects with and without caries experience. *J. Oral. Microbiol.* **11**, 1633194 (2019).
20. Bi, D. et al. A newly developed PCR-based method revealed distinct Fusobacterium nucleatum subspecies infection patterns in colorectal cancer. *Microb. Biotechnol.* https://doi.org/10.1111/1751-7915.13900 (2021).
21. Komiya, Y. et al. Patients with colorectal cancer have identical strains of. *Gut* **68**, 1335–1337 (2019).

22. Thibault, P. et al. Identification of the carbohydrate moieties and glycosylation motifs in Campylobacter jejuni flagellin. *J. Biol. Chem.* **276**, 34862–34870 (2001).
23. Wang, L., Lu, Z., Allen, K. N., Mariano, P. S. & Dunaway-Mariano, D. Human symbiont bacteroides thetaiotaomicron synthesizes 2-keto-3-deoxy-D-glycero-D- galacto-nononic acid (KDN). *Chem. Biol.* **15**, 893–897 (2008).
24. Haines-Menges, B. L., Whitaker, W. B., Lubin, J. B. & Boyd, E. F. Host sialic acids: a delicacy for the pathogen with discerning taste. *Microbiol. Spectr.* **3** https://doi.org/10.1128/microbiolspec.MBP-0005-2014 (2015).
25. Qin, R. et al. Role of glucose metabolic reprogramming in colorectal cancer progression and drug resistance. *Transl. Oncol.* **50**, 102156 (2024).
26. Mullen, N. J. & Singh, P. K. Nucleotide metabolism: a pan-cancer metabolic dependency. *Nat. Rev. Cancer* **23**, 275–294 (2023).
27. Hsu, J. M. et al. STT3-dependent PD-L1 accumulation on cancer stem cells promotes immune evasion. *Nat. Commun.* **9**, 1908 (2018).
28. Ito, M. et al. Association of Fusobacterium nucleatum with clinical and molecular features in colorectal serrated pathway. *Int. J. Cancer* **137**, 1258–1268 (2015).
29. Viljoen, K. S., Dakshinamurthy, A., Goldberg, P. & Blackburn, J. M. Quantitative profiling of colorectal cancer-associated bacteria reveals associations between fusobacterium spp., enterotoxigenic Bacteroides fragilis (ETBF) and clinicopathological features of colorectal cancer. *PLoS One* **10**, e0119462 (2015).
30. Zhou, Y. et al. Association of oncogenic bacteria with colorectal cancer in South China. *Oncotarget* **7**, 80794–80802 (2016).
31. Li, Y. Y. et al. Association of Fusobacterium nucleatum infection with colorectal cancer in Chinese patients. *World J. Gastroenterol.* **22**, 3227–3233 (2016).
32. Tahara, T. et al. Fusobacterium in colonic flora and molecular features of colorectal carcinoma. *Cancer Res.* **74**, 1311–1318 (2014).
33. Mima, K. et al. Fusobacterium nucleatum in colorectal carcinoma tissue and patient prognosis. *Gut* **65**, 1973–1980 (2016).
34. Lee, J. B. et al. Association between Fusobacterium nucleatum and patient prognosis in metastatic colon cancer. *Sci. Rep.* **11**, 20263 (2021).
35. Borozan, I. et al. Molecular and pathology features of colorectal tumors and patient outcomes are associated with cancer. *Epidemiol. Biomark. Prev.* **31**, 210–220 (2022).
36. Kostic, A. D. et al. Fusobacterium nucleatum potentiates intestinal tumorigenesis and modulates the tumor-immune microenvironment. *Cell Host Microbe* **14**, 207–215 (2013).
37. Queen, J. et al. Comparative analysis of colon cancer-derived fusobacterium nucleatum subspecies: inflammation and colon tumorigenesis in murine models. *mBio*, e0299121 https://doi.org/10.1128/mbio.02991-21 (2022).
38. Tomkovich, S. et al. Locoregional effects of microbiota in a preclinical model of colon carcinogenesis. *Cancer Res.* **77**, 2620–2632 (2017).
39. Drewes, J. L. et al. Epidemiology of bacterial biofilms on polyps and normal tissues in a screening colonoscopy cohort. *Gut Microbes* **17**, 2452233 (2025).
40. Murphy, C. L. et al. Mapping the colorectal tumor microbiota. *Gut Microbes* **13**, 1–10 (2021).
41. Liu, W. et al. Microbial community heterogeneity within colorectal neoplasia and its correlation with colorectal carcinogenesis. *Gastroenterology* **160**, 2395–2408 (2021).
42. Shah, M. S. et al. Leveraging sequence-based faecal microbial community survey data to identify a composite biomarker for colorectal cancer. *Gut* **67**, 882–891 (2018).
43. Russo, E. et al. From adenoma to CRC stages: the oral-gut microbiome axis as a source of potential microbial and metabolic biomarkers of malignancy. *Neoplasia* **40**, 100901 (2023).
44. Magoč, T. & Salzberg, S. L. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27**, 2957–2963 (2011).
45. Caporaso, J. G. et al. QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7**, 335–336 (2010).
46. Langmead, B., Wilks, C., Antonescu, V. & Charles, R. Scaling read aligners to hundreds of threads on general-purpose processors. *Bioinformatics* **35**, 421–432 (2019).
47. Wang, Q., Garrity, G. M., Tiedje, J. M. & Cole, J. R. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol* **73**, 5261–5267 (2007).
48. Grim, C. J. et al. High-resolution microbiome profiling for detection and tracking of *Salmonella enterica*. *Front. Microbiol.* **8**, 1587 (2017).
49. Daquigan, N., Seekatz, A. M., Greathouse, K. L., Young, V. B. & White, J. R. High-resolution profiling of the gut microbiome reveals the extent of. *NPJ Biofilms Microbiomes* **3**, 35 (2017).
50. Langille, M. G. et al. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat. Biotechnol.* **31**, 814–821 (2013).
51. Boutaga, K., van, Winkelhoff, A. J., Vandenbroucke-Grauls, C. M. & Savelkoul, P. H. Periodontal pathogens: a quantitative comparison of anaerobic culture and real-time PCR. *FEMS Immunol. Med. Microbiol.* **45**, 191–199 (2005).
52. Valm, A. M. et al. Systems-level analysis of microbial community organization through combinatorial labeling and spectral imaging. *Proc. Natl. Acad. Sci. USA* **108**, 4152–4157 (2011).
53. Weller, R., Glockner, F. O. & Amann, R. 16S rRNA-targeted oligonucleotide probes for the in situ detection of members of the phylum Cytophaga-Flavobacterium-Bacteroides. *Syst. Appl. Microbiol.* **23**, 107–114 (2000).
54. Manz, W., Amann, R., Ludwig, W., Wagner, M. & Schleifer, K. H. Phylogenetic oligodeoxynucleotide probes for the major subclasses of proteobacteria - problems and solutions. *Syst. Appl. Microbiol.* **15**, 593–600 (1992).
55. Kong, Y. H., He, M. L., McAlister, T., Seviour, R. & Forster, R. Quantitative fluorescence in situ hybridization of microbial communities in the rumens of cattle fed different diets. *Appl. Environ. Microbiol.* **76**, 6933–6938 (2010).

Acknowledgements

We honor the memory of the late Dr. Khean Lee Goh, who contributed to this work. We thank Kumar Thulasi, Han Ming Gan, Hoong Yin Chong, and Sandip Kumar for their technical assistance.We thank the Sidney Kimmel Comprehensive Cancer Center and the Johns Hopkins University Oncology Tissue Services (supported by NCI grant P30 CA006973), and the Johns Hopkins University Institute for Basic Biomedical Science Microscopy Facility for use of their Zeiss LSM 780 Confocal microscope (with Fluorescence Correlation Spectroscopy) (supported by NIH Grant S10 OD016374).This work was supported by the National Institute of Allergy and Infectious Diseases training grant T32-A1007291, the Biocodex Microbiota Foundation, the Burroughs Wellcome Fund Career Award for Medical Scientists [1022128], and the Black in Cancer award sponsored by the Emerald Foundation, Inc. (J.Q.); the U-RISE Program at the University of Maryland, Baltimore County (UMBC), which is supported by the National Institute of General Medical Sciences of the National Institutes of Health under Award Number T34-GM136497 (Z.C.); the National Institute of General Medical Sciences training grant 3-T32-GM136577 (H.M.); the University of Malaya Research Grant RP016A-13HTM (J.V.); National Cancer Institute grant R00 CA230192 (J.L.D.); the Bloomberg Philanthropies, and the Cancer Grand Challenges OPTIMISTICC team grant [A27140] funded by Cancer Research UK (C.L.S.). Fig. 2a, Fig. S6b, and Fig. S10a were created with https://biorender.com.

Author contributions

Conceptualization: J.Q., J.L.D., C.L.S.; Methodology: J.Q., J.W., J.R.W, J.L.D, C.L.S.; Software: J.R.W.; Formal Analysis: J.Q., J.R.W., J.L.D.; Validation: J.Q., J.R.W.; Investigation: J.Q., Z.C., H.M., A.N., T.S., J.F., M.M., T.I., J.V., A.R., M.F.L., J.W., J.R.W., J.L.D.; Resources: J.Q., Z.C., J.V., J.L.D.,

C.L.S.; Data Curation: J.W., J.R.W., J.L.D., C.L.S.; Writing – Original Draft: J.Q.; Writing – Review & Editing: all authors; Visualization: J.Q., Z.C., J.R.W., J.L.D.; Supervision: J.L.D., C.L.S.; Project Administration: J.W., C.L.S.; Funding Acquisition: J.Q., J.L.D., C.L.S.

Competing interests

C.L.S. has received research funding to Johns Hopkins University from Bristol-Myers Squibb and Janssen, and royalties from Up to Date outside the submitted work. J.R.W. reports equity ownership of Resphera Biosciences.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41522-025-00717-7.

Correspondence and requests for materials should be addressed to Julia L. Drewes or Cynthia L. Sears.