Indole Profiles Define Disease-Specific Virulence in *Fusobacterium nucleatum* Clinical Isolates.

Colin Scano[1], Ankan Choudhury[2], Macarena G. Rojo[3], Jessalyn Hawkins[1], Lauren Crowhurst[1], Greg Zaharas[1], Ramon Lavado[3], Leigh Greathouse[1,2*]

1. Biology Department, Baylor University
2. Robbins College of Health and Human Sciences, Baylor University
3. Environmental Science, Baylor University
*Corresponding Author: Leigh Greathouse (leigh_greathouse@baylor.edu)
Address: Baylor University, One Bear Place #97068, Waco, TX 76798-7068, USA

KEYWORDS: *Fusobacterium nucleatum*, biofilm, indole, tryptophan metabolism, indole derivatives

Repositories: The authors confirm that the data supporting the findings of this study are available within the article and its supplementary materials. All data generated or analyzed during this study have been deposited in the Texas Data Repository (https://doi.org/10.18738/T8/3PYJA9) and are available upon request.

ABSTRACT

Pathogenic bacteria adapt to distinct disease environments, but whether these adaptations create therapeutic vulnerabilities remains unclear. *Fusobacterium nucleatum* has emerged as a key microbial player in colorectal cancer (CRC), yet the molecular mechanisms underlying its niche-specific virulence remain poorly defined. In this study, we investigated whether *F. nucleatum* displays disease-specific vulnerabilities to indole derivatives across 16 clinical isolates from CRC (n=6), Crohn's disease (n=6), healthy individuals (n=3), and an oral lesion (n=1). Using HPLC, we found that CRC-associated isolates produced a 3-4-fold higher level of endogenous indole compared to healthy isolates ($p < 0.05$), including indole, indole-3-acetic acid (IAA), and indole-3-carboxylic acid (I3CA), which is a metabolite enriched in CRC patient samples. Exogenous indole derivatives (0.25–2 mM) selectively suppressed CRC isolates, exhibiting 40-50% reductions in biofilm formation (*$p < 0.001$) and bacterial invasion of colon cancer cells, while non-CRC isolates displayed minimal or no response. Mechanistically, indole derivatives triggered distinct, disease-specific changes in virulence gene expression, with naïve Bayes classifier achieving 90% accuracy in disease association (AUC = 0.90–0.92; accuracy = 0.74–0.75; $p < 2 \times 10^{-1}$). This disease-adapted vulnerability was independent of subspecies identity, suggesting niche-specific phenotypes. Furthermore, exogenous indole derivatives, I3A and indole, strengthened epithelial tight junction gene expression while inhibiting bacterial invasion, demonstrating dual host-protective and pathogen-suppressive effects. These findings establish a key disease-adapted indole sensitivity that could enable selective suppression of pathogenic populations of *F. nucleatum* while preserving beneficial commensals.

INTRODUCTION

The discovery that *Fusobacterium nucleatum (F. nucleatum),* selectively colonizes colorectal cancer (CRC) tumors (1,2) transformed our understanding of host-microbe interactions in cancer, revealing that pathogenic bacteria not only respond to disease environments but may actively drive oncogenesis. In the oral cavity, *F. nucleatum* functions as a keystone bridging organism, coordinating biofilm and dental plaque formation, virulence traits strongly associated with periodontitis (3,4). Extending beyond the oral cavity, these highly virulent oral strains are thought to serve as the primary source of CRC *F. nucleatum* populations (5), raising a fundamental question: do bacteria that disseminate from oral to tumor sites retain their original virulence programs, or do they undergo niche-specific adaptation?

Supporting this latter notion, *F. nucleatum* demonstrates selective enrichment in biofilm-positive colon tumors (6), and these mucus-invasive bacterial biofilms themselves are highly carcinogenic in mice models (7). Biofilm-positive tumors correlate with reduced E-cadherin expression, activation of IL-6/STAT3 signaling, and increased crypt cell proliferation in adjacent normal colonic mucosa (8), directly linking biofilm formation to tumor progression. Importantly, biofilm formation is not unique to CRC, it has also been implicated in ulcerative colitis, Crohn's disease, and other forms of inflammatory bowel disease (IBD) (9–11), where *F. nucleatum* is likewise enriched and

contributes to disease aggravation (12,13). This pattern suggests that *F. nucleatum* may adapt its biofilm-forming capacity to distinct inflammatory and neoplastic environments, indicating that disease-specific selection pressures shape pathogen behavior.

Beyond biofilm formation, *F. nucleatum* also has an arsenal of mechanisms that provide fitness advantages in the host, including type V secretion systems and outer membrane adhesion proteins (e.g., CmpA, RadD, Aid1) that mediate bacterial interactions (14,15), and host cell binding (e.g. RadD, Fap2, FadA) (16–19). Coordinated bacterial interactions mediated by *F. nucleatum* adhesins, including CmpA and RadD, are critical for dual-species biofilm formation with *Streptococcus gordonii* in the oral cavity (20) and for promoting biofilm development with *Clostridioides difficile* on the intestinal mucosa (21). These findings suggest that *F. nucleatum* adhesion proteins play a critical role in recruiting other bacterial species during multispecies biofilm assembly. In addition, the ability of *F. nucleatum* to adhere to host cells via its adhesins enables it to activate pro-inflammatory signaling pathways. Among these, the best-characterized adhesin is FadA, which binds to E-cadherin to facilitate *F. nucleatum* invasion (22), activate β-catenin signaling, and differentially regulate inflammatory and oncogenic responses (18,23) while exacerbating DSS-induced colitis (24). Collectively, these findings highlight that mechanisms regulating both biofilm formation and the expression of outer membrane adhesion proteins in *F. nucleatum* are central to understanding and mitigating its pathogenic role in Crohn's disease and CRC. Yet a critical knowledge gap remains: what environmental signals regulate these virulence programs in the diseased environment?

One candidate regulator is indole. Indole is a known chemical effector in enteric bacterial communities (25–27), and can also exert diverse biological effects on *E. coli*, including modulation of biofilm formation and motility (28,29) and suppression of virulence programs (30,31). Interestingly, early work showed that indole enhanced biofilm in the reference strain *Fusobacterium nucleatum* ATCC 25586 (32); however, whether this response is conserved across clinical isolates from different disease contexts remains unknown and is critical to resolve. Given indole's complex, concentration-dependent effects in other pathogens and virulence factor expression (28,33), it is critical to understand indole's influence on *F. nucleatum* pathogenicity in association with Crohn's disease and CRC, an area that has been largely overlooked in *F. nucleatum* research.

Intestinal indole is produced mainly by *Escherichia coli*, *Bacteroides spp*., along with other gut microbes, through a single enzymatic reaction catalyzed by the *tnaA*-encoded tryptophanase, which hydrolyzes tryptophan into indole, pyruvate, and ammonia (34–39). Beyond the parent indole, gut microbes generate a range of tryptophan catabolites via alternative pathways (37). Specifically, *Clostridium sporogenes* metabolizes tryptophan into tryptamine, indolelactic acid, and indole-3-propionic acid (IPA) (40), whereas *Peptostreptococcus spp.* produce indoleacrylic acid and IPA (41). Enteric pathogens are known to exploit intestinal chemistry to program virulence gene expression, leading to successful colonization of the gastrointestinal tract (GI) tract (42). However, indole production contributes to microbial community homeostasis by suppressing virulence and limiting host cell colonization by several pathogens, including enterohemorrhagic *E. coli* (43), *Salmonella enterica* (44), and

Vibrio cholerae (45). Indole appears to elicit a biphasic chemotactic response, functioning as a chemorepellent for certain enteric pathogens (46,47) while acting as a chemoattractant for other bacterial species (48). These varying effects of indole likely depend on its concentration, which can fluctuate over a wide range (0.2 to 6.5 mM) in the GI tract (49), and have been hypothesized to act similarly to quorum sensing molecules (50). While a few studies have demonstrated *F. nucleatum's* capacity to produce indole via a tryptophanase enzyme (32,51), its ability to generate other indole derivatives remains largely unexplored. Moreover, it remains unclear whether exogenous indole and its derivatives, produced by enteric bacteria, act as chemoattractants or chemorepellents for *F. nucleatum*, and whether these metabolites uniformly promote biofilm formation or differentially modulate virulence factor expression. Notably, in *Citrobacter rodentium*, both self- and microbiota-derived indole have been shown to suppress virulence factor expression *in vitro* and during murine infection (43). Elucidating whether *F. nucleatum* itself produces disease-relevant indole derivatives, and how both endogenous and microbiota-derived indoles influence its biofilm development and virulence factor regulation, will be essential to define their contribution to CRC and inflammatory diseases.

F. nucleatum is highly genetically diverse, and has been subdivided into four distinct subspecies for decades: *nucleatum*, *polymorphum*, *vincentii*, and *animalis* (52–55). Substantial evidence now supports elevating the subspecies to species status (56–58); however, the field has been slow to fully abandon the subspecies designation. Research has increasingly concentrated on *F. nucleatum animalis*, given its pronounced enrichment in CRC relative to the other subspecies (6,57,59,60). A recent study further stratified tumor-derived strains of *F. nucleatum* subsp*. animalis* into two distinct clades, *FnaC1* and *FnaC2*, based on pangenomic analyses of genetic determinants (59). *FnaC1* was found to be predominantly associated with the oral cavity, whereas *FnaC2* was enriched in CRC, offering new insights into the genetic basis of *F. nucleatum's* contribution to CRC. However, subsequent analyses revealed that *FnaC1* may in fact be a rediscovery of *Fusobacterium watanabei* (61), and the candidate operons attributed to *FnaC2* are not unique to *F. nucleatum* subsp. *animalis* and therefore cannot alone explain subspecies-specific pathogenicity. Nonetheless, although *F. nucleatum* subsp. *animalis* is the subspecies most consistently detected in CRC, it remains essential to investigate the virulence potential of all *F. nucleatum* subspecies to clarify their individual contributions to CRC**.** This is particularly relevant given that all subspecies can invade CRC tissue and each exhibits pathogenic activity within the oral microbiome (57), which likely serves as a reservoir for CRC-associated populations. Upon invasion, *F. nucleatum* adapts to host conditions by upregulating metabolic pathways, quorum-sensing systems, and amino acid metabolism (62). It further reprograms its gene expression during the transition from planktonic to biofilm formation, again characterized by enhanced amino acid metabolism (63). Together, these observations illustrate how *F. nucleatum* fine-tunes its metabolic and transcriptional programs in response to environmental cues, prompting the question of whether isolates from disease-associated niches, rather than specific subspecies, harbor distinct virulence adaptations, including differential responses to indole production, that distinguish them from those in healthy or less inflamed sites.

The clinical relevance of indole metabolism in CRC is underscored by recent metabolomic findings. Our previous meta-analysis of metagenomic data using PICRUSt revealed that, in two of the three cohorts (64–66), tryptophan biosynthesis was enriched in tumor samples alongside increased tryptophan metabolism **(Supplementary Figure 1)**. These findings suggest that indole production may also be elevated in CRC. Supporting this, recent metabolomic analyses demonstrated altered levels of multiple tryptophan derivatives, including indole-3-acetic acid (IAA), indole-3-propionic acid (IPA), indole-3-aldehyde (I3A), and indole-3-carboxylic acid (I3CA), in CRC and adenoma patients compared with healthy controls (67). Notably, overall indole levels were significantly increased in CRC, driven particularly by elevations in I3A and I3CA, whereas IPA levels were consistently reduced in both adenoma and CRC (67). The most intriguing finding was an overall enrichment of *Fusobacteria* in fecal samples from patients with adenomas and CRC. Although elevated indole derivatives are typically associated with reduced inflammation and improved intestinal barrier integrity in healthy individuals (39), their role in CRC, particularly in relation to pathogens like *F. nucleatum*, remains unclear. It is not yet known whether *F. nucleatum* produces these disease-associated indole derivatives, nor whether isolates from diseased hosts preferentially generate them compared with those from other clinical contexts.

This study addresses our key hypothesis: *F. nucleatum* isolates from disease-associated niches may harbor distinct metabolic and virulence adaptations, including differential responses to indole, that distinguish them from healthy-site isolates, independent of subspecies identity.

MATERIALS AND METHODS

<u>*Bacterial growth conditions*</u>: *F. nucleatum* cultures were prepared by streaking glycerol stocks onto Columbia blood agar plates containing L-cysteine (0.5 g/L), hemin (5 mg/L), vitamin K (10 µg/L), and 5% defibrinated sheep blood. Plates were incubated anaerobically (90% N_2, 5% CO_2, 5% H_2) at 37°C for 48 to 72 hours. Single colonies were transferred to anaerobically reduced Columbia broth complete (containing L-cysteine (0.5 g/L), hemin (5 mg/L), and vitamin K (10 µg/L)) and grown under identical anaerobic conditions at 37°C for 24 to 48 hours before dilution and use in experiments. Culture purity was regularly confirmed through Gram staining.

<u>*Biofilm formation Assay:*</u> Biofilm formation by *F. nucleatum* isolates was assessed using previously described methods (32) with the following modifications. Biofilm biomass was quantified through crystal violet (CV) staining in 96-well plates. Planktonic cells were removed from 96-well plates and biofilms were washed once with sterile water, then air-dried for 5 minutes. Biofilms were fixed with methanol at 37°C overnight. Fixed biofilms were stained with 0.1% CV for 1 minute, washed with deionized water until negative control wells were clear, and air-dried for 5 minutes. Biofilms were destained with 30% acetic acid for 5 minutes. Absorbance was measured at 560 nm using a Sunrise Tecan microplate reader with 10-second shaking before measurement and analyzed with SparkControl software (v2.3).

<u>*Time course Biofilm Assay:*</u> Single colonies were grown anaerobically in Columbia broth to stationary phase (48 hours), then diluted to 0.3 OD_{600}. The diluted culture was used to inoculate six 96-well plates simultaneously, with two wells per plate left uninoculated

for contamination monitoring and background correction. Plates were collected at specific time points (12, 24, 36, 48, 60, and 72 hours), measured for optical density (OD_{600}) using a Cerillo Stratus plate reader, and subsequently processed using the previously described biofilm formation assay. For each time point, six biological replicates were included per isolate, totaling six CRC isolates (n = 36), six Crohn's isolates (n = 36), three healthy isolates (n = 18), and one oral isolate (n = 6).

Growth curves and indole treatment: To assess the impact of indole derivatives on *F. nucleatum* growth, 16 isolates were cultured anaerobically in Columbia broth and treated with five compounds: indole (CAS# 120-72-9), indole propionic acid (IPA, CAS# 830-96-6), indole-3-acetic acid (IAA, CAS# 87-51-4), indole-3-aldehyde (I3A, CAS# 487-89-8), and indole-3-carboxylic acid (I3CA, CAS# 771-50-6). Stock solutions were prepared in DMSO and diluted to final concentrations of 0.25, 0.5, 1, and 2 mM. DMSO-only controls were used to normalize solvent concentration. Tryptophan (1.25–10 mM) was used to assess bacteria's ability to convert tryptophan into indole affected their growth, and chloramphenicol (12.5–100 µg/mL) as negative controls. Mid-log phase cultures were diluted to a standardized OD_{600} (0.3), and each treatment was plated in triplicate. Growth curves were generated in a Cerillo Stratus reader (96-well format) at 37°C under anaerobic conditions for 72 hours, with OD_{600} measured hourly. For each isolate, three biological replicates were included per indole concentration, along with nine biological replicates for the untreated control and three media-only replicates to monitor contamination.

Quantification of total indole production with Kovacs reagent: *F. nucleatum* isolates were streaked from frozen stocks, and liquid cultures were initiated from single colonies. Starter cultures were grown to stationary phase, then diluted to 0.3 OD_{600} for inoculation. Media conditions included Columbia broth, Columbia + 5 mM tryptophan, Columbia + 1% glucose, and BHI complete (containing 0.5 g/L L-cysteine, 5 mg/L hemin, 10 µg/L vitamin K). Each media type was aliquoted into deep-well 96-well plates (1.2 mL/well) and inoculated with diluted culture. Samples were collected every 12 hours over 72 hours and 100 µL of culture was frozen at –20°C for indole quantification. Indole quantification was performed according to previously established methods (68). Briefly, samples were thawed slowly at 4°C, then centrifuged at 3,000 x g for 15 minutes to pellet the bacteria. The supernatant was transferred to a new 96-well plate and mixed with an equal volume of Kovac's reagent. After a 5-minute color development period, the OD_{560} was measured with shaking for 10 seconds on a plate reader (Sunrise Tecan's Magellan™ universal reader). Indole concentrations were determined using a standard curve generated with synthetic indole at concentrations ranging from 0 to 2000 µM. For each isolate, three biological replicates were collected for each media type, and three technical replicates were included for the standard curve. Totaling six CRC isolates (n = 18), six Crohn's isolates (n = 18), three healthy isolates (n = 9), and one oral isolate (n = 3).

Indole Derivative quantification by HPLC: *F. nucleatum* isolates were streaked from frozen stocks and cultured in liquid medium from single colonies. Starter cultures were grown to stationary phase, then diluted to 0.3 OD_{600} for inoculation. Each isolate was subsequently inoculated into Columbia broth and incubated for 48 hours, after which cultures were centrifuged and supernatants collected. Indole derivatives in the

supernatant were analyzed using high-performance liquid chromatography (HPLC) with fluorescence detection, following modified methods from Tuomola et al. (69) and Kema et al. (70). Chromatographic separation was performed on an UltiMate 3000 HPLC system equipped with an RS fluorescence detector (Thermo Fisher Scientific, Waltham, MA) using an Acclaim™ VANQUISH C-18 column (2.2 µm, 150 × 2.1 mm; Thermo Fisher Scientific). The mobile phase consisted of two eluents: eluent A (50 mM potassium dihydrogen phosphate in water, adjusted to pH 3.3 with phosphoric acid) and eluent B (70:30 mixture of eluent A and acetonitrile). Separation was achieved using a linear gradient elution program: 80% A/20% B (0-2 min), transitioning to 25% A/75% B (2-10 min), followed by 100% B (10-15 min), and returning to initial conditions of 80% A/20% B (15-20 min). The flow rate was maintained at 0.5 mL/min, and the column temperature was held constant at 30°C. Indole derivatives were detected by monitoring native fluorescence using excitation and emission wavelengths of 270 nm and 350 nm, respectively. The following compounds were analyzed: indole, indole-3-acetic acid (IAA), indole-3-propionic acid (IPA), indole-3-aldehyde (I3A), and indole-3-carboxylic acid (I3CA). Compounds were identified through co-chromatography with authentic standards, confirmed by LC-MS, and quantification was performed by integrating peak areas. Method linearity was assessed using standard solutions of each indole derivative prepared at various concentrations. System suitability parameters including retention time stability, peak resolution, and detector response were monitored throughout the analysis.

RNA Isolation for Virulence Factor Expression: *F. nucleatum* isolates were streaked from frozen stocks and grown in liquid culture from single colonies. Starter cultures were diluted to 0.3 OD_{600} and used to inoculate three 10 mL cultures per isolate and condition, each treated with either 1 mM indole derivatives (indole, IAA, I3A, I3CA, IPA) or an equivalent volume of DMSO as a control. Cultures were incubated anaerobically for 72 h, pooled, and centrifuged at 8000 × g for 5 min. Pellets were resuspended in 1.5 mL chilled lysozyme (0.5 mg/mL; Fisher BioReagents™, Cat# BP535-5), vortexed gently, and incubated at 37°C for ≤5 min. Lysis was performed by adding sodium acetate (Invitrogen™, Cat# AM9740) and SDS (VWR, Cat# 97064-496) to final concentrations of 0.1 M and 1%, respectively. After mixing to a gelatinous consistency, an equal volume of Acid-Phenol:Chloroform (pH 4.5) with IAA (25:24:1; Invitrogen™, Cat# AM9720) was added, vortexed, incubated at 65°C for 15 min with intermittent vortexing, then chilled and centrifuged at 18,000 × g (4°C, 20 min). The aqueous layer was transferred to Phasemaker™ tubes (Invitrogen™, Cat# A33248) containing one-fifth volume chloroform, vortexed, and centrifuged at 16,000 × g (4°C, 20 min). The aqueous phase was precipitated with sodium acetate (0.75 M final) and two volumes of ethanol, then centrifuged at 18,000 × g (4°C, 20 min). Pellets were washed with 80% ethanol, air-dried, and resuspended in nuclease-free water. RNA quality was assessed using a Nanodrop spectrophotometer, and concentrations determined with the Qubit RNA BR Assay Kit (Invitrogen™, Cat# Q10210). DNA contamination was removed using the DNA-free DNase Treatment & Removal Kit (ThermoFisher). Ten micrograms of RNA were treated with TURBO DNA-free™ Kit (Invitrogen™, Cat# AM1906). For cDNA synthesis, 500 ng of RNA was reverse transcribed using the High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems™, Cat# 4368814). cDNA was stored at –20°C.

Real-time quantitative PCR to quantify virulence factor expression: Real-time quantitative PCR (qPCR) was performed using PowerUp™ SYBR™ Green Master Mix (Thermo Fisher Scientific, Cat. No. A25741) on a QuantStudio 6 Flex Real-Time PCR System (Applied Biosystems). Primers targeting key virulence factors (*tnaA*, *aid1*, *fomA*, *fap2*, *fadA*, *cmpA*, *radD*, and *16s* for reference control) of *F. nucleatum* were either designed in-house or sourced from published literature **(Supplementary Table 1)**. Prior to experimental analysis, primer efficiency and specificity were evaluated for each target gene using standard curves generated from known concentrations of genomic DNA extracted from the 16 *F. nucleatum* isolates. Each experimental reaction was performed in technical triplicates in a total volume of 10 µL containing 5 µL of 2X PowerUp™ SYBR™ Green Master Mix, 250 nM of each forward and reverse primer, and 1.5 µL of template DNA at the appropriate dilution. No-template controls were included in each run to monitor potential contamination. Thermal cycling conditions consisted of UDG activation at 50°C for 2 minutes, followed by DNA polymerase activation at 95°C for 10 minutes, and 40 cycles of denaturation at 95°C for 15 seconds and annealing/extension at 60°C for 1 minute. A melt curve analysis was performed following amplification by heating samples from 60°C to 95°C at a ramp rate of 0.1°C/second to confirm amplification specificity. Fold expression was calculated using the ΔΔCt method, where ΔΔCt = (Ct_target − Ct_16S)_treated − (Ct_target − Ct_16s)_control. Relative expression was quantified as 2^-ΔΔCt, indicating fold change normalized to 16S gene and relative to the control.

Cell culture and cell treatment: Caco-2 and HT-29 cell lines were received as a generous gift from Dr. Christie Sayes (Baylor), thawed from liquid nitrogen and cultured in T75 or T175 flasks using Dulbecco's Modified Eagle Medium (DMEM; Corning, Kennebunk, ME, USA) supplemented with 10% fetal bovine serum (FBS; Equitech-Bio, Kerrville, TX, USA) and 200 µg/mL penicillin-streptomycin (Lonza, Basel, Switzerland). Cells were maintained at 37 °C in 5% CO_2 and passaged at 70–90% confluency and mycoplasma testing was performed biweekly using MycoStrip kits (InvivoGen). Cells were seeded in 12-well plates at 0.1 × 10□ cells/well and grown to 95% confluence. After washing twice with Dulbecco's phosphate-buffered saline, prewarmed treatment medium was added, and cells were incubated at 37 °C with 5% CO_2 for 24 h. Treatment conditions included DMEM containing 200 µg/mL penicillin–streptomycin and 1 mM indole derivatives (dissolved in DMSO). Control media contained 0.1% DMSO without indole, and each treatment had four biological replicates. Lipopolysaccharide (LPS) served as a control due to its known ability to reduce barrier integrity via downregulation of target genes. After 24 hours of exposure, cells were lysed, and total RNA was isolated. RNA extraction and removal of genomic DNA contamination were performed as previously described. qPCR was performed using a QuantStudio 6 system (ThermoFisher) in 384-well plates (10 µL reaction volume): 2 µL cDNA, 5 µL PowerUp SYBR Green Master Mix (ThermoFisher), 0.2 µL each of forward and reverse primers (10 mM), and 3.6 µL nuclease-free water. Primer sequences are listed in **(Supplementary Table 1)**. Thermocycling conditions: 50 °C for 2 min, 95 °C for 10 min, then 40 cycles of 95 °C for 15 s and 55.5 °C for 1 min, followed by melting curve analysis. Cycle threshold (Ct) values from QuantStudio software were analyzed using the Livak–Schmittgen (ΔΔCt) method. Beta-actin served as the housekeeping gene.

Fold changes were calculated for five genes associated with tight and adherens junctions.

Bacterial Invasion Assay: Caco-2 cells were cultured as described above in 24-well plates. *Fusobacterium nucleatum* isolate SB-CTX3Tcol3 was grown to log phase in Columbia broth under anaerobic conditions. Cells at ~85% confluence were washed with PBS and incubated in antibiotic- and serum-free DMEM containing 1 mM indole derivatives. Control groups included DMEM with 200 µg/mL penicillin–streptomycin or vehicle-only media. Cells were then infected with *F. nucleatum* at a multiplicity of infection (MOI) of 50:1 bacteria:cell for 2 or 4 h. Following infection, cells were washed three times with PBS and incubated with DMEM containing 200 µg/mL penicillin–streptomycin at 37 °C for 1 h to eliminate extracellular bacteria. Cells were washed an additional three times with antibiotic-containing DMEM, lysed with sterile deionized water, and lysates were serially diluted in complete Columbia broth. Two dilution factors per condition were plated in triplicate on Columbia blood agar supplemented with 0.5 g/L L-cysteine, 5 mg/L hemin, 10 µg/L vitamin K, and 5% defibrinated sheep blood. Plates were incubated anaerobically for 48 h or until colony formation, and colony-forming units (CFU/mL) were calculated as the average across dilutions and replicates.

Statistical Methods: All analyses were conducted in R (v4.3.2) using RStudio (v2024.12.0.467). Detailed statistical analyses are provided in figure legends, with methods implemented as follows: Raw data were processed using the tidyverse package to standardize structure and address missing values. One-way ANOVA (aov()) was used to assess group differences, with significance set at $\alpha = 0.05$. Post-hoc comparisons were performed using Tukey's HSD (TukeyHSD()) to control for family-wise error. Data are shown as mean ± SEM unless noted. Plots were generated with ggplot2. For classification, we used Naïve Bayes models (caret, naivebayes) to predict disease status (CRC, Crohn's, Healthy) based on fold gene expression and gene direction data. Fold gene expression change was quantified using the ΔΔCt method. Gene direction was determined based on fold change values: genes with a fold change greater than 1.25 were classified as "Up", those between 0.75 and 1.25 as "No Change", and those below 0.75 as "Down". Oral lesion isolates were excluded due to small sample size. Models were randomly split, trained on 70% of the data and tested on 30%, with 10-fold cross-validation used to assess accuracy, AUC, precision, recall, F1 score, and detection rate **(Supplementary Figure 2)**. We validated model performance by checking class priors **(Supplementary Figure 3)**, calculating mutual information **(Supplementary Figure 4B),** Chi-squared and Cramér's V **(Supplementary Figure 5)**, analyzing feature independence via correlation **(Supplementary Figure 4A)**, and evaluating predictor importance through permutation-based accuracy loss **(Supplementary Figure 3)**.

RESULTS

Elevated indole production is a distinctive feature of colorectal cancer-associated *F. nucleatum* isolates.

We obtained 15 clinical isolates of *F. nucleatum* from collaborators, which were derived from patient biopsies representing distinct disease states **(Table 1)**. These isolates originated from a range of conditions, including negative, non-specific biopsies from colon cancer screenings (Healthy), active Crohn's disease in the sigmoid colon

(Crohn's), and positive colon cancer screenings (CRC). Based on the metadata on patient status at the time of sampling, the isolates were categorized into groups based on their disease association: Healthy (3 isolates), Crohn's (6 isolates), CRC (6 isolates). We also obtained the reference strain *F. nucleatum subsp. nucleatum* ATCC 25586, originally isolated from an oral lesion, as it is the most frequently cited representative in the literature (24,81,82). To assess and compare the indole production capabilities of these isolates, we sampled bacterial cultures throughout the growth phase of *F. nucleatum* and quantified indole production with Kovacs' reagent **(Figure 1A)**. Indole production increased markedly around 36 hours, suggesting this process primarily occurs during the transition from exponential to stationary phase. Notably, *F. nucleatum* isolates from individuals with CRC showed the greatest variability in indole production; while many produced amounts comparable to other isolates, four strains generated substantially higher levels (SB-CTX3Tcol3, SB-CTX47T, CC7_3JVN3C1, and CC2_3FMU1). Driven mainly by these four isolates, CRC-associated isolates exhibited significantly higher indole production across all growth times compared to the remaining isolates **(Figure 1B)**. This trend persisted even across different growth media (**Supplementary Figure 6**), suggesting CRC-associated isolates exhibit heightened indole production. When grouped by subspecies, the distinction seen between CRC-associated isolates and other disease groups became less pronounced **(Figure 1C)**. Whereas *vincentii* exhibited minimal indole production, the other subspecies produced markedly higher levels, with *polymorphum* surpassing all others across most time points. This observation, however, is based on a single isolate. These findings indicate that elevated indole production is a defining metabolic characteristic of CRC-associated isolates, independent of subspecies classification.

Disease-associated isolates exhibit elevated production of specific indole derivatives and have distinct indole production profiles.

After confirming elevated indole production in CRC-associated isolates using Kovacs' reagent, we employed targeted HPLC analysis to characterize the indole profiles, with a focus on CRC-associated derivatives reported in the literature (67). Spent media collected after 48 hours of growth was analyzed, with derivative identification and quantification performed against authenticated standards. Across all disease states, *F. nucleatum* isolates produced all five targeted indole derivatives—indole, I3A, I3CA, IAA, and IPA—with indole being the predominant metabolite **(Figure 2A)**. Although all targeted derivatives were detected, IPA levels were below the detection limit in most samples **(Supplementary Figure 7A).** *F. nucleatum* isolates from CRC, Crohn's, and oral lesions trended toward higher production of indole derivatives—indole, I3A, I3CA, and IAA—relative to healthy isolates. Most notably, the indole was significantly elevated in both CRC and Crohn's disease isolates compared to healthy isolates ($p<0.05$). The effects of indoles on host physiology and immunity are highly dependent on the specific indole species present, with individual derivatives exerting distinct and sometimes opposing roles in regulating inflammation, barrier function, and tumor progression (76,84–88). We therefore investigated whether *F. nucleatum* isolates from distinct disease states exhibited preferences for producing specific indole derivatives within their metabolic profiles. To address this, we performed a point-biserial correlation analysis between the indole derivatives and *F. nucleatum* isolates. Correlation analysis (**Figure 2B**) revealed distinct associations between specific *F.*

nucleatum isolates and their indole derivative profiles. CRC-associated isolates (CC2_6JVN3, CC7_3JVN3C1, SB-CTX47T, and SB-CTX3Tcol3) demonstrated strong positive correlations with specific indole derivatives. Among healthy isolates, indole derivative production exhibited a trend toward a negative correlation. The distinct indole production profiles of each isolate suggest a preference for certain indole derivative production over others. To further examine this preference, we calculated derivative-to-parent indole ratios for each isolate. These ratios normalize for baseline indole production and reflect the extent to which tryptophan metabolism is directed into specific derivative pathways, with higher ratios indicating greater relative production of that derivative. Examination of individual derivative-to-parent indole ratios **(Figure 2C)** revealed that CRC isolates—particularly CC2_6JVN3, SB-CTX47T, and SB-CTX3Tcol3—exhibited markedly elevated I3A ratios (0.52, 0.79, and 0.89, respectively). Given that I3A is enriched in adenoma and CRC patients (67), these findings suggest that *F. nucleatum* isolates may contribute to the elevated concentrations of this metabolite observed in disease contexts. Crohn's disease isolates displayed intermediate indole derivative production, with ratios generally lower than those of CRC isolates but higher than healthy controls. When isolates were grouped by disease state, differences in indole derivative production were more pronounced than when grouped by subspecies, indicating that disease association exerts a stronger influence on indole production **(Supplementary Figure 7B)**. Collectively, these results demonstrate that disease-associated *F. nucleatum* strains—particularly those isolated from CRC—exhibit distinct metabolic capacities for endogenous indole derivative production, especially I3A, in line with observations from fecal samples of CRC patients.

Exposure to indole derivatives alter *F. nucleatum* growth kinetics in a dose-dependent manner without bactericidal effects.

Indole affects bacterial physiology in a concentration-dependent manner (35). In the gut, fecal indole levels typically range from 0.3 to 6.6 mM (68), which are similar to the concentrations observed *in vitro* (0.5 to 5 mM) (89). Shifts in indole concentration can dramatically affect gut bacteria. At lower levels (0.5–2 mM), indole modulates *E. coli* virulence by influencing biofilm formation, gene expression, and stress responses (28,31), while at higher levels (4–5 mM) it inhibits bacterial growth and division (90). Before testing the exogenous effects of these indole derivatives on *F. nucleatum* isolate virulence, we first evaluated whether low concentrations—previously shown to modulate *E. coli* virulence—exhibited antimicrobial activity against our isolates. To determine how increasing indole levels within this biologically relevant range influence *F. nucleatum* growth, we continuously monitored the isolates over 72 hours with an automated plate reader while exposing them to escalating concentrations of indole. The resulting growth curves were analyzed using the Gompertz mathematical model. The Gompertz model is used to estimate growth parameters such as maximum growth rate, carrying capacity (OD_{max}), and inflection time, making it useful for comparing bacterial responses to increasing concentrations of indole. This analysis showed that increasing concentrations of indole derivatives restrict maximum growth (OD_{max}) but do not necessarily slow growth rate (μ) or delay inflection time (**Figure 3**), a kinetic profile distinct from the severe growth inhibition characterized by a decreased growth rate and a pronounced delay in inflection time observed with antibiotic treatment **(Supplementary Figure 8)**. CRC isolates exhibited the highest sensitivity to indole at 2

mM concentration, with total population (OD max) reductions of 45% for indole, 33% for I3A, 44% for I3CA, 40% for IAA, and 44% for IPA compared to the control (**Figure 3A and Table 2**). In contrast, Crohn's disease isolates showed substantially lower sensitivity at 2 mM (20% for indole, 15% for I3A, 1% for I3CA, and 0% for both IAA and IPA) **(Figure 3B and Table 3)**, as did healthy isolates (13% for indole, 0% for I3A, 0% for I3CA, IAA, and IPA) **(Figure 3C and Table 4)**. Oral lesion isolates demonstrated minimal sensitivity across all derivatives at 2 mM (0% for indole, 10% for I3A, and 0% for I3CA, IAA, and IPA) (**Figure 3D and Table 5**). Interestingly, the inflection time—where the growth curve shifts concavity, marking the transition from the exponential phase to a deceleration phase—remained largely unchanged for Crohn's disease, healthy, and oral isolates within the 0.25 to 1 mM indole range. At 2 mM indole, the isolates exhibited a slight increase in inflection time, suggesting a minimal impact on growth. The CRC isolates exhibited minimal to no variation in inflection time across all concentrations for I3CA, IAA, and IPA treatments. However, indole at 1mM and 2 mM and I3A at 2 mM displayed negative inflection times for the CRC isolates. The Gompertz model failed to converge robustly for inflection time in several CRC isolate treatments, resulting in mathematical anomalies **(negative values, Table 2)**. This suggests the model assumptions were not well-suited to describe the growth dynamics in the presence of high indole concentrations. However, the consistent reduction in the final population size (OD_{max}) for CRC isolates, without a corresponding decrease in the maximum growth rate, is characteristic of an early cessation of exponential growth. This phenomenon is reminiscent of the 'indole pulse' observed in *E. coli*, where elevated indole concentrations trigger an earlier entry into stationary phase (91). Rather than completely suppressing bacterial proliferation, indole derivatives appear to have altered the fundamental growth dynamics of the bacteria, forcing CRC-associated isolates into stationary phase early, resulting in a lower maximum population without demonstrating true bactericidal properties. These results suggest indoles may be affecting metabolic pathways or cellular processes predominantly in CRC-associated isolates that influence growth patterns while still allowing bacterial survival and limited proliferation as an adaptation to its environment.

F. nucleatum **isolates exhibit disease-associated differences in biofilm formation independent of subspecies identity.**

F. nucleatum isolates display subspecies-specific virulence profiles, characterized by significant variation in biofilm-forming capacity (92); however, in vivo inflammatory responses in mouse models do not consistently correlate with subspecies identity (93). These data suggest that the virulence of individual isolates associated with disease states cannot be fully accounted for by subspecies designation alone. We hypothesized that isolates from disease states adapt their virulence mechanisms to the environment and would therefore exhibit distinct biofilm-forming capacities when grouped by disease. To test this, we monitored biofilm development at 12-hour intervals over 72 hours. At 12 hours, several isolates showed minimal biofilm formation, with CC2_3FMU1 (CRC) displaying the highest early biofilm formation **(Figure 4A)**. By 36 and 48 hours, when the isolates transitioned from the exponential to the stationary phase, most had reached a stable phase of biofilm development, with EAVG002-7_1 (Crohn's), SB-CTX3Tcol3 (CRC), and EAVG_014 (Healthy) consistently forming the most robust biofilms, while isolates such as CC2_6JVN3 (CRC), CC53 (CRC), and CC7_3JVN3C1 (CRC)

produced minimal biofilm. By 72 hours, a clear distinction emerged between high and low biofilm-forming isolates **(Supplementary Figure 9)**, with the high biofilm formers predominantly associated with Crohn's disease **(Figure 4B)**. When stratified by subspecies, the pronounced differences in biofilm-forming capacity observed across disease groups diminished, with statistical significance primarily driven by *F. nucleatum* subsp. *polymorphum*, which showed the lowest biofilm production at 12 and 24 hours but eventually reached levels comparable to the other subspecies at later time points **(Figure 4C)**. Interestingly, the CRC-associated isolates generally exhibited the lowest baseline biofilm formation, with the notable exception of SB-CTX3Tcol3, which was among the strongest biofilm former and indole producer. Collectively, these data demonstrate that *F. nucleatum* isolates display substantial variability in biofilm-forming capacity, mirroring differences in their indole derivative production profiles, with this variability more closely linked to disease association than to subspecies identity.

Biofilm formation by CRC-associated *F. nucleatum* isolates is positively correlated with indole production.

Bacteria utilize indole production and biofilm formation as stress-adaptive strategies, enhancing their survival in hostile environments. Consequently, indole synthesis and the transition to biofilm formation often occur simultaneously during specific growth phases (94). To better understand the relationship between indole production and biofilm formation over time, we performed correlation analysis on the two independent time-course studies. This analysis revealed that *F. nucleatum* isolates simultaneously produce indole and form biofilms upon entering the stationary phase (36–48 hours), a stage triggered by cellular stress from reduced nutrient availability. At the 36-hour time point, clear differences emerged in the relationship between indole production and biofilm formation across the disease groups **(Figure 5)**. In healthy-associated isolates, total indole production was negatively correlated with biofilm formation ($R^2 = 0.72$, $p < 0.001$), suggesting that higher indole levels may be linked to reduced biofilm biomass **(Figure 5C)**. Crohn's-associated isolates showed only a weak, non-significant negative correlation ($R^2 = 0.04$, $p = 0.051$), indicating little evidence of a consistent relationship between indole production and biofilm formation in this group **(Figure 5B)**. In contrast, CRC-associated isolates displayed a positive correlation ($R^2 = 0.50$, $p < 0.001$), suggesting that greater indole production may be associated with enhanced biofilm formation at this time point **(Figure 5A)**. At 48 hours, the negative correlation between total indole production and biofilm formation strengthened for healthy-associated isolates ($R^2 = 0.88$, $p < 0.001$). Crohn's-associated isolates continued to show a moderate negative correlation ($R^2 = 0.10$, $p < 0.001$), and the oral associated isolates exhibited no significant correlation ($R^2 < 0.01$, $p = 1$) **(Figure 5D)**. However, it should be noted that there is only one oral isolate represented in this study. The positive correlation between indole production and biofilm formation at 48 hours strengthened for CRC-associated isolates ($R^2 = 0.70$, $p < 0.001$) **(Figure 5A)**. These results indicate that indole production and biofilm formation tend to be inversely related in healthy- and Crohn's-associated isolates, whereas CRC-associated isolates show a positive correlation across all time points. Together, these findings suggest that endogenous indole production may influence biofilm formation in *F. nucleatum*, and that CRC-associated isolates could be particularly responsive to indole, raising the

possibility that exogenous indole exposure also may enhance their biofilm development compared with other other isolates.

Exposure to exogenous indole derivatives significantly diminish biofilm formation in CRC-associated *F. nucleatum* isolates.

To determine whether exogenous indole derivatives influence biofilm formation, bacterial isolates were cultured with increasing concentrations of the derivatives, and biofilm formation was quantified after a 72-hour incubation period. We selected a concentration range (0.5–2 mM) that reflects levels produced by strong indole producers such as *E. coli*, which are known to modulate biofilm formation, gene expression, and stress responses in *E. coli* (28,31). Importantly, these concentrations did not exert antimicrobial effects on our isolates **(Figure 3)**. Treatment with exogenous indole and its derivatives (I3A, I3CA, IAA, and IPA) led to distinct alterations in biofilm formation, with responses differing according to disease association **(Figure 6)**. CRC-associated isolates demonstrated significant biofilm inhibition at higher concentrations of several indole derivatives. Indole treatment at 2.0 mM reduced biofilm formation by approximately 25% ($p<0.01$) compared to untreated controls. Similarly, I3CA and IAA at 2.0 mM substantially inhibited biofilm formation by approximately 40% ($p<0.01$). Interestingly, I3A exhibited a biphasic response in CRC isolates, with enhanced biofilm formation (approximately 25% increase) at 0.5 mM followed by statistically significant inhibition at higher concentrations ($p<0.01$). Healthy-associated isolates exhibited variable responses with modest effects overall. A slight but significant increase in biofilm formation was observed at 1.0 mM indole ($p<0.05$), while other derivatives showed minimal impact across most concentrations. Unlike CRC and oral isolates, healthy-derived strains demonstrated neither strong stimulation nor inhibition patterns, suggesting a more neutral sensitivity to indole signaling at these concentrations. Crohn's disease-associated isolates showed intermediate responses with greater variability. The effects were generally less pronounced than those observed in CRC or oral isolates, with most treatments resulting in slight, non-significant deviations from baseline. I3CA at 2 mM showed moderate inhibitory effects but did not reach statistical significance for the Crohn’s associated isolates. The orally derived *F. nucleatum* isolate displayed a consistent stimulatory response to indole derivatives across all tested concentrations. This effect was most prominent at lower concentrations (0.5-1.0 mM), with statistical significance ($p<0.05$) observed for indole, I3A, I3CA, and IAA treatments. Even at higher concentrations (2.0 mM), where CRC isolates showed inhibition, oral isolates maintained biofilm levels above untreated controls. The most pronounced stimulation was observed with indole at 1.0 mM and I3CA at 0.5 mM, both inducing approximately 20% enhancement in biofilm formation. These results reveal that exogenous indole and its derivatives modulate *F. nucleatum* biofilm formation in a disease-context-dependent manner, with CRC isolates being particularly susceptible to biofilm inhibition at higher indole concentrations, while our single oral isolate demonstrated enhanced biofilm formation across all concentrations tested.

Indole derivatives elicit distinct virulence expression patterns in *F. nucleatum* based on isolate origin and disease association.

To assess the indole derivative influence on bacterial virulence potential, we analyzed transcriptional alterations in key virulence-associated genes (*tnaA*, *fomA*,

fadA, *fap2*, *radD*, *cmpA*, and *aid1*) across our *F. nucleatum* isolates following 72-hour indole exposure. Exposure to exogenous indole derivatives elicited markedly distinct transcriptional responses among the *F. nucleatum* isolates analyzed. RT-qPCR analysis revealed that the expression patterns of key virulence-associated genes varied not only with the specific indole derivative but also in an isolate-dependent manner (**Figure 7A**). Specifically, *aid1* showed a diverse expression profile with significantly higher expression in isolates EAVG_029, EAVG_030, and EAVG002-7_1 under indole exposure, yet its expression was significantly reduced in CC2_3FMU1 and CC2_6JVN3. Genes involved in adhesion and immune modulation, including *fomA*, *radD*, and *fadA*, also exhibited heterogeneous responses. Notably, *fadA* expression was decreased across all treatments in subspecies *vincentii*, while it was increased in several tumor-associated isolates (CC53, CC7_3JVN3C1, and SB-CTX3Tcol3). *fap2* expression was generally low or undetectable, yet select isolates (EAVG_012, EAVG_029, and CC53) showed significant expression in response to certain indole derivatives. Examining the disease-association expression patterns, *fomA* showed opposite responses to indole: its expression was reduced in healthy isolates but increased in most CRC-associated isolates, suggesting an adaptive environmental response. Due to the heterogeneous response in virulence factor expression, it was difficult to discern a clear pattern in virulence factor change. To evaluate whether indole-induced responses differed among disease groups, we applied a Naïve Bayes classification model to test whether expression patterns could predict the disease association of each isolate **(Figure 7B)**. This approach provided insight into how virulence gene expression varied across disease groups and evaluated whether indole-induced directional changes could be used to classify isolates by disease association. To capture directionality, we applied a threshold where fold > 1.25 was considered “Up”, 0.75 < fold ≤ 1.25 as “No change”, and fold < 0.75 as “Down”. Using both the raw fold change data and the gene direction threshold from RT-qPCR, the Naïve Bayes model was able to predict disease association with high accuracy (74% and p<2.2e-16, 76% and p<2e-16, respectively) **(Figure 7B, confusion matrix)**. The two models also exhibited a strong true positive rate, suggesting the models were very good at correctly identifying positive cases (AUC = 0.9019 and AUC = 0.9221 respectively). Together, these results suggest that indole derivatives modulate *F. nucleatum* virulence factor expression in a disease-dependent manner, with isolates from the same disease environment showing similar responses to indole.

Indole derivatives alter tight junction integrity in CRC cells and restrict bacterial invasion.

Bacterial interactions with the epithelium are central to infection, as invasion allows pathogens to both evade host immunity and disseminate deeper into the tissue. In CRC, invasion is linked to biofilm formation, with tumor biofilms resulting in deeper penetration into the tumor mass (8). Intracellular invasion by *F. nucleatum* through its virulence factors drives oncogenic signaling and cancer cell migration (18,95,96), and enables its persistence in liver metastases, where it co-occurs with other Gram-negative anaerobes in primary and matched tumors (97). Our data have shown that both endogenous and exogenous indole derivatives modulate *F. nucleatum* virulence by altering biofilm formation and virulence factor expression. However, such transcriptional changes do not necessarily manifest as functional virulence traits, such as active

cellular invasion. To continue to evaluate the impact of indole derivatives on virulence, we selected the CRC isolate *F. nucleatum* (SB-CTX3Tcol3), which displayed the highest levels of indole production and biofilm formation, indicating that it would be most responsive to indole-induced modulation. A bacterial invasion assay using SB-CTX3Tcol3 (CRC) and Caco-2 cells showed that indole treatment after 2 or 4 hours markedly inhibited invasion, producing an effect comparable to the presence of antibiotics in the media **(Figure 8A)**. Notably, I3CA exhibited the strongest inhibitory effect, reducing invasion by ~20% after just 2 hours. By 4 hours, indole, I3A, and I3CA suppressed invasion by more than 50%, whereas IAA and IPA had no detectable impact on SB-CTX3Tcol3 invasion. These data demonstrate direct alteration of *F. nucleatum* isolate virulence and highlight that indole-induced effects are indole species-specific. *F. nucleatum* relies on virulence factors such as FadA, which binds the adherens junction protein E-cadherin to promote invasion (18,98). Since indoles enhance barrier integrity by regulating tight and adherens junctions (99), we analyzed transcript levels of these junctional proteins in Caco-2 cells to determine whether such changes could explain the reduced invasion. When Caco-2 cells were treated with indole derivatives at the same concentrations used in the invasion assay, transcripts for tight and adherens junction genes were predominantly upregulated across most derivatives **(Figure 8B)**. Notably, I3A exerted the strongest effect on transcript levels, followed by indole. In contrast, I3CA, IAA, and IPA showed little to no impact on transcript expression. HT-29 cells displayed similar, though less pronounced, trends, suggesting that these effects extend to other cell lines. The analysis suggests that indole derivatives act via at least two distinct mechanisms to inhibit *F. nucleatum* invasion: Indole and I3A appear to reduce invasion by strengthening the host barrier through the transcriptional upregulation of Caco-2 junction genes **(Figure 8B)**. In contrast, the potent inhibitory effect of I3CA occurs despite minimal host cell transcriptional change, indicating a separate mechanism potentially involving direct action on bacterial virulence factors.

DISCUSSION

We sought to address a fundamental question in cancer microbiology: do pathogenic bacteria that colonize tumor environments retain the characteristics of their original niche populations, or do they undergo disease-specific adaptation that alters their metabolic and virulence programs to allow them to establish a new niche? Using *F. nucleatum* as a model, a pathogen enriched in CRC tumors that can originate from oral habitats, we profiled indole production and response across 16 clinical isolates from CRC, Crohn's disease, healthy individuals, and oral lesions. Our findings reveal that: (i) CRC isolates produced 3-4 fold-higher endogenous indole levels than healthy isolates, with preferential production of I3A, a tryptophan metabolite elevated in CRC patient samples, suggesting *F. nucleatum* contributes to disease-associated metabolite shifts; (ii) exogenous indole derivatives selectively suppressed CRC isolates at physiologically-relevant concentrations (1-2 mM), reducing biofilm formation by 40-50% and epithelial invasion by ~50%, while minimally affecting Crohn's or healthy isolates, revealing a therapeutic target; (iii) this disease-specific vulnerability was independent of subspecies identity and manifested as distinct transcriptional responses in virulence genes, with our Naïve Bayes classifier achieving over 90% accuracy in predicting disease association; (iv) indole derivatives strengthened epithelial tight junctions while simultaneously

inhibiting bacterial invasion, providing dual host-protective and pathogen-suppressive effects. These findings reveal a fundamental observation: pathogenic bacteria that have adapted to disease environments acquire metabolic sensitivities that can be targeted for selective suppression or elimination, establishing an opportunity for microbiome interventions that target pathogenic populations while preserving beneficial commensals.

Indole production is a defining feature of gut-resident bacteria, with the gut microbiota solely responsible for converting tryptophan into indole and its derivatives, as these metabolites are absent in gnotobiotic mice lacking a microbiome (100). Over 85 bacterial species, including numerous pathogens, have been identified as indole producers (34), an area which our laboratory is actively investigating. Our study demonstrates that clinical isolates of *F. nucleatum* can produce several indole derivatives linked to adenoma formation and CRC progression (67). Clinical isolates associated with CRC produced significantly higher levels of endogenous indole and its derivatives compared to isolates from other disease contexts. Furthermore, CRC-associated isolates displayed heightened sensitivity to exogenous indole derivatives, as escalating concentrations promoted premature entry into the stationary phase and markedly reduced biofilm formation. These observations prompt an important question: why would pathogens like *F. nucleatum* produce indoles that seemingly suppress virulence-associated behaviors that otherwise would enhance their fitness within the host environment? As mentioned previously, indole has a biphasic chemoattraction effect in *E. coli*, where indole can induce multidrug resistance (31) and promote biofilm formation (28), but at higher concentrations it can inhibit bacterial growth and division (90). More specifically, this phenomenon, termed the 'indole pulse' in *E. coli*, occurs when elevated indole concentrations trigger an earlier entry into stationary phase, thereby enhancing the long-term viability of wild-type *E. coli* cells during extended stationary phase (91). At the highest concentration of exogenous indoles tested (2 mM), a similar pattern was observed specifically among the CRC-associated isolates. At this concentration, CRC-associated isolates exhibited reductions in total population (OD_{600}) of 45% for indole, 33% for I3A, 44% for I3CA, 40% for IAA, and 44% for IPA compared with the untreated control. Similar to observations in *E. coli*, the indole-induced early entry into stationary phase at a lower population density may enhance long-term viability and promote fitness under these conditions. These concentration-dependent effects underscore that indole acts as a tunable signaling molecule rather than solely an anti-virulence agent. At concentrations of 500-600 µM, indole represses several virulence-associated phenotypes in EHEC; however, at concentrations of 125 µM or lower, it stimulates the expression of type III secretion system-related proteins and virulence phenotypes (30). This biphasic response highlights indole's dual regulatory role in EHEC virulence. Therefore, the endogenous indole-sensing system likely differs from exogenous intercellular signal transduction systems. Indole concentrations in the gut fluctuate dynamically in response to changes in the abundance of indole-producing bacteria and variations in the intestinal environment. An in vivo mouse study further demonstrated that dietary tryptophan levels significantly influence bacterial tryptophanase activity and, consequently, indole production within the gut microbiota (101). The divergent indole responses we observed likely reflect evolutionary adaptation to distinct ecological niches. CRC-associated isolates, which

encounter altered metabolite profiles in the tumor microenvironment, may have evolved heightened indole sensitivity as an environmental cue for regulating virulence programs. In contrast, Crohn's associated isolates, which colonize chronically inflamed tissue, exhibit indole insensitivity, perhaps reflecting adaptation to persistent inflammatory conditions where constitutive biofilm formation confers a survival advantage.

Indole production also maintains microbial community homeostasis by suppressing virulence and limiting host cell colonization by several pathogens, including enterohemorrhagic *E. coli* (43), *Salmonella enterica* (44), and *Vibrio cholerae* (45). Building on this idea, we investigated whether indole derivatives could inhibit invasion of our isolates into CRC cell lines. At a concentration of 1 mM, the indole derivatives indole, I3CA, and I3A reduced invasion of the high-indole, high-biofilm-forming isolate SB-CTX3Tcol3 (CRC) into Caco-2 cells by approximately 50% after 4 hours. At this concentration, indole derivatives significantly altered the expression of *F. nucleatum* virulence genes involved in host cell binding. Analysis of indole-induced transcriptional responses using a Naïve Bayes classification model revealed that virulence gene expression patterns could accurately predict disease association. Collectively, these results suggest that isolates belonging to the same disease category exhibit conserved transcriptional responses to indole signaling.

Given that many of these virulence factors facilitate host cell adhesion, such as FadA binding to E-cadherin (18), we next investigated the effects of indole derivatives on tight and adherens junction proteins in CRC cell lines. Indole derivative treatment of Caco-2 cells significantly increased the expression of genes associated with tight and adherens junctions, indicating that strengthened intercellular adhesion could contribute to the reduced invasion phenotype. *F. nucleatum* invades host epithelial and endothelial cells primarily through its adhesin FadA, which binds to E-cadherin on the cell surface (18), triggering β-catenin signaling and cytoskeletal rearrangements that facilitate bacterial internalization via a zipper-like mechanism (98,102). Strengthened cell-cell junctions have been shown to reduce *Campylobacter jejuni* invasion (103), a pathogen that also employs a zipper-like entry mechanism (104), suggesting that enhanced epithelial cohesion could similarly limit *F. nucleatum* invasion. To elucidate how indoles inhibit this process, future mechanistic studies incorporating dual RNA-seq during active invasion will be essential to define the host-pathogen molecular pathways by which indoles impede bacterial entry (105).

The differential responses to indole exposure among *F. nucleatum* isolates from CRC, Crohn's disease, oral sites, and healthy individuals underscore the importance of understanding how pathogens adapt to the disease environment. These findings suggest potential for targeted therapeutic strategies in CRC, where inhibiting biofilm formation through indole production could reduce pathogenicity (38). Moreover, the disease-specific regulation of virulence factors by indoles presents a promising avenue for targeted therapeutic intervention. Guided antimicrobial peptides (AMPs) designed to target virulence factors consistently upregulated in CRC-associated isolates following indole exposure, such as FomA, which exhibits opposite regulation in healthy versus CRC-associated isolates, may enable selective eradication of highly virulent strains, an approach currently under investigation in our laboratory (80). We expect that future research will focus on the triggers that allow pathogen adaptation and on exploiting their

disease-adapted vulnerabilities to formulate strategies that can transform how we prevent and treat microbiome-associated disease, including cancer.

ACKNOWLEDGMENTS

The authors thank the Molecular Biosciences Core (MBC) at Baylor University for access to instrumentation and expert technical assistance. An earlier pre-print version of this manuscript is available on bioRxiv (106).

DISCLOSURE STATEMENT

The authors declare that they have no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

FUNDING

Funding for this project was provided by a grant to Leigh Greathouse from the Baylor ONE-URC Grant Award.

CRediT AUTHORSHIP CONTRIBUTION STATEMENT

C.S. Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Software, Validation, Visualization, Writing - original draft, Writing - review & editing. **A.C.** Conceptualization, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – review & editing. **M.R.** Investigation, Methodology, Writing – review & editing. **J.H.** Data curation, Investigation, Writing – review & editing. **L.C.** Investigation, Writing – review & editing. **G.Z.** Investigation, Writing – review & editing. **R.L.** Conceptualization, Resources, Writing – review & editing. **L.G.** Corresponding author, Conceptualization, Funding acquisition, Methodology, Resources, Supervision, Writing – original draft, Writing – review & editing.

DATA AVAILABILITY STATEMENT

The authors confirm that the data supporting the findings of this study are available within the article and its supplementary materials. All data generated or analyzed during this study have been deposited in the Texas Data Repository (https://doi.org/10.18738/T8/3PYJA9) and are available upon request.

REFERENCES

1. Kostic AD, Gevers D, Pedamallu CS, Michaud M, Duke F, Earl AM, et al. Genomic analysis identifies association of Fusobacterium with colorectal carcinoma. Genome Res. 2012 Feb;22(2):292–8.

2. Castellarin M, Warren RL, Freeman JD, Dreolini L, Krzywinski M, Strauss J, et al. Fusobacterium nucleatum infection is prevalent in human colorectal carcinoma. Genome Res. 2012 Feb;22(2):299–306.

3. Chen Y, Huang Z, Tang Z, Huang Y, Huang M, Liu H, et al. More Than Just a Periodontal Pathogen –the Research Progress on Fusobacterium nucleatum. Front Cell Infect Microbiol [Internet]. 2022 [cited 2023 May 19];12. Available from: https://www.frontiersin.org/articles/10.3389/fcimb.2022.815318

824 4. Patini R, Staderini E, Lajolo C, Lopetuso L, Mohammed H, Rimondini L, et al. Relationship
825 between oral microbiota and periodontal disease: a systematic review. Eur Rev Med
826 Pharmacol Sci. 2018 Sep;22(18):5775–88.

827 5. Abed J, Maalouf N, Manson AL, Earl AM, Parhi L, Emgård JEM, et al. Colon Cancer-
828 Associated Fusobacterium nucleatum May Originate From the Oral Cavity and Reach Colon
829 Tumors via the Circulatory System. Front Cell Infect Microbiol [Internet]. 2020 [cited 2022
830 Nov 30];10. Available from:
831 https://www.frontiersin.org/articles/10.3389/fcimb.2020.00400

832 6. Queen J, Cing Z, Minsky H, Nandi A, Southward T, Ferri J, et al. Fusobacterium nucleatum is
833 enriched in invasive biofilms in colorectal cancer. bioRxiv. 2025 Jan 1;2024.12.30.630810.

834 7. Tomkovich S, Dejea CM, Winglee K, Drewes JL, Chung L, Housseau F, et al. Human colon
835 mucosal biofilms from healthy or colon cancer hosts are carcinogenic. J Clin Invest.
836 129(4):1699–712.

837 8. Dejea CM, Wick EC, Hechenbleikner EM, White JR, Mark Welch JL, Rossetti BJ, et al.
838 Microbiota organization is a distinct feature of proximal colorectal cancers. Proc Natl Acad
839 Sci U S A. 2014 Dec 23;111(51):18321–6.

840 9. Prudent V, Demarre G, Vazeille E, Wery M, Quenech'Du N, Ravet A, et al. The Crohn's
841 disease-related bacterial strain LF82 assembles biofilm-like communities to protect itself
842 from phagolysosomal attack. Commun Biol. 2021 May 25;4(1):627.

843 10. Jandl B, Dighe S, Baumgartner M, Makristathis A, Gasche C, Muttenthaler M.
844 Gastrointestinal Biofilms: Endoscopic Detection, Disease Relevance, and Therapeutic
845 Strategies. Gastroenterology. 2024 Nov 1;167(6):1098-1112.e5.

846 11. Swidsinski A, Weber J, Loening-Baucke V, Hale LP, Lochs H. Spatial Organization and
847 Composition of the Mucosal Flora in Patients with Inflammatory Bowel Disease. J Clin
848 Microbiol. 2005 Jul;43(7):3380–9.

849 12. Connolly JP, Kelly L. The physical biogeography of Fusobacterium nucleatum in health and
850 disease. mBio. 16(4):e02989-24.

851 13. Wang Z, Li B, Bao L, Chen Y, Yang J, Xu F, et al. Fusobacterium Nucleatum Aggravates
852 Intestinal Barrier Impairment and Colitis Through IL-8 Induced Neutrophil Chemotaxis by
853 Activating Epithelial Cells. J Inflamm Res. 2024 Nov 8;17:8407–20.

854 14. Umana A, Sanders BE, Yoo CC, Casasanta MA, Udayasuryan B, Verbridge SS, et al.
855 Reevaluating the Fusobacterium Virulence Factor Landscape [Internet]. bioRxiv; 2019 [cited
856 2024 Oct 17]. p. 534297. Available from:
857 https://www.biorxiv.org/content/10.1101/534297v1

858 15. Kaplan A, Kaplan CW, He X, McHardy I, Shi W, Lux R. Characterization of aid1, a novel gene
859 involved in Fusobacterium nucleatum interspecies interactions. Microb Ecol. 2014
860 Aug;68(2):379–87.

861 16. Zhang L, Leng XX, Qi J, Wang N, Han JX, Tao ZH, et al. The adhesin RadD enhances
862 Fusobacterium nucleatum tumour colonization and colorectal carcinogenesis. Nat
863 Microbiol. 2024 Sep;9(9):2292–307.

864 17. Abed J, Emgård JEM, Zamir G, Faroja M, Almogy G, Grenov A, et al. Fap2 Mediates
865 Fusobacterium nucleatum Colorectal Adenocarcinoma Enrichment by Binding to Tumor-
866 Expressed Gal-GalNAc. Cell Host Microbe. 2016 Aug 10;20(2):215–25.

867 18. Rubinstein MR, Wang X, Liu W, Hao Y, Cai G, Han YW. Fusobacterium nucleatum promotes
868 colorectal carcinogenesis by modulating E-cadherin/β-catenin signaling via its FadA adhesin.
869 Cell Host Microbe. 2013 Aug 14;14(2):195–206.

870 19. Ma CT, Luo HS, Gao F, Tang QC, Chen W. Fusobacterium nucleatum promotes the
871 progression of colorectal cancer by interacting with E-cadherin. Oncol Lett. 2018 Aug
872 1;16(2):2606–12.

873 20. Lima BP, Shi W, Lux R. Identification and characterization of a novel Fusobacterium
874 nucleatum adhesin involved in physical interaction and biofilm formation with
875 Streptococcus gordonii. MicrobiologyOpen. 2017 Feb 7;6(3):e00444.

876 21. Engevik MA, Danhof HA, Auchtung J, Endres BT, Ruan W, Bassères E, et al. Fusobacterium
877 nucleatum adheres to Clostridioides difficile via the RadD adhesin to enhance biofilm
878 formation in intestinal mucus. Gastroenterology. 2021 Mar;160(4):1301-1314.e8.

879 22. Xu M, Yamada M, Li M, Liu H, Chen SG, Han YW. FadA from Fusobacterium nucleatum
880 Utilizes both Secreted and Nonsecreted Forms for Functional Oligomerization for
881 Attachment and Invasion of Host Cells. J Biol Chem. 2007 Aug;282(34):25000–9.

882 23. Guo P, Tian Z, Kong X, Yang L, Shan X, Dong B, et al. FadA promotes DNA damage and
883 progression of Fusobacterium nucleatum-induced colorectal cancer through up-regulation
884 of chk2. J Exp Clin Cancer Res CR. 2020 Sep 29;39:202.

885 24. Li D, Li Z, Wang L, Zhang Y, Ning S. Oral inoculation of Fusobacterium nucleatum
886 exacerbates ulcerative colitis via the secretion of virulence adhesin FadA. Virulence.
887 15(1):2399217.

888 25. Lee JH, Wood TK, Lee J. Roles of Indole as an Interspecies and Interkingdom Signaling
889 Molecule. Trends Microbiol. 2015 Nov 1;23(11):707–18.

890 26. Li X, Zhang B, Hu Y, Zhao Y. New Insights Into Gut-Bacteria-Derived Indole and Its
891 Derivatives in Intestinal and Liver Diseases. Front Pharmacol. 2021 Dec 13;12:769501.

892 27. Gupta R, Rhee KY, Beagle SD, Chawla R, Perdomo N, Lockless SW, et al. Indole modulates
893 cooperative protein–protein interactions in the flagellar motor. PNAS Nexus. 2022 May
894 1;1(2):pgac035.

895 28. Martino PD, Fursy R, Bret L, Sundararaju B, Phillips RS. Indole can act as an extracellular
896 signal to regulate biofilm formation of Escherichia coli and other indole-producing bacteria.
897 Can J Microbiol. 2003 Jul;49(7):443–9.

898 29. Yang J, Chawla R, Rhee KY, Gupta R, Manson MD, Jayaraman A, et al. Biphasic chemotaxis of
899 Escherichia coli to the microbiota metabolite indole. Proc Natl Acad Sci. 2020 Mar
900 17;117(11):6114–20.

901 30. Hirakawa H, Kodama T, Takumi-Kobayashi A, Honda T, Yamaguchi A. Secreted indole serves
902 as a signal for expression of type III secretion system translocators in enterohaemorrhagic
903 Escherichia coli O157:H7. Microbiol Read Engl. 2009 Feb;155(Pt 2):541–50.

904 31. Hirakawa H, Inazumi Y, Masaki T, Hirata T, Yamaguchi A. Indole induces the expression of
905 multidrug exporter genes in Escherichia coli. Mol Microbiol. 2005 Feb;55(4):1113–26.

906 32. Sasaki-Imamura T, Yano A, Yoshida Y. Production of Indole from l-Tryptophan and Effects of
907 These Compounds on Biofilm Formation by Fusobacterium nucleatum ATCC 25586. Appl
908 Environ Microbiol. 2010 Jul;76(13):4260–8.

909 33. Li J, Xie L, Lin F, Ling B. Indole derivatives display antimicrobial and antibiofilm effects
910 against extensively drug-resistant Acinetobacter baumannii. Microbiol Spectr. 2025 Apr
911 15;13(5):e03388-24.

912 34. Lee JH, Lee J. Indole as an intercellular signal in microbial communities. FEMS Microbiol Rev.
913 2010 Jul 1;34(4):426–44.

914 35. Li G, Young KD. Indole production by the tryptophanase TnaA in Escherichia coli is
915 determined by the amount of exogenous tryptophan. Microbiol Read Engl. 2013 Feb;159(Pt
916 2):402–10.

917 36. Sasaki-Imamura T, Yoshida Y, Suwabe K, Yoshimura F, Kato H. Molecular basis of indole
918 production catalyzed by tryptophanase in the genus Prevotella. FEMS Microbiol Lett. 2011
919 Sep 1;322(1):51–9.

920 37. Roager HM, Licht TR. Microbial tryptophan catabolites in health and disease. Nat Commun.
921 2018 Aug 17;9(1):3294.

922 38. Sinha AK, Laursen MF, Brinck JE, Rybtke ML, Hjørne AP, Procházková N, et al. Dietary fibre
923 directs microbial tryptophan metabolism via metabolic interactions in the gut microbiota.
924 Nat Microbiol. 2024 Aug;9(8):1964–78.

925 39. Ye X, Li H, Anjum K, Zhong X, Miao S, Zheng G, et al. Dual Role of Indoles Derived From
926 Intestinal Microbiota on Human Health. Front Immunol. 2022 Jun 17;13:903526.

927 40. Dodd D, Spitzer MH, Van Treuren W, Merrill BD, Hryckowian AJ, Higginbottom SK, et al. A
928 gut bacterial pathway metabolizes aromatic amino acids into nine circulating metabolites.
929 Nature. 2017 Nov 30;551(7682):648–52.

930 41. Wlodarska M, Luo C, Kolde R, d'Hennezel E, Annand JW, Heim CE, et al. Indoleacrylic Acid
931 Produced by Commensal Peptostreptococcus Species Suppresses Inflammation. Cell Host
932 Microbe. 2017 Jul 12;22(1):25-37.e6.

933 42. Pacheco AR, Sperandio V. Enteric Pathogens Exploit the Microbiota-generated Nutritional
934 Environment of the Gut. Microbiol Spectr. 2015 Jun;3(3).

935 43. Kumar A, Sperandio V. Indole Signaling at the Host-Microbiota-Pathogen Interface. mBio.
936 10(3):e01031-19.

937 44. Kohli N, Crisp Z, Riordan R, Li M, Alaniz RC, Jayaraman A. The microbiota metabolite indole
938 inhibits Salmonella virulence: Involvement of the PhoPQ two-component system. PloS One.
939 2018;13(1):e0190613.

940 45. Howard MF, Bina XR, Bina JE. Indole Inhibits ToxR Regulon Expression in Vibrio cholerae.
941 Infect Immun. 2019 Mar;87(3):e00776-18.

942 46. Tsang N, Macnab R, Koshland DE. Common mechanism for repellents and attractants in
943 bacterial chemotaxis. Science. 1973 Jul 6;181(4094):60–3.

944 47. Tso WW, Adler J. Negative chemotaxis in Escherichia coli. J Bacteriol. 1974 May
945 1;118(2):560–76.

946 48. Muskavitch MA, Kort EN, Springer MS, Goy MF, Adler J. Attraction by repellents: an error in
947 sensory information processing by bacterial mutants. Science. 1978 Jul 7;201(4350):63–5.

948 49. Mao H, Cremer PS, Manson MD. A sensitive, versatile microfluidic assay for bacterial
949 chemotaxis. Proc Natl Acad Sci U S A. 2003 Apr 29;100(9):5449–54.

950 50. Wu W, Jing J, Chen Z, Guo X, Qu Y. Indole signaling enhances biofilm formation and quorum
951 sensing in sequencing biofilm batch reactors. J Environ Chem Eng. 2024 Jun 1;12(3):112494.

952 51. Zhou P, G. C. B, Stolte F, Wu C. Use of CRISPR interference for efficient and rapid gene
953 inactivation in Fusobacterium nucleatum. Appl Environ Microbiol. 2024 Jan 8;90(2):e01665-
954 23.

955 52. Lawson PA, Gharbia SE, Shah HN, Clark DR, Collins MD. Intrageneric relationships of
956 members of the genus Fusobacterium as determined by reverse transcriptase sequencing
957 of small-subunit rRNA. Int J Syst Bacteriol. 1991 Jul;41(3):347–54.

958 53. Gharbia SE, Shah HN. Fusobacterium nucleatum subsp. fusiforme subsp. nov. and
959 Fusobacterium nucleatum subsp. animalis subsp. nov. as additional subspecies within
960 Fusobacterium nucleatum. Int J Syst Bacteriol. 1992 Apr;42(2):296–8.

961 54. Karpathy SE, Qin X, Gioia J, Jiang H, Liu Y, Petrosino JF, et al. Genome sequence of
962 Fusobacterium nucleatum subspecies polymorphum - a genetically tractable fusobacterium.
963 PloS One. 2007 Aug 1;2(7):e659.

964 55. Kook JK, Park SN, Lim YK, Choi MH, Cho E, Kong SW, et al. Fusobacterium nucleatum subsp.
965 fusiforme Gharbia and Shah 1992 is a Later Synonym of Fusobacterium nucleatum subsp.
966 vincentii Dzink et al. 1990. Curr Microbiol. 2013 Apr 1;66(4):414–7.

967 56. Krieger M, Guo M, Merritt J. Reexamining the role of Fusobacterium nucleatum subspecies
968 in clinical and experimental studies. Gut Microbes. 16(1):2415490.

969 57. Krieger M, AbdelRahman YM, Choi D, Palmer EA, Yoo A, McGuire S, et al. Stratification of
970 Fusobacterium nucleatum by local health status in the oral cavity defines its subspecies
971 disease association. Cell Host Microbe. 2024 Apr 10;32(4):479-488.e4.

972 58. Kook JK, Park SN, Lim YK, Cho E, Jo E, Roh H, et al. Genome-Based Reclassification of
973 Fusobacterium nucleatum Subspecies at the Species Level. Curr Microbiol. 2017
974 Oct;74(10):1137–47.

975 59. Zepeda-Rivera M, Minot SS, Bouzek H, Wu H, Blanco-Míguez A, Manghi P, et al. A distinct
976 Fusobacterium nucleatum clade dominates the colorectal cancer niche. Nature. 2024
977 Apr;628(8007):424–32.

978 60. Ye X, Wang R, Bhattacharya R, Boulbes DR, Fan F, Xia L, et al. Fusobacterium Nucleatum
979 Subspecies Animalis Influences Proinflammatory Cytokine Expression and Monocyte
980 Activation in Human Colorectal Tumors. Cancer Prev Res (Phila Pa). 2017 Jul 4;10(7):398–
981 409.

982 61. Sivertsen A, Forni D, Molteni C, Bivand J, Dimmen G, Bruvold TS, et al. Reassessing
983 taxonomy and virulence in the Fusobacterium nucleatum group—rebuttal of Fusobacterium
984 animalis clades “Fna C1” and “Fna C2,” genome announcement for Fusobacterium
985 watanabei, and description of Fusobacterium paranimalis sp. nov. mBio. 2025 Jul
986 31;16(9):e00941-25.

987 62. Cochrane K, Robinson AV, Holt RA, Allen-Vercoe E. A survey of Fusobacterium nucleatum
988 genes modulated by host cell infection. Microb Genomics. 2020 Feb;6(2):e000300.

989 63. Zhao T, Chen J, Liu S, Yang J, Wu J, Miao L, et al. Transcriptome analysis of *Fusobacterium*
990 *nucleatum* reveals differential gene expression patterns in the biofilm versus planktonic
991 cells. Biochem Biophys Res Commun. 2022 Feb 19;593:151–7.

992 64. Feng Q, Liang S, Jia H, Stadlmayr A, Tang L, Lan Z, et al. Gut microbiome development along
993 the colorectal adenoma–carcinoma sequence. Nat Commun. 2015 Mar 11;6(1):6528.

994 65. Zeller G, Tap J, Voigt AY, Sunagawa S, Kultima JR, Costea PI, et al. Potential of fecal
995 microbiota for early-stage detection of colorectal cancer. Mol Syst Biol. 2014 Nov
996 28;10(11):766.

997 66. Vogtmann E, Hua X, Zeller G, Sunagawa S, Voigt AY, Hercog R, et al. Colorectal Cancer and
998 the Human Gut Microbiome: Reproducibility with Whole-Genome Shotgun Sequencing.
999 PLoS ONE. 2016 May 12;11(5):e0155362.

1000 67. Sun XZ, Zhao DY, Zhou YC, Wang QQ, Qin G, Yao SK. Alteration of fecal tryptophan
1001 metabolism correlates with shifted microbiota and may be involved in pathogenesis of
1002 colorectal cancer. World J Gastroenterol. 2020 Dec 7;26(45):7173–90.

1003 68. Darkoh C, Chappell C, Gonzales C, Okhuysen P. A rapid and specific method for the
1004 detection of indole in complex biological samples. Appl Environ Microbiol. 2015
1005 Dec;81(23):8093–7.

1006 69. Tuomola M, Vahva M, Kallio H. High-Performance Liquid Chromatography Determination of
1007 Skatole and Indole Levels in Pig Serum, Subcutaneous Fat, and Submaxillary Salivary Glands.
1008 J Agric Food Chem. 1996 Jan 1;44(5):1265–70.

1009 70. Kema IP, Meijer WG, Meiborg G, Ooms B, Willemse PHB, De Vries EGE. Profiling of
1010 Tryptophan-related Plasma Indoles in Patients with Carcinoid Tumors by Automated, On-
1011 Line, Solid-Phase Extraction and HPLC with Fluorescence Detection. Clin Chem. 2001 Oct
1012 1;47(10):1811–20.

1013 71. Cochrane KLS. Elucidating potential virulence factors in Fusobacterium nucleatum. 2016;
1014 Available from: http://hdl.handle.net/10214/9623

1015 72. Robinson AV, Vancuren SJ, Marcone M, Allen-Vercoe E. Characterization of diet-linked
1016 amino acid pool influence on Fusobacterium spp. growth and metabolism. mSphere. 2025
1017 Feb 13;10(3):e00789-24.

1018 73. Kasper SH, Morell-Perez C, Wyche TP, Sana TR, Lieberman LA, Hett EC. Colorectal cancer-
1019 associated anaerobic bacteria proliferate in tumor spheroids and alter the
1020 microenvironment. Sci Rep. 2020 Mar 24;10(1):5321.

1021 74. Vinogradov E, St Michael F, Cairns C, Cox AD. The structure of the LPS O-chain from five
1022 *Fusobacterium nucleatum* strains CTX47T, CC2_6JVN3, CC2_3FMU1, CC2_1JVN3, HM-996,
1023 containing alditol and phosphate in the main chain and development of mouse monoclonal
1024 antibodies specific to the O-antigens. Carbohydr Res. 2022 Nov 1;521:108648.

1025 75. Dharmani P, Strauss J, Ambrose C, Allen-Vercoe E, Chadee K. Fusobacterium nucleatum
1026 Infection of Colonic Cells Stimulates MUC2 Mucin and Tumor Necrosis Factor Alpha. Infect
1027 Immun. 2011 Jul;79(7):2597–607.

1028 76. Strauss J, Kaplan GG, Beck PL, Rioux K, Panaccione R, DeVinney R, et al. Invasive potential of
1029 gut mucosa-derived fusobacterium nucleatum positively correlates with IBD status of the
1030 host. Inflamm Bowel Dis. 2011 Sep 1;17(9):1971–8.

1031 77. Strauss J, White A, Ambrose C, McDonald J, Allen-Vercoe E. Phenotypic and genotypic
1032 analyses of clinical *Fusobacterium nucleatum* and *Fusobacterium periodonticum* isolates
1033 from the human gut. Anaerobe. 2008 Dec 1;14(6):301–9.

1034 78. Cochrane K, McGuire AM, Priest ME, Abouelleil A, Cerqueira GC, Lo R, et al. Complete
1035 Genome Sequences and Analysis of the Fusobacterium nucleatum subspecies animalis 7-1
1036 Bacteriophage Funu1 and Funu2. Anaerobe. 2016 Apr;38:125.

1037 79. Tomkovich S, Yang Y, Winglee K, Gauthier J, Mühlbauer M, Sun X, et al. Locoregional effects
1038 of microbiota in a preclinical model of colon carcinogenesis. Cancer Res. 2017 May
1039 15;77(10):2620–32.

1040 80. Choudhury A, Scano C, Barton A, Greathouse KL. Targeted therapy of oncomicrobe F.
1041 nucleatum with bioengineered probiotic expressing guided antimicrobial peptide (gAMP)
1042 [Internet]. bioRxiv; 2024 [cited 2025 Apr 16]. p. 2024.10.29.620919. Available from:
1043 https://www.biorxiv.org/content/10.1101/2024.10.29.620919v1

1044 81. Huang L, Ge S, Yang K, Duan L, Gao L, Li YZ, et al. Effects of oral gavage with periodontal
1045 pathogens and plaque biofilm on gut microbiota ecology and intestinal tissue architecture
1046 in mice: a mechanistic study. Front Cell Infect Microbiol. 2025 Aug 8;15:1589055.

1047 82. Wang Q, Hu T, Zhang Q, Zhang Y, Dong X, Jin Y, et al. *Fusobacterium nucleatum* promotes
1048 colorectal cancer through neogenesis of tumor stem cells. J Clin Invest [Internet]. 2025 Feb
1049 3 [cited 2025 Oct 4];135(3). Available from:
1050 https://www.jci.org/articles/view/181595?utm_source=chatgpt.com

1051 83. Karta J, Meyers M, Rodriguez F, Koncina E, Gilson C, Klein E, et al. Fusobacterium nucleatum
1052 interacts with cancer-associated fibroblasts to promote colorectal cancer. EMBO J. 2025
1053 Aug 22;44(19):5375–93.

1054 84. Bansal T, Alaniz RC, Wood TK, Jayaraman A. The bacterial signal indole increases epithelial-
1055 cell tight-junction resistance and attenuates indicators of inflammation. Proc Natl Acad Sci.
1056 2010 Jan 5;107(1):228–33.

1057 85. Bender MJ, McPherson AC, Phelps CM, Pandey SP, Laughlin CR, Shapira JH, et al. Dietary
1058 tryptophan metabolite released by intratumoral Lactobacillus reuteri facilitates immune
1059 checkpoint inhibitor treatment. Cell. 2023 Apr 27;186(9):1846-1862.e26.

1060 86. Hezaveh K, Shinde RS, Klötgen A, Halaby MJ, Lamorte S, Ciudad MT, et al. Tryptophan-
1061 derived microbial metabolites activate the aryl hydrocarbon receptor in tumor-associated
1062 macrophages to suppress anti-tumor immunity. Immunity. 2022 Feb 8;55(2):324-340.e8.

1063 87. Li M, Ding Y, Wei J, Dong Y, Wang J, Dai X, et al. Gut microbiota metabolite indole-3-acetic
1064 acid maintains intestinal epithelial homeostasis through mucin sulfation. Gut Microbes.
1065 16(1):2377576.

1066 88. Dai Z, Deng KL, Wang XM, Yang DX, Tang CL, Zhou YP. Bidirectional effects of the tryptophan
1067 metabolite indole-3-acetaldehyde on colorectal cancer. World J Gastrointest Oncol. 2024
1068 Jun 15;16(6):2697–715.

1069 89. Gaimster H, Summers D. Regulation of Indole Signalling during the Transition of E. coli from
1070 Exponential to Stationary Phase. PLoS ONE. 2015 Sep 2;10(9):e0136691.

1071 90. Chant EL, Summers DK. Indole signalling contributes to the stable maintenance of
1072 Escherichia coli multicopy plasmids. Mol Microbiol. 2007;63(1):35–43.

1073 91. Gaimster H, Cama J, Hernández-Ainsa S, Keyser UF, Summers DK. The Indole Pulse: A New
1074 Perspective on Indole Signalling in Escherichia coli. PLoS ONE. 2014 Apr 2;9(4):e93168.

1075 92. Muchova M, Balacco DL, Grant MM, Chapple ILC, Kuehne SA, Hirschfeld J. Fusobacterium
1076 nucleatum Subspecies Differ in Biofilm Forming Ability in vitro. Front Oral Health [Internet].
1077 2022 [cited 2022 May 30];3. Available from:
1078 https://www.frontiersin.org/article/10.3389/froh.2022.853618

1079 93. Queen J, Domingue JC, White JR, Stevens C, Udayasuryan B, Nguyen TTD, et al. Comparative
1080 Analysis of Colon Cancer-Derived Fusobacterium nucleatum Subspecies: Inflammation and
1081 Colon Tumorigenesis in Murine Models. mBio. 2022 Feb 8;13(1):e02991-21.

1082 94. Hashidoko Y, Kim D. Bidirectional Cell-Cell Communication via Indole and Cyclo(Pro-Tyr)
1083 Modulates Interspecies Biofilm Formation. Appl Environ Microbiol. 87(22):e01277-21.

1084 95. Casasanta MA, Yoo CC, Udayasuryan B, Sanders BE, Umaña A, Zhang Y, et al. Fusobacterium
1085 nucleatum host-cell binding and invasion induces IL-8 and CXCL1 secretion that drives
1086 colorectal cancer cell migration. Sci Signal. 2020 Jul 21;13(641):eaba9157.

1087 96. Fusobacterium nucleatum promotes tumor progression in KRAS p.G12D-mutant colorectal
1088 cancer by binding to DHX15 | Nature Communications [Internet]. [cited 2025 Oct 4].
1089 Available from: https://www.nature.com/articles/s41467-024-45572-
1090 w?utm_source=chatgpt.com

1091 97. Bullman S, Pedamallu CS, Sicinska E, Clancy TE, Zhang X, Cai D, et al. Analysis of
1092 Fusobacterium persistence and antibiotic response in colorectal cancer. Science. 2017 Dec
1093 15;358(6369):1443–8.

1094 98. Fardini Y, Wang X, Témoin S, Nithianantham S, Lee D, Shoham M, et al. Fusobacterium
1095 nucleatum adhesin FadA binds vascular-endothelial cadherin and alters endothelial
1096 integrity. Mol Microbiol. 2011 Dec;82(6):1468–80.

1097 99. Yu M, Wang Q, Ma Y, Li L, Yu K, Zhang Z, et al. Aryl Hydrocarbon Receptor Activation
1098 Modulates Intestinal Epithelial Barrier Function by Maintaining Tight Junction Integrity. Int J
1099 Biol Sci. 2018;14(1):69–77.

1100 100. Wikoff WR, Anfora AT, Liu J, Schultz PG, Lesley SA, Peters EC, et al. Metabolomics analysis
1101 reveals large effects of gut microflora on mammalian blood metabolites. Proc Natl Acad Sci
1102 U S A. 2009 Mar 10;106(10):3698–703.

1103 101. Botsford JL, Demoss RD. Escherichia coli Tryptophanase in the Enteric Environment. J
1104 Bacteriol [Internet]. 1972 Jan [cited 2025 Oct 5]; Available from:
1105 https://journals.asm.org/doi/10.1128/jb.109.1.74-80.1972

1106 102. Han YW, Shi W, Huang GTJ, Kinder Haake S, Park NH, Kuramitsu H, et al. Interactions
1107 between Periodontal Bacteria and Human Oral Epithelial Cells: Fusobacterium nucleatum
1108 Adheres to and Invades Epithelial Cells. Infect Immun. 2000 Jun;68(6):3140–6.

1109 103. Hatayama S, Shimohata T, Amano S, Kido J, Nguyen AQ, Sato Y, et al. Cellular Tight
1110 Junctions Prevent Effective Campylobacter jejuni Invasion and Inflammatory Barrier
1111 Disruption Promoting Bacterial Invasion from Lateral Membrane in Polarized Intestinal
1112 Epithelial Cells. Front Cell Infect Microbiol [Internet]. 2018 Jan 30 [cited 2025 Oct 5];8.
1113 Available from: https://www.frontiersin.org/journals/cellular-and-infection-
1114 microbiology/articles/10.3389/fcimb.2018.00015/full

1115 104. Cróinín TÓ, Backert S. Host Epithelial Cell Invasion by Campylobacter jejuni: Trigger or
1116 Zipper Mechanism? Front Cell Infect Microbiol [Internet]. 2012 [cited 2023 Sep 26];2.
1117 Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3417527/

1118 105. Westermann AJ, Vogel J. Host-Pathogen Transcriptomics by Dual RNA-Seq. Methods Mol
1119 Biol Clifton NJ. 2018;1737:59–75.

1120 106. Scano C, Choudhury A, Rojo MG, Lavado R, Zaharas G, Hawkins J, et al. Comparative
1121 Analysis of Indole Derivative Production Among Clinical Isolates of Fusobacterium
1122 nucleatum [Internet]. bioRxiv; 2025 [cited 2025 Jul 17]. p. 2025.05.16.654558. Available
1123 from: https://www.biorxiv.org/content/10.1101/2025.05.16.654558v3

1124

Table 1: Sources and References of Clinical *F. nucleatum* Isolates

STRAIN	ALT NAMES BASED ON PMID #	SAMPLING SITE	DISEASE	SUBSPECIES	ISOLATE SUPPLIERS	Paper Reference
F. nucleatum animalis EAVG_003	11_3_2	Nonspecific colon biopsy	**Healthy**	*animalis*	Dr. Emma Allen-Vercoe, University of Guelph	(71–74)
F. nucleatum EAVG_014	3_1_27	Rectum biopsy	**Healthy**	*vincentii*	Dr. Emma Allen-Vercoe, University of Guelph	*(71,75)*
F. nucleatum EAVG_012	4_1_13	Ascending colon biopsy	**Healthy**	*vincentii*	Dr. Emma Allen-Vercoe, University of Guelph	*(71,75,76)*
F. nucleatum EAVG_027	2_1_50A	Distal colon biopsy	**Crohn's remission**	*animalis*	Dr. Emma Allen-Vercoe, University of Guelph	*(71,76)*
F. nucleatum subsp. vincentii	1_1_2	Distal colon biopsy	**Crohn's remission**	*vincentii*	Dr. Emma Allen-Vercoe, University of Guelph	
F. nucleatum EAVG_005	13_3C	Sigmoid colon biopsy	**Crohn's remission**	*nucleatum*	Dr. Emma Allen-Vercoe, University of Guelph	(71,75,77,78)
F. nucleatum EAVG_029	4_1_31B	Ascending colon biopsy	**Crohn's active**	*vincentii*	Dr. Emma Allen-Vercoe, University of Guelph	
F. nucleatum EAVG_030	3_1_36B3	Ascending colon biopsy	**Crohn's active**	*vincentii*	Dr. Emma Allen-Vercoe, University of Guelph	
F. nucleatum subsp. nucleatum CC2_3FMU1		Tumor biopsy	**Colorectal cancer**	*nucleatum*	Dr. Emma Allen-Vercoe, University of Guelph	(2,71,74,79)
F. nucleatum subsp. animalis CC2_6JVN3		Tumor biopsy	**Colorectal cancer**	*animalis*	Dr. Emma Allen-Vercoe, University of Guelph	(71–74)
F. nucleatum subsp. vincentii CC53		Tumor biopsy	**Colorectal cancer**	*vincentii*	Dr. Emma Allen-Vercoe, University of Guelph	*(2,5,71,78)*
F. nucleatum subsp. animalis CC7_3JVN3C1		Tumor biopsy	**Colorectal cancer**	*animalis*	Dr. Emma Allen-Vercoe, University of Guelph	*(71,79)*
F. nucleatum (SB-CTX47T)		Tumor biopsy	**Colorectal cancer**	*polymorphum*	Dr. Susan Bullman, MD Anderson Cancer Center	*(74)*
F. nucleatum (SB-CTX3Tcol3)		Tumor biopsy	**Colorectal cancer**	*nucleatum*	Dr. Susan Bullman, MD Anderson Cancer Center	*(80)*
F. nucleatum EAVG002-7_1-HM_73	7_1	Ascending colon biopsy	**Crohn's active**	*animalis*	Dr. Emma Allen-Vercoe, University of Guelph	*(71,77,78)*
F. nucleatum subsp. nucleatum Knorr 25586		ATCC	**Oral Lesion**	*nucleatum*	ATCC	(24,81–83)

Table 1. Sources and references of clinical *Fusobacterium nucleatum* isolates.
A summary of clinical *F. nucleatum* isolates obtained from various disease contexts, including healthy controls, Crohn's disease (remission and active), colorectal cancer, and oral lesions. Isolates are categorized by strain, alternative names (based on PubMed identifiers), sampling site, associated disease state, and subspecies classification. Subspecies was determined with subspecies specific primers (Supplemental Table 1). Supplier information and relevant literature references are also provided.

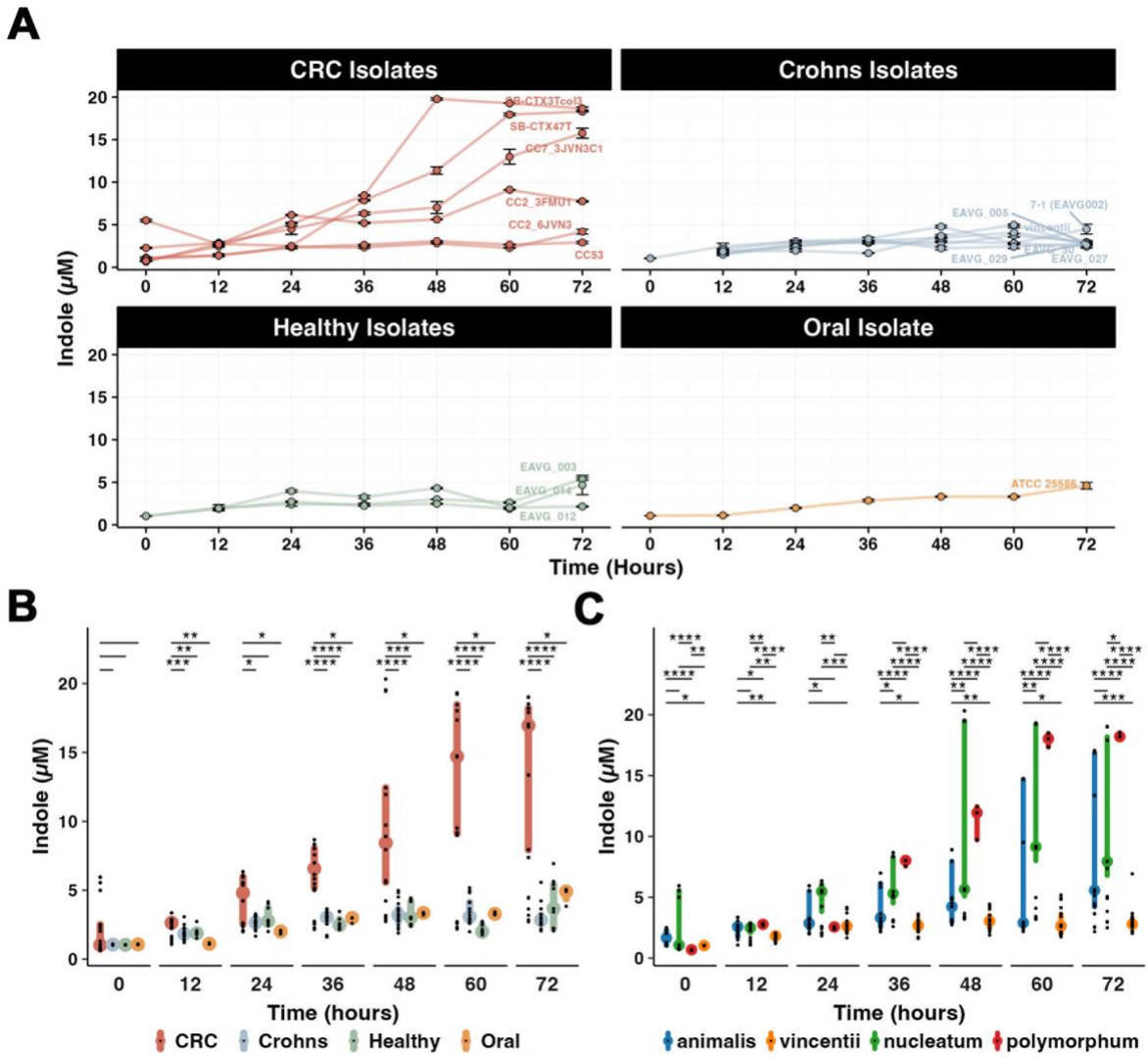


Figure 1: Elevated indole production is a distinctive feature of colorectal cancer-associated *F. nucleatum* isolates. (A) Indole production by each *F. nucleatum* isolate over time, quantified using Kovac's reagent and calibrated against an indole standard curve. Kovac's reagent (p-dimethylaminobenzaldehyde in acidic solution) reacts with the indole ring at position 3 to form a conjugated imine, producing a pink/red color proportional to indole concentration. Data are shown as mean indole concentration (μM) ± SEM for each isolate **(B, C)** Indole production of *F. nucleatum* isolates over time, grouped by disease association **(B)** and subspecies **(C)**. Differences in indole production between CRC-associated and non-CRC isolates **(B)** and subspecies **(C)** were tested by adjusted ANOVA (Benjamini–Hochberg correction), followed by Tukey's HSD for pairwise comparisons. Data are presented as median with interquartile range (IQR). Significance is indicated as: $p < 0.05$; **p** < 0.01; ***p*** < 0.001; **p** < 0.0001. Isolate groups: CRC (n=6), Crohn's (n=6), healthy (n=3), oral lesion (n=1). Subspecies groups: *animalis* (5), *vincentii* (6), *nucleatum* (4), *polymorphum* (1).

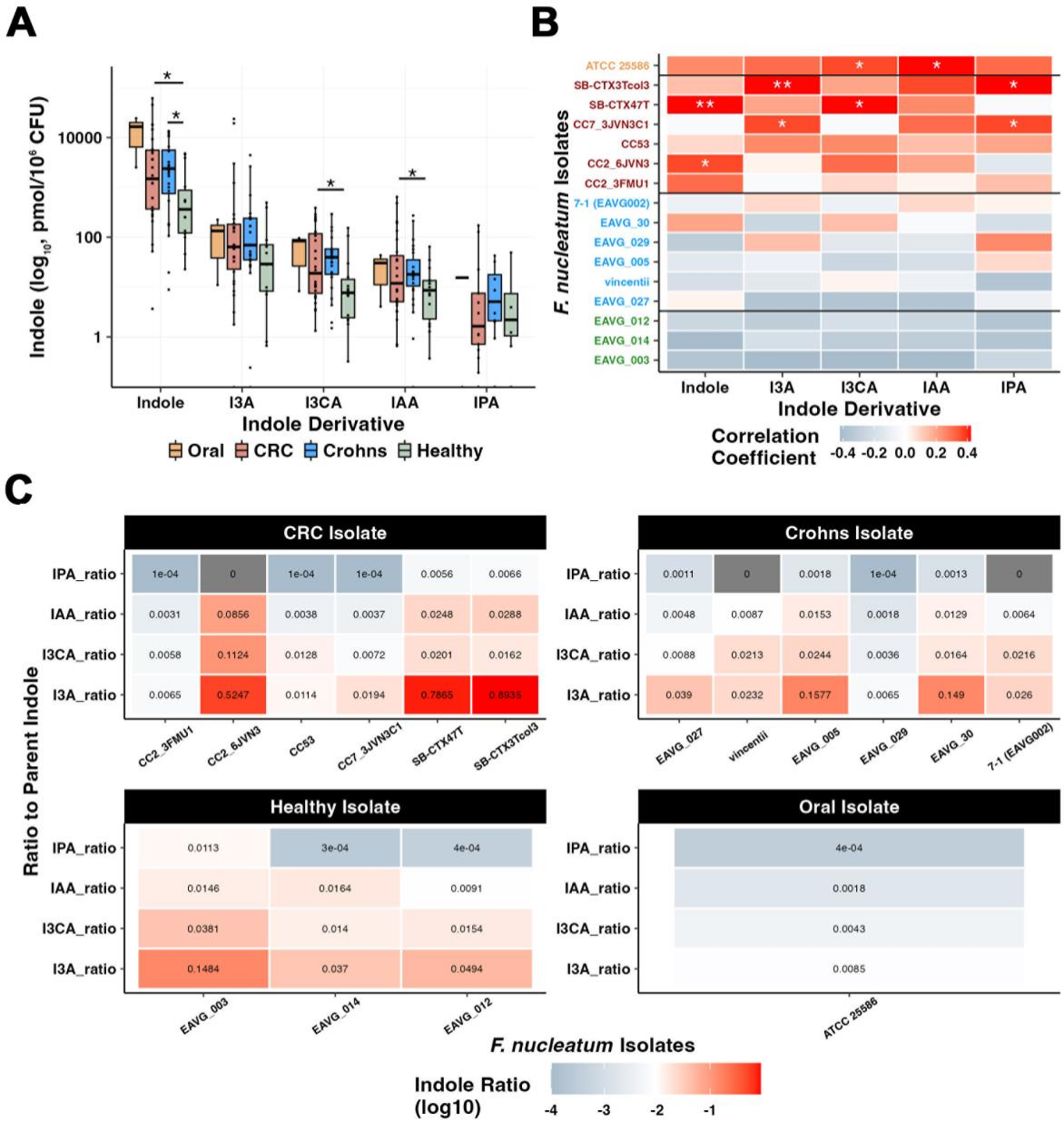


Figure 2: Disease-associated isolates exhibit elevated production of specific indole derivatives and have distinct indole production profiles. (A) Quantification of indole and indole derivatives in Columbia Broth spent media from *F. nucleatum* clinical isolates by HPLC. Box plots illustrate the production of five indole derivatives - indole, indole-3-aldehyde (I3A), indole-3-carboxylic acid (I3CA), indole-3-acetic acid (IAA), and indole-3-propionic acid (IPA) - measured in picomoles (pmol) per million colony-forming units (CFU). P-values indicate statistically significant differences between groups, determined by one-way ANOVA. **(B)** Point-biserial correlation analysis was conducted using *F. nucleatum* isolates as the categorical variable and indole derivative production as the continuous variable. The resulting heatmap displays Pearson correlation coefficients (ρ), illustrating the associations between individual *F. nucleatum* isolates and the levels of

specific indole derivatives. The y-axis labels are color-coded based on the disease association of each isolate: green for healthy, blue for Crohn's disease, red for CRC, and orange for oral isolate. Positive correlations are indicated in shades of red, while negative correlations are shown in shades of blue. Asterisks (*) denote statistically significant correlations (* $p < 0.05$; ** $p < 0.01$). **(C)** To determine indole derivative production preference by isolates, ratios of indole derivatives to the parent indole were generated for all *F. nucleatum* isolates. Each cell shows the ratio value, with colors ranging from light blue (low ratios) to red (high ratios). Isolates grouped by disease association (CRC: n=6, Crohn's: n=6, healthy: n=3, oral lesion: n=1).

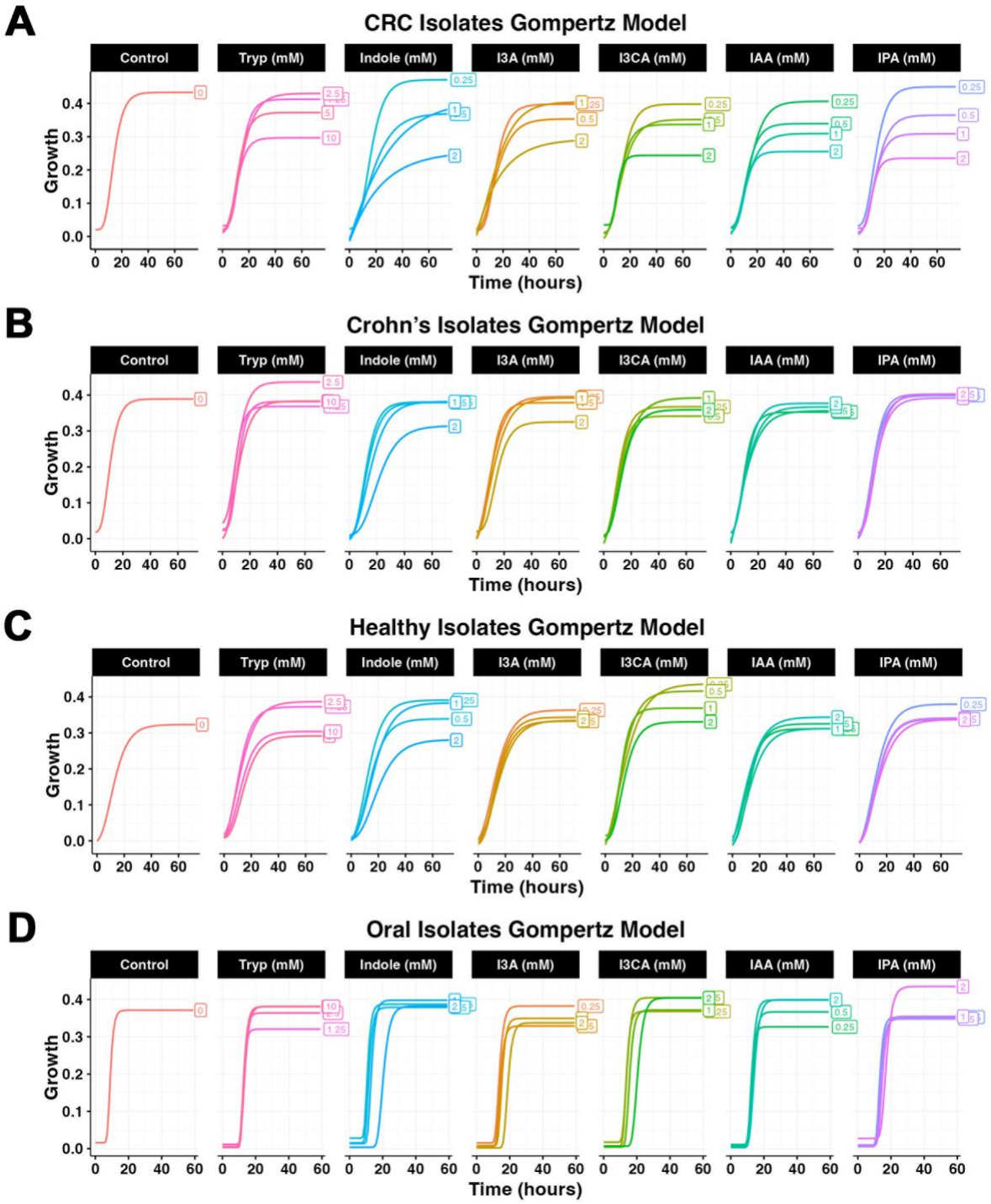


Figure 3 Exposure to indole derivatives alters *F. nucleatum* growth kinetics in a dose-dependent manner without bactericidal effects. Gompertz growth models comparing bacterial growth responses to indole derivatives across CRC **(A)**, Crohn's **(B)**, Healthy **(C)**, and Oral isolates **(D)**. Growth (OD) versus time curves show responses to Indole, I3A, I3CA, IAA, and IPA (all in mM) relative to no treatment. Chloramphenicol (µg/mL) is used as negative controls to display strong inhibition of growth. Tryptophan

was included in these growth experiments to determine whether the bacteria's ability to convert tryptophan to indole derivatives impacts their growth. Growth was monitored for 72 hours, with concentrations of treatments indicated on each curve. Gompertz model fits generated key growth parameters including inflection time, growth rate, upper asymptote (OD max), initial value, R-squared, and maximum growth rate to quantitatively compare derivatives effects across isolates (Table 3-6). Isolates grouped by disease association (CRC: n=6, Crohn's: n=6, healthy: n=3, oral lesion: n=1).

Table 2: CRC Gompertz Model Parameters							
	Concentration	Inflection Time	Growth Rate	OD MAX	Initial Value	R squared	Max Growth Rate
Control	0	11.96	0.20	0.43	0.02	0.35	0.03
Tryp (mM)	1.25	11.86	0.20	0.41	0.03	0.32	0.03
	2.5	11.03	0.13	0.43	0.00	0.39	0.02
	5	10.04	0.19	0.37	0.02	0.27	0.02
	10	9.35	0.21	0.30	0.02	0.23	0.02
Indole (mM)	0.25	13.62	0.15	0.47	0.02	0.40	0.02
	0.5	11.84	0.09	0.37	-0.03	0.28	0.01
	1	-21.20	0.04	0.41	-0.78	0.23	0.02
	2	-74.89	0.04	0.26	-6.13	0.14	0.10
I3A (mM)	0.25	12.49	0.16	0.40	0.02	0.30	0.02
	0.5	11.49	0.12	0.35	0.01	0.23	0.02
	1	9.46	0.09	0.40	-0.02	0.25	0.01
	2	-92.97	0.05	0.29	-47.33	0.14	0.96
I3CA (mM)	0.25	11.74	0.17	0.40	0.01	0.29	0.02
	0.5	10.28	0.13	0.35	-0.01	0.25	0.02
	1	10.29	0.18	0.34	0.01	0.28	0.02
	2	9.16	0.34	0.24	0.03	0.12	0.03
IAA (mM)	0.25	11.73	0.14	0.41	0.01	0.32	0.02
	0.5	10.13	0.17	0.34	0.03	0.24	0.02
	1	9.02	0.14	0.31	0.00	0.22	0.02
	2	9.37	0.22	0.26	0.02	0.15	0.02
IPA (mM)	0.25	11.18	0.16	0.40	0.03	0.26	0.02
	0.5	11.24	0.15	0.36	0.00	0.25	0.02
	1	10.45	0.18	0.31	0.01	0.18	0.02
	2	9.80	0.28	0.24	0.02	0.14	0.02

Table 2: Growth kinetics of CRC-associated *F. nucleatum* in response to 125 tryptophan metabolites modeled by the Gompertz equation. Gompertz model parameters—including inflection time, growth rate, OD max, initial value, R^2, and maximum growth rate—for *F. nucleatum* cultures treated with various concentrations of tryptophan (Trp) and its metabolites: indole, indole-3-aldehyde (I3A), indole-3-carboxylic acid (I3CA), indole-3-acetic acid (IAA), and indole-3-propionic acid (IPA).

Table 3: Crohn's Gompertz Model Parameters							
	Concentration	Inflection Time	Growth Rate	OD MAX	Initial Value	R squared	Max Growth Rate
Control	0	8.96	0.22	0.39	0.02	0.33	0.03
Tryp (mM)	1.25	7.92	0.27	0.37	0.02	0.28	0.03
	2.5	8.55	0.19	0.44	0.04	0.31	0.03
	5	9.06	0.16	0.38	-0.01	0.37	0.02
	10	9.80	0.23	0.38	0.03	0.33	0.03
Indole (mM)	0.25	9.82	0.17	0.38	0.00	0.34	0.02
	0.5	10.29	0.15	0.38	0.00	0.37	0.02
	1	11.72	0.12	0.38	-0.01	0.42	0.02
	2	17.44	0.10	0.31	0.01	0.47	0.01
I3A (mM)	0.25	7.67	0.14	0.40	-0.02	0.31	0.02
	0.5	9.10	0.20	0.38	0.02	0.29	0.03
	1	8.91	0.12	0.39	-0.02	0.36	0.02
	2	12.28	0.16	0.33	0.02	0.41	0.02
I3CA (mM)	0.25	8.74	0.18	0.37	0.00	0.31	0.02
	0.5	9.65	0.21	0.34	0.01	0.29	0.03
	1	8.78	0.12	0.39	-0.04	0.35	0.02
	2	11.03	0.15	0.36	0.00	0.37	0.02
IAA (mM)	0.25	8.22	0.21	0.35	0.01	0.31	0.03
	0.5	5.53	0.11	0.36	-0.08	0.30	0.02
	1	7.56	0.13	0.37	-0.03	0.30	0.02
	2	9.17	0.17	0.38	0.01	0.42	0.02
IPA (mM)	0.25	8.79	0.16	0.40	0.00	0.35	0.02
	0.5	9.30	0.19	0.40	0.01	0.36	0.03
	1	9.78	0.15	0.39	-0.01	0.34	0.02
	2	11.07	0.16	0.40	0.02	0.41	0.02

Table 3: Growth kinetics of Crohn's-associated *F. nucleatum* in response to tryptophan metabolites modeled by the Gompertz equation. Gompertz model parameters—including inflection time, growth rate, OD max, initial value, R^2, and maximum growth rate—for *F. nucleatum* cultures treated with various concentrations of tryptophan (Trp) and its metabolites: indole, indole-3-aldehyde (I3A), indole-3-carboxylic acid (I3CA), indole-3-acetic acid (IAA), and indole-3-propionic acid (IPA).

Table 4: Healthy Gompertz Model Parameters							
	Concentration	Inflection Time	Growth Rate	OD MAX	Initial Value	R squared	Max Growth Rate
Control	0	9.90	0.14	0.32	-0.01	0.26	0.02
Tryp (mM)	1.25	8.90	0.16	0.37	0.01	0.26	0.02
	2.5	8.84	0.12	0.39	-0.01	0.34	0.02
	5	12.09	0.14	0.29	0.01	0.26	0.01
	10	10.39	0.14	0.30	0.01	0.20	0.02
Indole (mM)	0.25	10.02	0.13	0.39	-0.01	0.24	0.02
	0.5	11.24	0.13	0.34	0.00	0.28	0.02
	1	13.43	0.11	0.38	0.00	0.42	0.02
	2	15.77	0.10	0.28	0.00	0.25	0.01
I3A (mM)	0.25	9.72	0.12	0.36	-0.01	0.29	0.02
	0.5	11.28	0.13	0.33	-0.01	0.32	0.02
	1	10.89	0.14	0.34	-0.01	0.30	0.02
	2	11.99	0.11	0.33	0.00	0.35	0.01
I3CA (mM)	0.25	9.13	0.10	0.44	-0.06	0.38	0.02
	0.5	9.60	0.13	0.42	-0.01	0.32	0.02
	1	9.96	0.22	0.37	0.02	0.27	0.03
	2	11.35	0.15	0.33	0.00	0.31	0.02
IAA (mM)	0.25	8.69	0.15	0.31	0.00	0.20	0.02
	0.5	10.04	0.15	0.33	-0.01	0.31	0.02
	1	9.27	0.12	0.31	-0.03	0.31	0.01
	2	6.84	0.12	0.34	-0.04	0.22	0.02
IPA (mM)	0.25	9.17	0.13	0.38	-0.02	0.25	0.02
	0.5	10.60	0.14	0.34	-0.01	0.22	0.02
	1	10.04	0.13	0.34	-0.01	0.23	0.02
	2	9.91	0.11	0.34	-0.03	0.25	0.01

Table 4: Growth kinetics of Healthy-associated *F. nucleatum* in response to tryptophan metabolites modeled by the Gompertz equation. Gompertz model parameters—including inflection time, growth rate, OD max, initial value, R^2, and maximum growth rate—for *F. nucleatum* cultures treated with various concentrations of tryptophan (Trp) and its metabolites: indole, indole-3-aldehyde (I3A), indole-3-carboxylic 141 acid (I3CA), indole-3-acetic acid (IAA), and indole-3-propionic acid (IPA).

Table 5: Oral Gompertz Model Parameters							
	Concentration	Inflection Time	Growth Rate	OD MAX	Initial Value	R squared	Max Growth Rate
Control	0	8.54	0.73	0.37	0.02	0.63	0.10
Tryp (mM)	1.25	12.19	0.84	0.32	0.00	0.48	0.10
	2.5	12.27	0.68	0.36	0.01	0.59	0.09
	5	12.27	0.62	0.38	0.00	0.66	0.09
	10	12.80	0.56	0.38	0.01	0.76	0.08
Indole (mM)	0.25	10.30	0.68	0.39	0.03	0.62	0.09
	0.5	11.08	0.61	0.38	0.02	0.66	0.08
	1	12.57	0.58	0.40	0.01	0.75	0.08
	2	19.46	0.38	0.38	0.00	0.88	0.05
I3A (mM)	0.25	13.07	0.55	0.38	0.02	0.70	0.07
	0.5	13.61	0.60	0.33	0.00	0.64	0.07
	1	14.61	0.61	0.35	0.01	0.75	0.08
	2	18.15	0.53	0.34	0.00	0.89	0.06
I3CA (mM)	0.25	13.09	0.67	0.37	0.01	0.66	0.09
	0.5	13.33	0.52	0.40	0.02	0.76	0.07
	1	14.95	0.56	0.37	0.00	0.78	0.08
	2	19.67	0.40	0.40	0.01	0.90	0.06
IAA (mM)	0.25	11.69	0.75	0.33	0.00	0.54	0.09
	0.5	12.73	0.59	0.37	0.01	0.65	0.08
	1	12.12	0.57	0.40	0.01	0.71	0.08
	2	13.40	0.46	0.40	0.00	0.77	0.07
IPA (mM)	0.25	12.59	0.65	0.35	0.01	0.67	0.08
	0.5	13.11	0.57	0.35	0.01	0.71	0.07
	1	14.03	0.56	0.35	0.01	0.75	0.07
	2	16.30	0.41	0.43	0.03	0.79	0.06

Table 5: Growth kinetics of Oral-associated *F. nucleatum* in response to tryptophan 143 metabolites modeled by the Gompertz equation. Gompertz model parameters— 144 including inflection time, growth rate, OD max, initial value, R^2, and maximum growth 145 rate—for *F. nucleatum* cultures treated with various concentrations of tryptophan (Trp) 146 and its metabolites: indole, indole-3-aldehyde (I3A), indole-3-carboxylic acid.

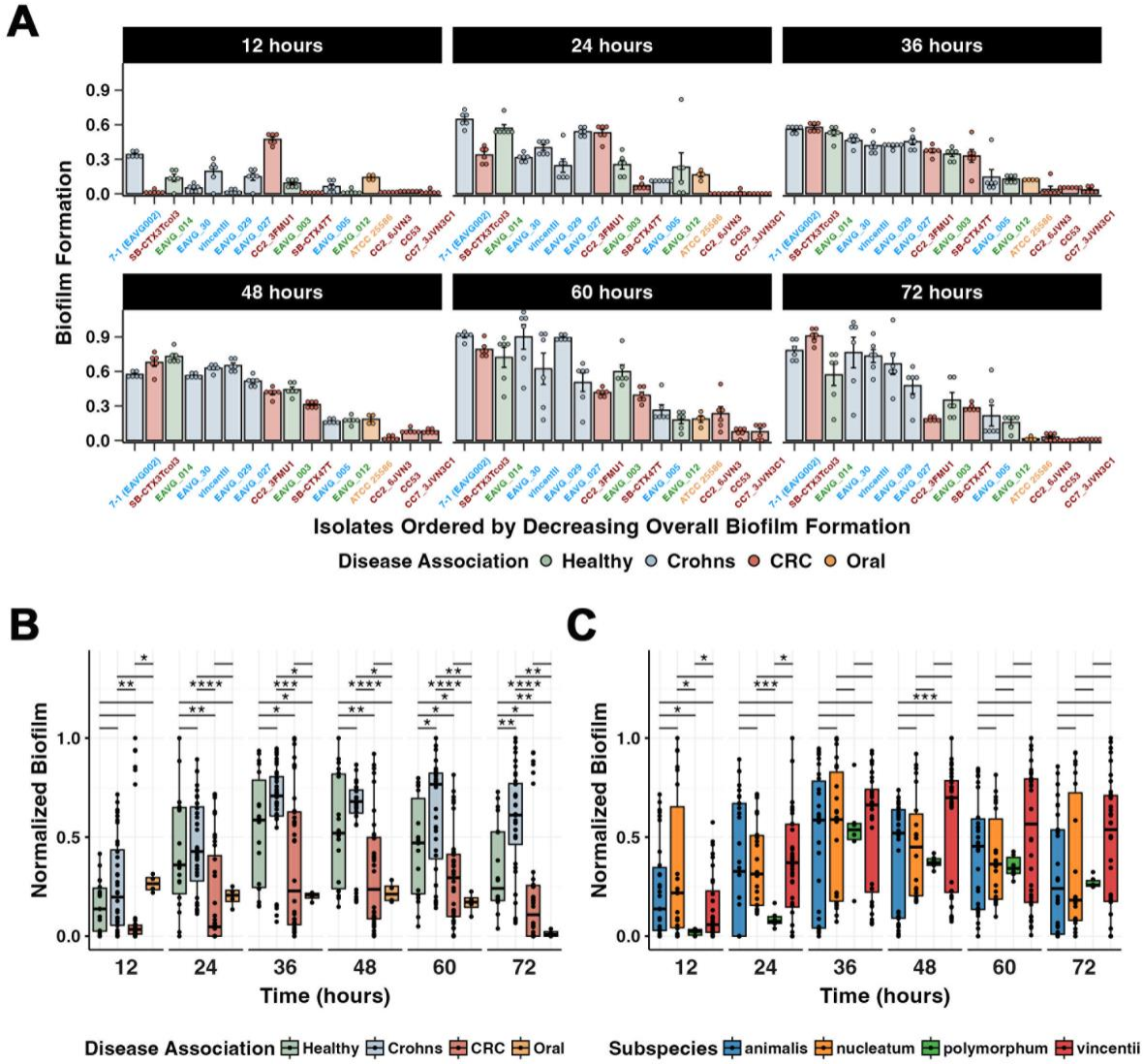


Figure 4: *F. nucleatum* isolates exhibit disease-associated differences in biofilm formation independent of subspecies identity. (A) Biofilm formation was quantified at six time points (12, 24, 36, 48, 60, and 72 hours) across four groups (Healthy, Crohn's, CRC, and Oral). Measurements were normalized to the maximum biofilm value at each time point. Group differences in total biofilm formation were analyzed by adjusted ANOVA (Benjamini–Hochberg correction), followed by Tukey's HSD test for pairwise comparisons. Statistical significance is denoted as $p < 0.05$ (*), $p < 0.01$ (**), $p < 0.001$ (***), and $p < 0.0001$ (****). **(B, C)** Boxplots showing normalized biofilm formation of *Fusobacterium nucleatum* isolates grouped by disease association **(B)** and subspecies **(C)**. Biofilm values were normalized to the maximum biofilm measurement at each time point, with numeric values indicating the proportion relative to that time point. Isolate groups: CRC (n=6), Crohn's (n=6), healthy (n=3), oral lesion (n=1).

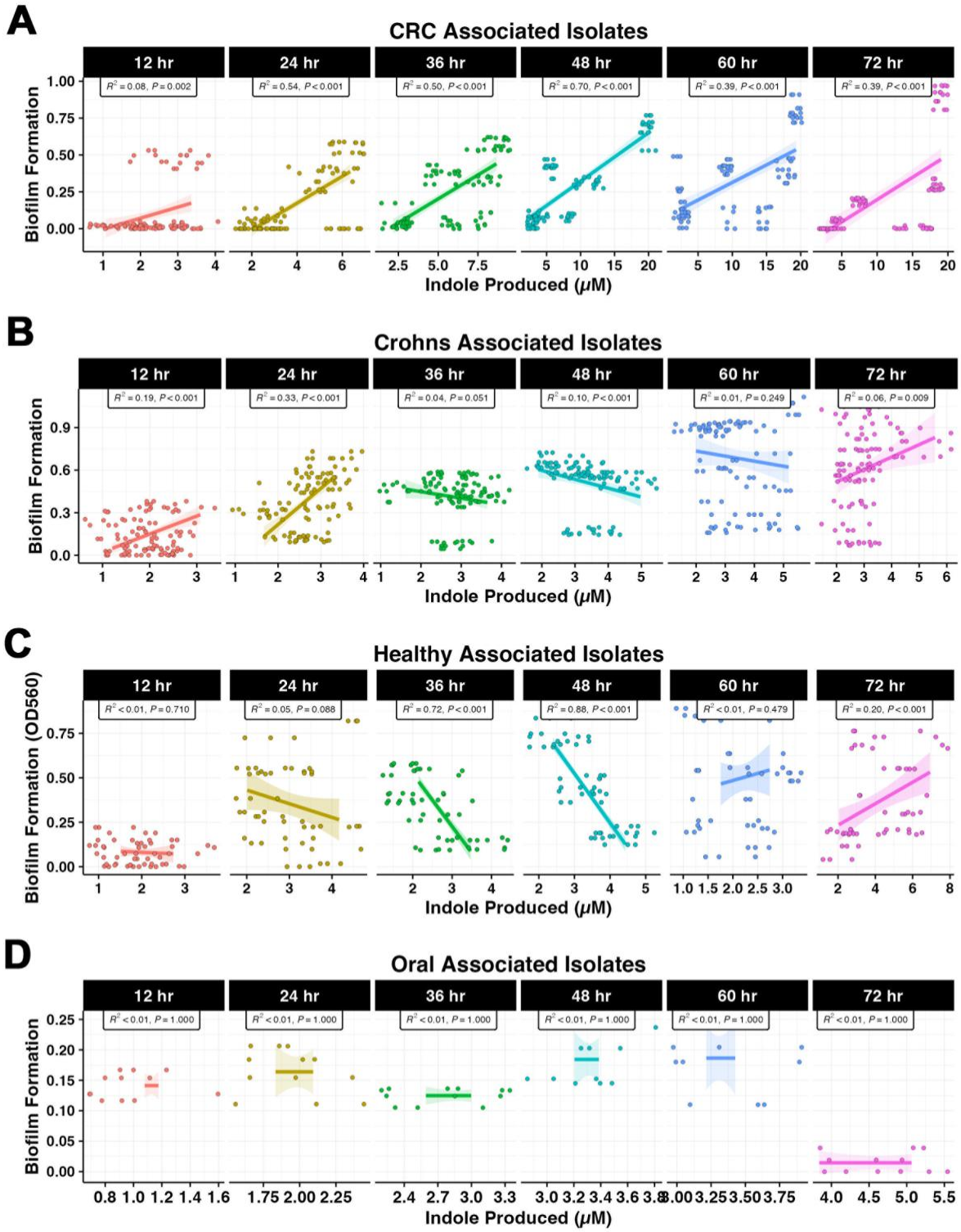


Figure 5: Biofilm formation by CRC-associated *F. nucleatum* isolates is positively correlated with indole production. The relationship between total indole production

(μM) measured with Kovac's reagent and biofilm formation was analyzed over time in bacterial isolates from different disease conditions: **(A)** CRC, **(B)** Crohn's, **(C)** healthy, and **(D)** oral-associated isolates. Measurements were collected at 12, 24, 36, 48, 60, and 72 hours. Each point represents data from a single bacterial isolate, with biofilm formation and total indole production determined through independent parallel experiments conducted under identical growth conditions. Pearson correlation analysis was performed, and regression lines with 95% confidence intervals (shaded areas) are shown for each time point. Isolates grouped by disease association (CRC: n=6, Crohn's: n=6, healthy: n=3, oral lesion: n=1).

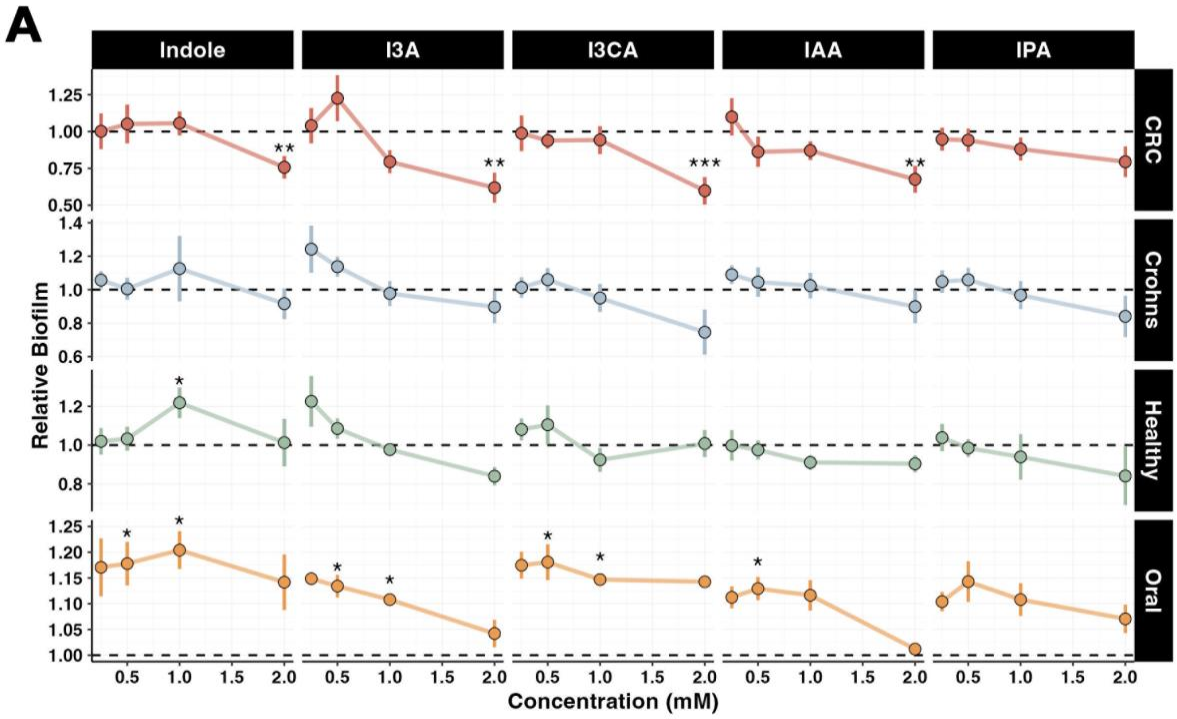
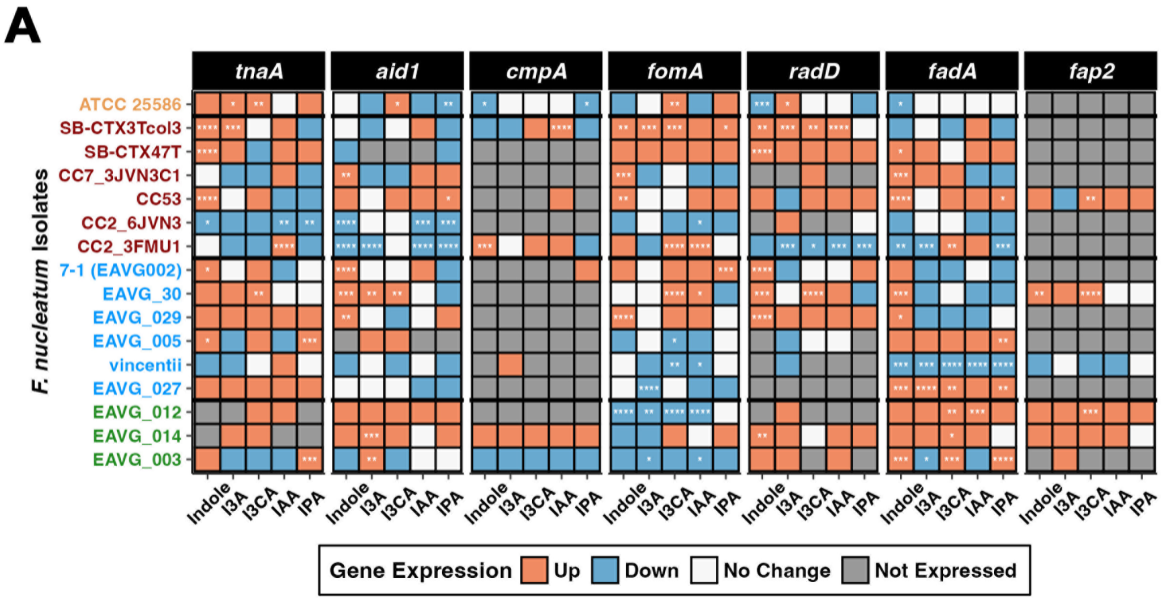
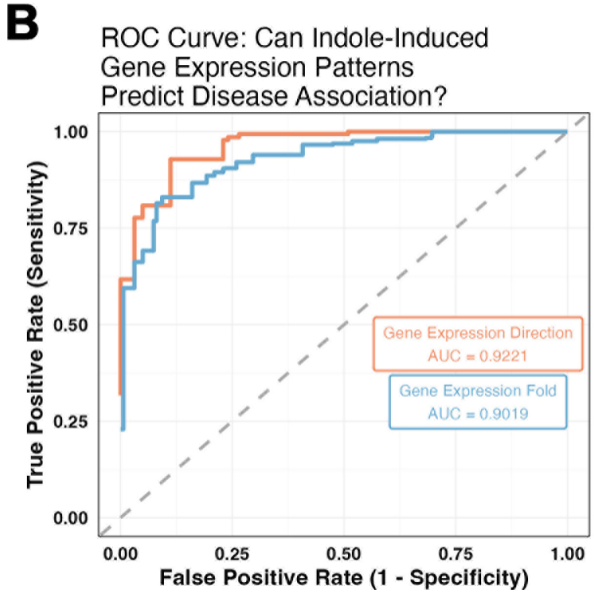


Figure 6: Exposure to exogenous indole derivatives significantly diminish biofilm formation in CRC-associated *F. nucleatum* isolates: Indole-induced changes in biofilm formation were assessed at concentrations ranging from 0.5 to 2.0 mM. Baseline biofilm formation is indicated by dotted lines, and relative changes from baseline biofilm (relative) across treatment concentration were quantified. Statistical comparisons between treated and control samples were performed using adjusted ANOVA with Benjamini-Hochberg correction, followed by Tukey's HSD test for pairwise comparisons. Statistical significance levels are denoted by asterisks (* $p<0.05$, ** $p<0.01$, *** $p<0.001$, **** $p<0.0001$). Isolates grouped by disease association (CRC: n=6, Crohn's: n=6, healthy: n=3, oral lesion: n=1).





Gene Expression Direction: Confusion Matrix and Statistics

	CRC	Crohn's	Healthy
CRC	119	32	19
Crohn's	37	127	1
Healthy	6	2	61
Accuracy:	0.7599		
P-Value:	< 2e-16		

Gene Expression Fold: Confusion Matrix and Statistics

	CRC	Crohn's	Healthy
CRC	118	1	26
Crohn's	44	149	24
Healthy	0	11	31
Accuracy:	0.7376		
P-Value:	< 2.2e-16		

Figure 7: Indole derivatives elicit distinct virulence expression patterns in *F. nucleatum* based on isolate origin and disease association. (A) Fold expression was calculated using the ΔΔCt method, where ΔΔCt = (Ct_target − Ct_16s gene) _treated − (Ct_target − Ct_16s)_control. Relative fold expression was quantified as 2^-ΔΔCt, indicating fold change normalized to 16s gene and relative to the control. Heatmap shows relative expression changes in virulence genes after exposure to indole derivatives for each isolate. Each cell indicates the direction of gene expression in response to treatment relative to untreated controls. Threshold is as follows: fold>1.25 = "Up", 1.25>fold>0.75 = “No change”, fold<0.75 = “Down”. Color coding: red = upregulation, blue = downregulation, white = no significant change, gray = gene not expressed. Pairwise t-tests (vs. DMSO) were performed within each isolate and target group. Bonferroni

correction was applied for multiple comparisons. Asterisks represent statistical significance: (* $p<0.05$, ** $p<0.01$, *** $p<0.001$, **** $p<0.0001$). **(B)** A naïve Bayes classification model was used to determine whether there exists a specific indole response by each disease state. Due to having only 1 isolate, the oral lesion was excluded from the classifier. The gene expression fold data and the gene expression direction calculated with the threshold mentioned above was split into training (70%) and test data (30%). Disease states were classified using a Naive Bayes model with kernel density estimation implemented in R. The Naive Bayes classifier was trained using the formula disease ~., indicating that all available features were utilized as predictors of disease status. Isolates grouped by disease association (CRC: n=6, Crohn's: n=6, healthy: n=3).

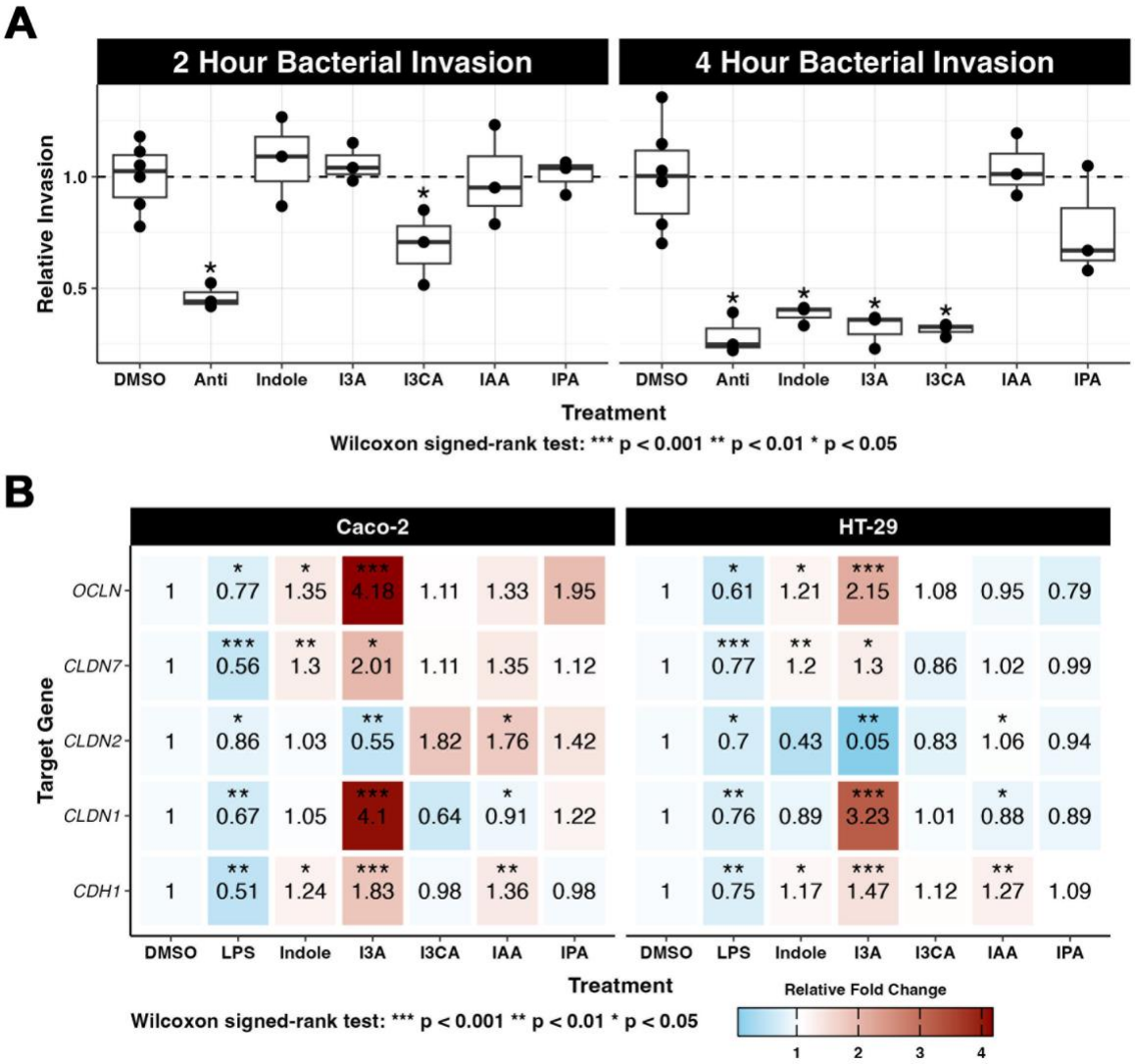


Figure 8: Indole derivatives alter tight junction integrity in CRC cells and restrict bacterial invasion. (A) Invasion assay of *F. nucleatum* (SB-CTX3Tcol3) isolated from a CRC patient into Caco-2 cells. *F. nucleatum* cultures were grown to exponential phase, adjusted to a MOI of 50, and added to Caco-2 cells treated with 1 mM indole derivatives (indole, I3A, I3CA, IAA, IPA), vehicle control (0.1% DMSO), or a combination of penicillin (200 U/mL) and streptomycin (200 µg/mL) as a negative control (Anti). After 2 or 4 hours of incubation, cells were washed and further treated with penicillin/streptomycin for 1 hour to eliminate extracellular bacteria, then washed and lysed with sterile deionized water. Lysates were serially diluted, plated under anaerobic conditions, and colony-forming units (CFU/mL) were quantified **(B)** HT-29 and Caco-2 cells were treated for 24 h with 1 mM indole, indole-3-aldehyde (I3A), indole-3-carboxylic acid (I3CA), indole-3-acetic acid (IAA), indole-3-propionic acid (IPA), DMSO (0.1%) as vehicle control, or lipopolysaccharide (LPS, 10 µg/mL) as a positive control for barrier disruption. Total RNA was extracted using TRIzol, reverse-transcribed into cDNA, and analyzed by qPCR for the expression of tight junction genes claudin-1 (CLDN1), claudin-2 (CLDN2), and claudin-7 (CLDN7), as well as adherens junction genes occludin (OCLN) and E-cadherin

(CDH1). Relative fold expression change (treatment vs. DMSO control) is shown in the heat map, with pairwise comparisons assessed by Wilcoxon signed-rank test.