Comparative Analysis of Indole Derivative Production Among Clinical Isolates of *Fusobacterium nucleatum*

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

Objectives: To investigate how indole production modulates *Fusobacterium nucleatum* virulence across different disease states.

Methods: Total indole production was measured in 16 *F. nucleatum* isolates (CRC: n=6, Crohn's: n=6, healthy: n=3, oral: n=1) using Kovac's reagent. Disease-associated indole derivatives—indole, indole-3-acetic acid (IAA), indole-3-aldehyde (I3A), indole-3-carboxaldehyde (I3CA), and indole-3-propionic acid (IPA)—were quantified via HPLC. Isolates were treated with exogenous derivatives to assess impacts on biofilm formation and virulence gene expression. A naïve Bayes classifier trained on virulence factor data evaluated niche-specific responses.

Results: CRC isolates produced significantly more total indole than healthy isolates ($p<0.05$). Indole derivatives minimally affected non-CRC isolate growth but reduced CRC isolate carrying capacity. Total indole positively correlated with biofilm formation in CRC isolates ($R^2=0.50$-0.70; $p<0.001$). Crohn's isolates formed most biofilm but were unresponsive to indoles; CRC isolates showed reduced formation. A Naïve Bayes classifier accurately predicted disease association from indole-induced virulence factor expression (AUC=0.90-0.92, Accuracy=0.74-0.75; $p<2e-16$).

Conclusions: CRC-associated isolates show unique metabolic profiles with elevated indole production and respond distinctively to indole exposure, suggesting potential therapeutic targets.

INTRODUCTION

Since the discovery that *F. nucleatum* selectively colonizes colorectal cancer (CRC) tissue (1,2), numerous studies have investigated how this bacterium both directly and indirectly contributes to CRC pathogenesis. Without known DNA-damaging toxins, *F. nucleatum* relies on type V secretion systems with outer membrane proteins that facilitate bacterial interactions (e.g. CmpA, RadD, aid1) (3,4), and host cell binding (e.g. RadD, Fap2, FadA) (5–8). Despite extensive research on host responses to *F. nucleatum* virulence factors, the regulatory mechanisms governing virulence expression differences among clinical isolates and subspecies remain poorly characterized. Given indoles' established role in modulating microbial virulence, investigating how *F. nucleatum* senses, metabolizes, and regulates indole production is essential for understanding and potentially controlling its pathogenic behavior.

Indole is a key bacterial signaling molecule that regulates collective behaviors and modulates epithelial responses to gut pathogens (9). *F. nucleatum*, a producer of indoles via Trp metabolism, has been shown to modulate its biofilm formation—a hallmark of CRC (10,11)—in response to indole exposure (12). More than 85 gram-positive and gram-negative species are known to hydrolyze tryptophan into the parent compound indole via the tryptophanase enzyme (tnaA) (13). Indole derivatives, through activation of aryl hydrocarbon receptor (AhR) and pregnane X receptor (PXR), regulate intestinal function and are increasingly investigated for their potential to modulate chemotherapy efficacy in cancer treatment (14). However, the impact of increased indole concentrations on CRC-associated pathogens remains poorly understood. While elevated indole levels have been shown to enhance biofilm formation in an oral *F. nucleatum* isolate (12), the mechanisms by which indole derivatives influence biofilm development and virulence in CRC-associated *F. nucleatum* strains are still unclear.

F. nucleatum is a central component of oral biofilms, bridging commensals and pathogens (15), and is believed to be the source of colon-derived *F. nucleatum* (16). In the CRC microenvironment, *F. nucleatum* is selectively enriched in biofilm-positive colon tumors (11). These *F. nucleatum*-positive biofilms are associated with reduced E-cadherin expression, increased IL-6/STAT3 signaling, and heightened crypt cell proliferation in adjacent normal colonic mucosa (10). Understanding the mechanisms that regulate *F. nucleatum* biofilm formation is therefore critical to elucidating its role in CRC progression.

This study aimed to characterize indole production profiles of clinical *F. nucleatum* isolates across various disease associations, with particular emphasis on determining whether increased indole derivative concentrations modulate isolate virulence. We conducted targeted HPLC analysis across 16 *F. nucleatum* clinical isolates from healthy individuals or those with disease and evaluated how indole exposure affects biofilm formation and virulence factor expression. Results revealed distinct indole production profiles among isolates, with CRC-associated isolates demonstrating significantly elevated production of indole derivatives. Notably, CRC-associated isolates exhibited heightened sensitivity to indole exposure, resulting in concentration-dependent decreases in biofilm formation. In contrast, the oral lesion

strain ATCC 25586, which is extensively cited in literature, displayed increased biofilm formation in response to indole exposure, indicating environment-specific differential responses to indole among isolates. These findings reveal a novel direction for investigating the regulatory mechanisms underlying *F. nucleatum* virulence, with potential implications for therapeutic interventions targeting indole-mediated pathways.

MATERIALS AND METHODS

Bacterial growth conditions: *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 48-72 hours. Single colonies were transferred to anaerobically reduced Columbia broth and grown under identical anaerobic conditions at 37°C for 24-48 hours before dilution and use in experiments. Culture purity was regularly confirmed through Gram staining.

Biofilm formation Assay: Biofilm formation by *F. nucleatum* isolates was assessed using previously described methods (12) 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% crystal violet for 1 minute, washed with deionized water until 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).

Time course Biofilm Assay 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 timepoints (12, 24, 36, 48, 60, and 72 hours), measured for optical density (OD600) using a Cerillo Stratus plate reader, and subsequently processed using the previously described biofilm formation assay.

Growth curves and indole treatment: To assess the impact of indole derivatives on *F. nucleatum* growth, sixteen isolates were cultured anaerobically in Columbia broth and treated with five compounds: indole, indole propionic acid (IPA), indole-3-acetic acid (IAA), indole-3-aldehyde (I3A), and indole-3-carboxylic acid (I3CA). 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.

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 (with 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 (0–72 h), and 100 µL of culture was frozen at –20°C for indole quantification. Indole quantification was performed according to previously established methods (17). Briefly, samples were thawed slowly at 4°C, then centrifuged at 3,000 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 OD560 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.

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. (18) and Kema et al. (19). 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 OD_{600} = 0.3 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 1/5 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 2x volumes 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). 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 (Table 2). 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.

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 means ± 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-change or binary gene expression. Oral lesion isolates were excluded due to low sample size. Models were 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. We validated model performance by checking class priors, calculating mutual information and Cramér's V, analyzing feature independence via correlation, and evaluating predictor importance through permutation-based accuracy loss.

RESULTS

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

Given the role of indole in modulating bacterial virulence, we investigated whether our *F. nucleatum* isolates differ in their capacity to produce indole. We obtained 15 clinical isolates of *F. nucleatum* from patient biopsies representing various disease states (**Table 1**), along with an oral lesion isolate, ATCC *F. nucleatum subsp. nucleatum* strain 25586 heavily cited in the literature. These isolates were categorized into groups based on their disease association: Healthy (3 isolates), Crohn's (6 isolates), CRC (6 isolates), or Oral Lesion (1 isolate). To investigate differences in total indole production profiles, we sampled bacterial cultures throughout the growth phase of *F. nucleatum*. Consistent with existing literature, our *F. nucleatum* isolates displayed varying levels of total indole production **(Figure 1A)**. Heightened indole production was observed at approximately 36 hours, suggesting that production primarily occurs during the transition from the exponential to stationary phase. Interestingly, *F. nucleatum* isolates associated with CRC patients exhibited the greatest variance within the disease categories for indole production. However, when categorized by disease association, the CRC isolates exhibited significantly higher indole production across all growth times compared to the other isolates **(Figure 1B)**. Despite variation in total indole production across different growth media, CRC-associated isolates consistently exhibited elevated indole production (**Supplemental Figure 1E**). These findings indicate that elevated indole production is a distinguishing metabolic feature of CRC-associated *F. nucleatum* isolates, potentially contributing to their enhanced virulence and selective advantage within the tumor microenvironment.

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

Recent metabolomic analysis revealed differential abundance of 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 to healthy controls (20). While overall indole/tryptophan ratios decreased, CRC patients showed elevated levels of specific compounds—I3A and I3CA. Conversely, IPA levels declined in both adenoma and CRC patients. The most intriguing finding was an overall enrichment of *Fusobacteria* in fecal samples from patients with adenomas and CRC. We aimed to investigate whether *F. nucleatum* isolates from

diseased environments preferentially produce specific disease-associated indole derivatives compared to strains from healthy environments. The indole production profiles of 16 *F. nucleatum* isolates were characterized using targeted HPLC analysis of spent media collected after 48 hours of growth, with derivative identification and quantification performed against authenticated standards. *F. nucleatum* isolates from different disease states showed distinct patterns of indole derivative production (**Figure 2A**). The parent indole compound was the predominant indole derivative produced across all isolates. Although all targeted derivatives were detected, most samples had IPA levels below the detection limit. *F. nucleatum* isolates from CRC, Crohn's, and oral lesions trended toward higher production of key indole derivatives—indole, I3A, I3CA, and IAA—relative to healthy isolates, consistent with total indole production data in **Figure 1**. Most notably, the parent indole derivative was significantly elevated in both CRC and Crohn's disease isolates compared to healthy isolates ($p < 0.05$). A similar trend was observed for I3CA and IAA, where disease-associated isolates, particularly those from CRC patients, showed increased production compared to healthy controls ($p < 0.05$). 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. For the 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 other derivatives. To explore this preference, the ratio of each indole derivative to the parent indole was calculated for each isolate. Examination of individual isolate-to-parent compound ratios (**Figure 2C**) revealed that CRC isolates, particularly CC2_6JVN3, 13 and 14, showed markedly higher I3A ratios (0.7865 and 0.8935, respectively) compared to healthy and oral isolates. Crohn's disease isolates displayed intermediate levels of indole derivatives production, with generally lower ratios than CRC isolates but higher than healthy controls. Isolates grouped by disease showed a more pronounced difference in indole derivative production than isolates grouped by subspecies, suggesting that disease association has a stronger influence on indole production patterns **(Supplemental Figure 2)**. These findings indicate that disease-associated *F. nucleatum* strains—especially those isolated from CRC patients—exhibit distinct metabolic capacities for indole production and can generate the indole derivatives also found in fecal samples from colorectal cancer.

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

Fecal indole concentrations vary widely from person to person, with levels ranging from 0.3 to 6.64 millimolar (mM) (17). We sought to identify an indole concentration within this range that could be used for exposure without exerting antimicrobial effects.

To assess the impact of increasing indole levels on *F. nucleatum* isolate growth, we continuously monitored the isolates for 72 hours using an automated plate reader while exposing them to increasing indole concentrations. The resulting growth curves were analyzed with the Gompertz mathematical model. The Gompertz model is used to

estimate growth parameters such as maximum growth rate, carrying capacity (ODmax), 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**). CRC isolates exhibited the highest sensitivity to indole at 2mM 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 3**). 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), as did healthy subject isolates (13% for indole, 0% for I3A, 0% for I3CA, IAA, and IPA). 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 3B-D and Tables 4-6**). 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, 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 negative inflection times in **Table 2** (-21.20, -74.89, -92.97) represent a mathematical anomaly when fitting Gompertz growth models to the bacterial growth data. Negative values likely appear when bacteria begin immediate rapid growth, face severe inhibition **(**e.g chloramphenicol treatment**, Supplemental Figure 3)**, or display atypical growth patterns that don't conform well to the assumptions of the Gompertz model's typical sigmoidal curve shape. Based on the growth patterns, indoles at 2 mM induced modifications to the bacterial growth curve, but the effects differed significantly from severe inhibition. Rather than completely suppressing bacterial proliferation, indole derivatives appear to have altered the fundamental growth dynamics of the bacteria 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.

Biofilm formation profiles distinguish disease-associated *F. nucleatum* isolates.

While *F. nucleatum* subspecies clearly differ in their ability to form biofilms, we wanted to determine whether isolates from CRC patients have an enhanced propensity to form biofilm. To test this, we monitored biofilm development in all isolates at 12-hour intervals over 72 hours. At 12 hours, several isolates showed minimal biofilm formation, with CC2_3FMU1 (CRC-associated) displaying the highest early biofilm formation **(Figure4A)**. 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 **(Supplemental Figure 4A)**. By 72 hours, a clear distinction emerged between high and low biofilm-forming isolates **(Supplemental Figure 4B)**, with the high biofilm formers predominantly

associated with Crohn's disease **(Figure 4B)**. 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 producers. Collectively, these data demonstrate that *F. nucleatum* isolates exhibit considerable variability in biofilm formation capability, paralleling their diverse indole derivative production profiles.

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

Bacteria are known to use indole production and biofilm formation as stress responses, increasing their survival in hostile environments. We sought to determine if total indole production corresponds with biofilm formation among our disease categories. This relationship was examined through linear correlation analysis comparing total indole against biofilm formation at defined intervals throughout the growth cycle from our two independent experiments. *F. nucleatum* isolates produce indole and form biofilms upon entering the stationary phase (at 36- and 48-hours), a phase triggered by cellular stress due to reduced nutrient availability. Looking at this 36-hour time point, a negative correlation between total indole production and biofilm formation was observed in healthy-associated isolates ($R^2 = 0.72$, $p < 0.001$) **(Figure 5C)**. Crohn's-associated isolates showed a weaker and non-significant negative correlation ($R^2 = 0.04$, $p = 0.051$) **(Figure 5B and D)**. However, a strong positive correlation was observed for the CRC-associated isolates ($R^2 = 0.5$, $p < 0.001$) 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 again exhibited no significant correlation ($R^2 < 0.01$, $p = 1$). 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 ($R^2 = 0.70$, $p < 0.001$). These results indicate that total indole production and biofilm formation are inversely related in healthy-associated and Crohn's-associated isolates, while CRC-associated isolates show a positive correlation at both 36 and 48 hours. Thus, total indole production may act as a biofilm-promoting signal in the tumor microenvironment, potentially enhancing bacterial persistence and pathogenicity in colorectal cancer.

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

To determine whether 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. Administration of exogenous indole and its derivatives (I3A, I3CA, IAA, and IPA) resulted in distinct biofilm formation responses that varied based on the 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. 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 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 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 indole exposure. Exposure to indole or 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 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 *vincentii*, while being 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—reduced in healthy isolates and increased in most CRC-associated isolates. To better assess

whether indole-induced responses differed by disease association, we used a naïve Bayes classification model to test if indole-induced expression patterns could predict disease origin for each isolate **(Figure 7B)**. Using the raw fold change data and the gene direction threshold, 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). The two models also exhibited a strong true positive rate, suggesting the models were very good at correctly identifying positive cases (AUC = 0.9019, AUC = 0.9221 respectively). Together, these results suggest that indole derivatives modulate *F. nucleatum* virulence factor expression in a strain-dependent manner, with isolates from the same disease environment showing similar responses to indole.

Naïve Bayes posterior probabilities and feature importance reveal class-specific predictive confidence based on gene expression direction and fold change.

To assess the model's confidence in classifying samples based on gene fold-change and directionality, we examined the posterior probabilities of true positive predictions across the three disease classes: CRC, Crohn's, and Healthy **(Figure 8A)**. Posterior probability reflects the naïve Bayes model's confidence, with values near 1 indicating high certainty and values near 1/3 (the random chance threshold in a three-class model) indicating low confidence. The gene direction model **(Figure 8A, left)** accurately classifies CRC, Crohn's, and Healthy isolates, showing strong confidence across groups (CRC avg = 0.662, Crohn's avg = 0.644, Healthy avg = 0.757). Healthy samples had the highest classification accuracy and lowest false positive rates (avg wrong CRC = 0.196, avg wrong Crohn's = 0.0468). Misclassifications typically fell below the 1/3 threshold (e.g., true CRC: avg wrong Crohn's = 0.271, avg wrong Healthy = 0.067; true Crohn's: avg wrong CRC = 0.292, avg wrong Healthy = 0.064), indicating the model appropriately expressed uncertainty when incorrect. The gene fold model **(Figure 8B, left)** also performed well, particularly for Crohn's (avg correct = 0.890; avg wrong Healthy = 0.0512, avg wrong CRC = 0.0593), and showed moderate confidence for CRC (avg correct = 0.697; avg wrong Crohn's = 0.286, avg wrong Healthy = 0.0167). However, it performed less reliably for Healthy samples, with more variable predictions (avg correct = 0.387; avg wrong Crohn's = 0.323, avg wrong CRC = 0.290). Nevertheless, gene directionality data appeared to be more effective at predicting all disease classes compared to other fold data. These data suggest that indole-induced virulence factor expression response may be specific to the isolate's disease origin. To identify which gene features most strongly influence model performance, we applied Leave-One-Feature-Out (LOFO) importance—also known as permutation importance—which quantifies the individual contribution of each feature by measuring the change in model accuracy when it is removed. In the directionality model, FomA and RadD caused the greatest drop in accuracy when removed **(Figure 8A, right)**, indicating their high predictive importance, while aid1 emerged as the most influential feature in the fold-change model **(Figure 8B, right)**. Taken together, these findings suggest that indole may drive the expression of specific virulence factors in a manner tailored to the environmental niche.

DISCUSSION

A key unanswered question is how *F. nucleatum* transitions from an oral commensal organism to a virulent intestinal pathogen, and which molecular mechanisms drive its dual role in host health. *F. nucleatum* exhibits significant heterogeneity, with strains categorized into subspecies based on genomic and physiological analyses (21–28)—so pronounced that reclassification into separate species has been proposed (29). At the subspecies level, *F. nucleatum* subspecies *animalis* specifically correlates with increased CRC-specific mortality and distinctive somatic gene mutations (30). Recently, subspecies *animalis* has been subdivided into two distinct clades characterized by the differential presence of key virulence factors (31), with clade Fna C2—which notably possess fap2 and cmpA genes—were found to be significantly enriched within biofilm-positive CRC tumor microenvironments (11). Given that specific virulence factors in *F. nucleatum* serve as biological markers of its association with CRC, elucidating the mechanisms that regulate the expression of these virulence factors is essential to comprehend the complex community interactions between this organism and disease pathology. Our study offers new insight by highlighting an underexplored virulence mechanism. To our knowledge, this is the first comparative analysis of indole production across clinical *F. nucleatum* isolates, revealing that indole derivatives may modulate virulence in a disease- and environment-specific manner through modulation of biofilm formation and virulence factor expression. This is supported by our data showing distinct indole production profiles that correlate with disease association and differential responses to indole exposure.

The human gut contains numerous signals that regulate host-microbiota and microbiota-microbiota interactions to maintain GI homeostasis (32–36). Microbiota-derived tryptophan metabolites, such as indole, support intestinal barrier function and modulate host immunity (9,37), including direct effects on the tumor immune microenvironment (38,39). These effects vary by indole species: indole-3-acetic acid (IAA) suppresses immune responses in pancreatic cancer by activating the aryl hydrocarbon receptor (AhR) in tumor-associated macrophages, reducing $CD8^+$ T cell activation (39). Supplementation with IAA or IAA-producing bacteria alleviates colitis in mice, while loss of AhR or reduced IAA levels (such as from a high-fat diet) worsens inflammation and barrier dysfunction (40). Certain indole derivatives exhibit dose-dependent bidirectional effects on host cells, with indole-3-acetaldehyde (IAAD) inhibiting cell proliferation and inducing apoptosis at low concentrations, thereby preventing CRC development, while at higher concentrations (>25 µmol/L), it promotes epithelial-mesenchymal transition and invasion in CRC cells (41). Similarly, indole-3-aldehyde (I3A) reduces colitis induced inflammation and helps restore intestinal epithelial barrier function (42), in part by activating AhR, which can also promote IL-22 secretion (43). While IL-22 is protective in some contexts, it can promote tumor growth in colorectal cancer (44), highlighting the context-dependent roles of indole derivatives.

Metabolism of tryptophan into indoles is a key metabolic pathway commonly altered in gastrointestinal diseases (20,45–47). Specifically, in multiple IBD and CRC studies tryptophan and indole levels are significantly different from controls, with higher fecal tryptophan but lower serum tryptophan levels suggesting possible malabsorption (48–53). A meta-analysis of fecal samples revealed elevated tryptophan metabolism and decreased tryptophan biosynthesis in tumor samples compared to non-tumor

controls, suggesting possible increased indole production in the tumor microenvironment (54–56) (**Supplemental Figure 9**). Due to the established enrichment of *F. nucleatum* in CRC tissue (1,2,11,31), we investigated whether CRC-associated *F. nucleatum* isolates could contribute to elevated tryptophan metabolism. Comparative analysis revealed that CRC-associated *F. nucleatum* isolates produced significantly higher levels of total indole than isolates from other disease groups, suggesting a potential link between CRC and *F. nucleatum*-mediated indole production. Supporting this, indole derivatives (indole, I3A, I3CA) enriched in CRC fecal samples (20) positively correlated with indole levels in CRC-derived isolates. Notably, I3A—elevated in CRC samples—was preferentially produced by three CRC-associated *F. nucleatum* isolates (CC2_6JVN3, SB-CTX47T, SB-CTX3Tcol3). These findings suggest a potential metabolic adaptation in CRC-associated isolates toward increased indole production, likely driven by microenvironmental cues.

Fecal indole levels range from 0.3–6.6 mM (17), with *in vitro* studies showing similar concentrations (0.5–5 mM) (57). At low levels (0.5–1 mM), indole modulates *E. coli* virulence by affecting biofilm formation, gene expression, and stress responses (58,59), while higher levels (4–5 mM) inhibit growth and division (60). We examined whether low concentrations could similarly modulate *F. nucleatum* growth kinetics without exerting antimicrobial effects. At 0.25–2 mM, indole primarily reduced carrying capacity (OD max) without altering inflection time, with CRC isolates showing the greatest sensitivity as compared to other non-CRC isolates. Overall, minimal growth effects at 1 mM across all derivatives suggest phenotypic modulation rather than direct toxicity.

Biofilm analysis revealed distinct high and low biofilm-forming isolates. SB-CTX3Tcol3 (CRC) and EAVG002-7_1 (Crohn's) peaked at 36 hours, identifying them as strong biofilm producers and suggesting potential for persistent virulence. EAVG002-7_1, part of a specific *F. nucleatum* clade associated with CRC (31), has been shown to induce colonic tumors in mouse models where most isolates fail to colonize (61,62). Interestingly, Crohn's-associated isolates were the strongest biofilm formers, while only CC2_3FMU1 and SB-CTX3Tcol3 showed strong formation among CRC isolates. When analyzing the 36 and 48-hour time points—which mark the beginning of stationary phase when indole production and biofilm formation are typically activated—only CRC isolates demonstrated this positive correlation with total indole production. These results suggest that CRC-associated *F. nucleatum* isolates possess a unique metabolic relationship between total indole signaling and biofilm development that distinguishes them from non-CRC isolates, which may provide a fitness advantage in the malignant microenvironment.

Recent studies in the field have highlighted virulence clades of *F. nucleatum* based on the presence and absence of certain virulence factors, potentially serving as biological markers of its association with CRC (31). During active invasion, *F. nucleatum* reprograms its metabolic and functional pathways to adapt to the host-imposed environment, marked by the upregulation of metabolic processes, quorum sensing, and amino acid metabolism (63). In addition, *F. nucleatum* switches it's the expression profile during the transition from planktonic to the biofilm phase, another pathway

mediated by upregulation of amino acid metabolism (64). These data support the concept that *F. nucleatum* tailors its metabolic and transcriptional profiles in response to environmental cues. However, the transcriptional regulation of *F. nucleatum* in response to specific indole derivatives, compounds found to be abundant in the CRC environment (20), remains poorly understood. Analysis of key virulence factors (Aid1, FomA, Fap2, FadA, CmpA, and RadD) following exposure to five indole derivatives (indole, I3A, I3CA, IAA, and IPA) revealed distinct, isolate-specific transcriptional responses. Interestingly, disease-specific responses were particularly evident for FomA: most CRC-associated isolates exhibited increased FomA expression across nearly all indole treatments, whereas most health-associated isolates showed a corresponding decrease in expression. This disease-specific response presents an opportunity to selectively target CRC-associated *F. nucleatum* isolates by exploiting their heightened FomA expression in response to indole. Our lab has developed guided antimicrobial peptides (gAMPs) against FomA , and these findings may enhance their efficacy by enabling more precise eradication of pathogenic isolates linked to CRC through combination of indole and gAMP treatment.

To further validate the disease-specific expression patterns of virulence factors in response to indole derivatives, we employed a Naïve Bayes classification model trained on indole-induced expression profiles (gene directionality and fold expression data), with the objective of predicting the disease association of each isolate. Both models accurately classified CRC, Crohn's, and healthy isolates (fold expression: 74%, $P < 2.2e-16$; directionality: 76%, $P < 2e-16$). These findings suggest that both indole production and indole sensitivity are shaped by environmental pressures, contributing to niche-specific adaptation in *F. nucleatum*. Contrary to our initial hypothesis that indole derivatives would enhance all aspects of *F. nucleatum* virulence, CRC-associated isolates showed reduced biofilm formation alongside increased virulence gene expression. A similar inverse relationship has been reported in *P. aeruginosa*, where disruption of the Entner-Doudoroff pathway decreased biofilm but increased lethality and immune activation in infection models (66). Indole exposure may likewise trigger metabolic pathways that suppress biofilm while enhancing immune resilience in *F. nucleatum*. This response appears niche-specific, with oral isolates rapidly forming biofilms under distinct environmental pressures, while CRC isolates suppress biofilm formation in favor of niche-specific virulence behaviors. Further transcriptomic and proteomic analyses—particularly during active invasion of oral versus colorectal cancer cells—are needed to define the transcriptional programs driving these niche-specific behaviors.

Despite the novel insights presented, this study has several notable limitations. All experiments were performed in monoculture, lacking the polymicrobial and host-interactive context of the gut or tumor microenvironment, where microbial crosstalk and host interactions significantly influence bacterial behavior. The inclusion of only a single oral isolate limits conclusions about niche-specific responses. Although correlations between indole production, biofilm formation, and virulence gene expression are compelling, underlying regulatory mechanisms remain uncharacterized. Additionally, the positive correlation between total indole levels and biofilm formation in CRC-associated isolates suggests the existence of other, untested indole derivatives that may enhance

biofilm formation in these strains—contrasting with the derivatives examined in this study, which were associated with reduced biofilm production. The use of rich artificial media further reduces physiological relevance, and the absence of host models restricts translational applicability. It is likely isolate behavior may differ in the presence of indoles within a more biologically relevant polymicrobial fecal environment or during active host cell invasion—conditions currently under investigation in our lab.

The differential responses to indole exposure among *F. nucleatum* isolates from CRC, Crohn's disease, oral sites, and healthy individuals underscore the importance of investigating indole's role in disease pathogenesis and persistence. These findings suggest potential for targeted therapeutic strategies in colorectal, where inhibiting biofilm formation or indole production could reduce pathogenicity (67). Moreover, the use of gAMPs targeting virulence factors that are consistently upregulated in CRC-associated isolates following indole exposure—such as FomA—may enhance the selective eradication of highly virulent strains. Future research should examine how indole and its derivatives simultaneously modulate both bacterial and host transcriptomes during active infection and explore how these metabolites shape microbial community dynamics in polymicrobial environments.

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

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

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

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

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

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

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

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

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

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

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

14. Wang G, Fan Y, Zhang G, Cai S, Ma Y, Yang L, et al. Microbiota-derived indoles alleviate intestinal inflammation and modulate microbiome by microbial cross-feeding. Microbiome. 2024 Mar 19;12(1):59.

15. Sakanaka A, Kuboniwa M, Shimma S, Alghamdi SA, Mayumi S, Lamont RJ, et al. Fusobacterium nucleatum Metabolically Integrates Commensals and Pathogens in Oral Biofilms. mSystems. 2022 Jul 19;7(4):e00170-22.

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

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

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

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

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

21. Loesche WJ, Gibbons RJ. Amino acid fermentation by Fusobacterium nucleatum. Arch Oral Biol. 1968 Feb;13(2):191–202.

22. Dzink JL, Socransky SS. Amino acid utilization by Fusobacterium nucleatum grown in a chemically defined medium. Oral Microbiol Immunol. 1990 Jun;5(3):172–4.

23. Suchett-Kaye G, Décoret D, Barsotti O. Clonal analysis by ribotyping of Fusobacterium nucleatum isolates obtained from healthy young adults with optimal plaque control. J Periodontal Res. 1998 May;33(4):179–86.

24. Thurnheer T, Guggenheim B, Gruica B, Gmür R. Infinite Serovar and Ribotype Heterogeneity Among Oral *Fusobacterium nucleatum* Strains? Anaerobe. 1999 Apr 1;5(2):79–92.

25. Morris ML, Andrews RH, Rogers AH. Investigations of the Taxonomy and Systematics of Fusobacterium nucleatum Using Allozyme Electrophoresis. Int J Syst Evol Microbiol. 1997;47(1):103–10.

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

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

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

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

30. Borozan I, Zaidi SH, Harrison TA, Phipps AI, Zheng J, Lee S, et al. Molecular and pathology features of colorectal tumors and patient outcomes are associated with Fusobacterium nucleatum and its subspecies animalis. Cancer Epidemiol Biomark

Prev Publ Am Assoc Cancer Res Cosponsored Am Soc Prev Oncol. 2022 Jan;31(1):210–20.

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

32. Wang T, Zheng X, Ji H, Wang TL, Xing XH, Zhang C. Dynamics of transcription–translation coordination tune bacterial indole signaling. Nat Chem Biol. 2020 Apr;16(4):440–9.

33. Cameron E, Sperandio V. Frenemies: Signaling and nutritional integration in pathogen-microbiota-host interactions. Cell Host Microbe. 2015 Sep 9;18(3):275–84.

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

35. Bäumler AJ, Sperandio V. Interactions between the microbiota and pathogenic bacteria in the gut. Nature. 2016 Jul;535(7610):85–93.

36. Falà AK, Álvarez-Ordóñez A, Filloux A, Gahan CGM, Cotter PD. Quorum sensing in human gut and food microbiomes: Significance and potential for therapeutic targeting. Front Microbiol [Internet]. 2022 Nov 25 [cited 2025 Feb 18];13. Available from: https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2022.1002185/full

37. Wang D, Ding X, Rather PN. Indole Can Act as an Extracellular Signal in Escherichia coli. J Bacteriol. 2001 Jul;183(14):4210–6.

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

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

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

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

42. Wang M, Guo J, Hart AL, Li JV. Indole-3-Aldehyde Reduces Inflammatory Responses and Restores Intestinal Epithelial Barrier Function Partially via Aryl Hydrocarbon Receptor (AhR) in Experimental Colitis Models. J Inflamm Res. 2023 Dec 6;16:5845–64.

43. Zelante T, Iannitti RG, Cunha C, De Luca A, Giovannini G, Pieraccini G, et al. Tryptophan catabolites from microbiota engage aryl hydrocarbon receptor and balance mucosal reactivity via interleukin-22. Immunity. 2013 Aug 22;39(2):372–85.

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

45. Papadimitriou N, Gunter MJ, Murphy N, Gicquiau A, Achaintre D, Brezina S, et al. Circulating tryptophan metabolites and risk of colon cancer: results from case-control and prospective cohort studies. Int J Cancer. 2021 Nov 1;149(9):1659–69.

46. Liu Y, Pei Z, Pan T, Wang H, Chen W, Lu W. Indole metabolites and colorectal cancer: Gut microbial tryptophan metabolism, host gut microbiome biomarkers, and potential intervention mechanisms. Microbiol Res. 2023 Jul 1;272:127392.

47. Venkateswaran N, Lafita-Navarro MC, Hao YH, Kilgore JA, Perez-Castro L, Braverman J, et al. MYC promotes tryptophan uptake and metabolism by the kynurenine pathway in colon cancer. Genes Dev. 2019 Sep 1;33(17–18):1236–51.

48. Bosch S, de Meij TGJ, de Boer NK. Altered Tryptophan Levels in Patients With Inflammatory Bowel Disease Owing to Colonic Leakage, Metabolism, or Malabsorption? Gastroenterology. 2018 May 1;154(6):1855–6.

49. Cheng J, Jin H, Hou X, Lv J, Gao X, Zheng G. Disturbed tryptophan metabolism correlating to progression and metastasis of esophageal squamous cell carcinoma. Biochem Biophys Res Commun. 2017 May 6;486(3):781–7.

50. Nikolaus S, Schulte B, Al-Massad N, Thieme F, Schulte DM, Bethge J, et al. Increased Tryptophan Metabolism Is Associated With Activity of Inflammatory Bowel Diseases. Gastroenterology. 2017 Dec;153(6):1504-1516.e2.

51. Wang S, van Schooten FJ, Jin H, Jonkers D, Godschalk R. The Involvement of Intestinal Tryptophan Metabolism in Inflammatory Bowel Disease Identified by a Meta-Analysis of the Transcriptome and a Systematic Review of the Metabolome. Nutrients. 2023 Jun 26;15(13):2886.

52. Bustamante S, Yau Y, Boys V, Chang J, Paramsothy S, Pudipeddi A, et al. Tryptophan Metabolism “Hub” Gene Expression Associates with Increased Inflammation and Severe Disease Outcomes in COVID-19 Infection and Inflammatory Bowel Disease. Int J Mol Sci. 2022 Nov 26;23(23):14776.

53. Crotti S, D'Angelo E, Bedin C, Fassan M, Pucciarelli S, Nitti D, et al. Tryptophan metabolism along the kynurenine and serotonin pathways reveals substantial differences in colon and rectal cancer. Metabolomics. 2017 Dec;13(12):148.

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

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

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

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

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

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

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

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

62. Kostic AD, Chun E, Robertson L, Glickman JN, Gallini CA, Michaud M, et al. Fusobacterium nucleatum potentiates intestinal tumorigenesis and modulates the tumor immune microenvironment. Cell Host Microbe. 2013 Aug 14;14(2):207–15.

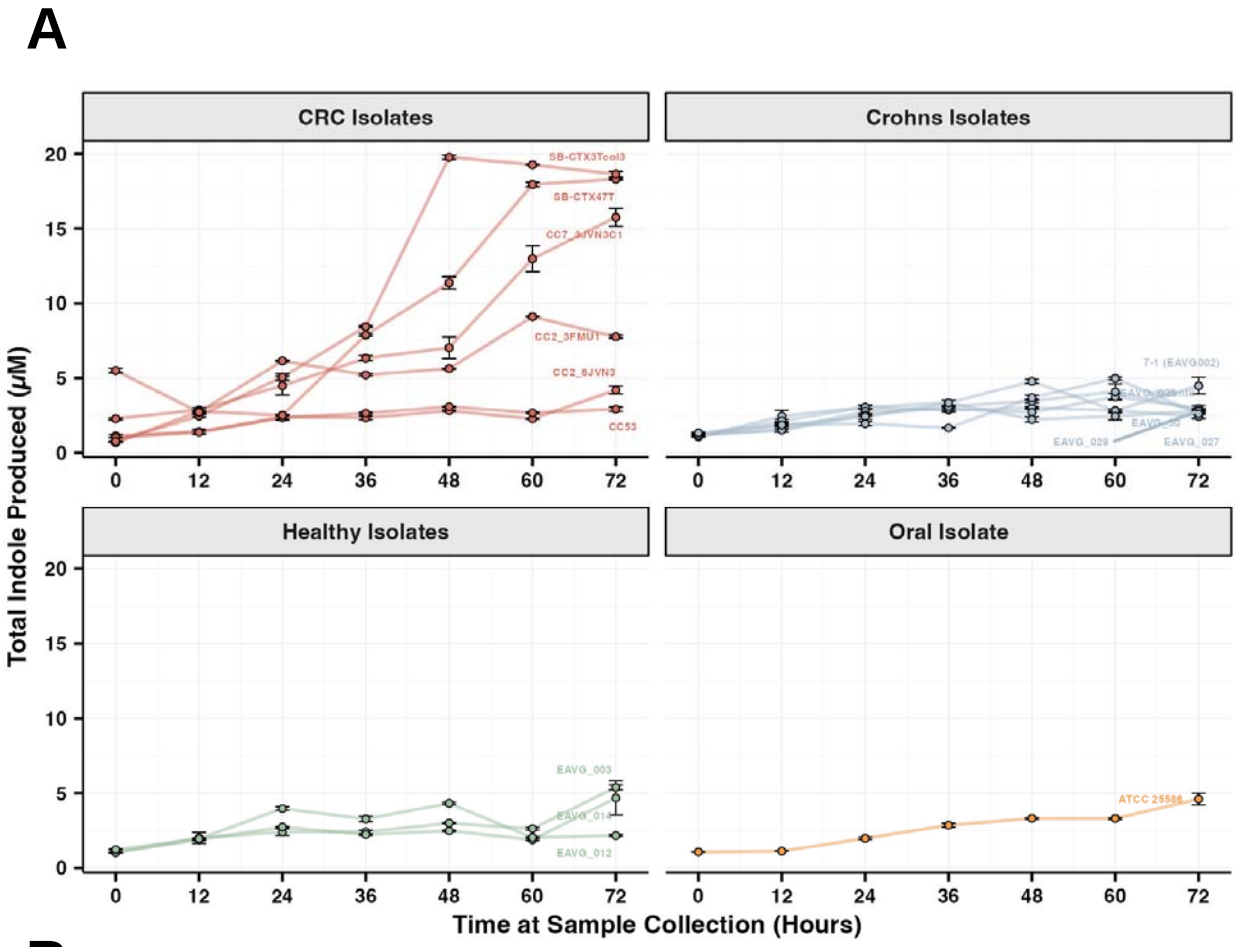
63. Cochrane K, Robinson AV, Holt RA, Allen-Vercoe E. A survey of Fusobacterium nucleatum genes modulated by host cell infection. Microb Genomics. 2020 Feb;6(2):e000300.

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

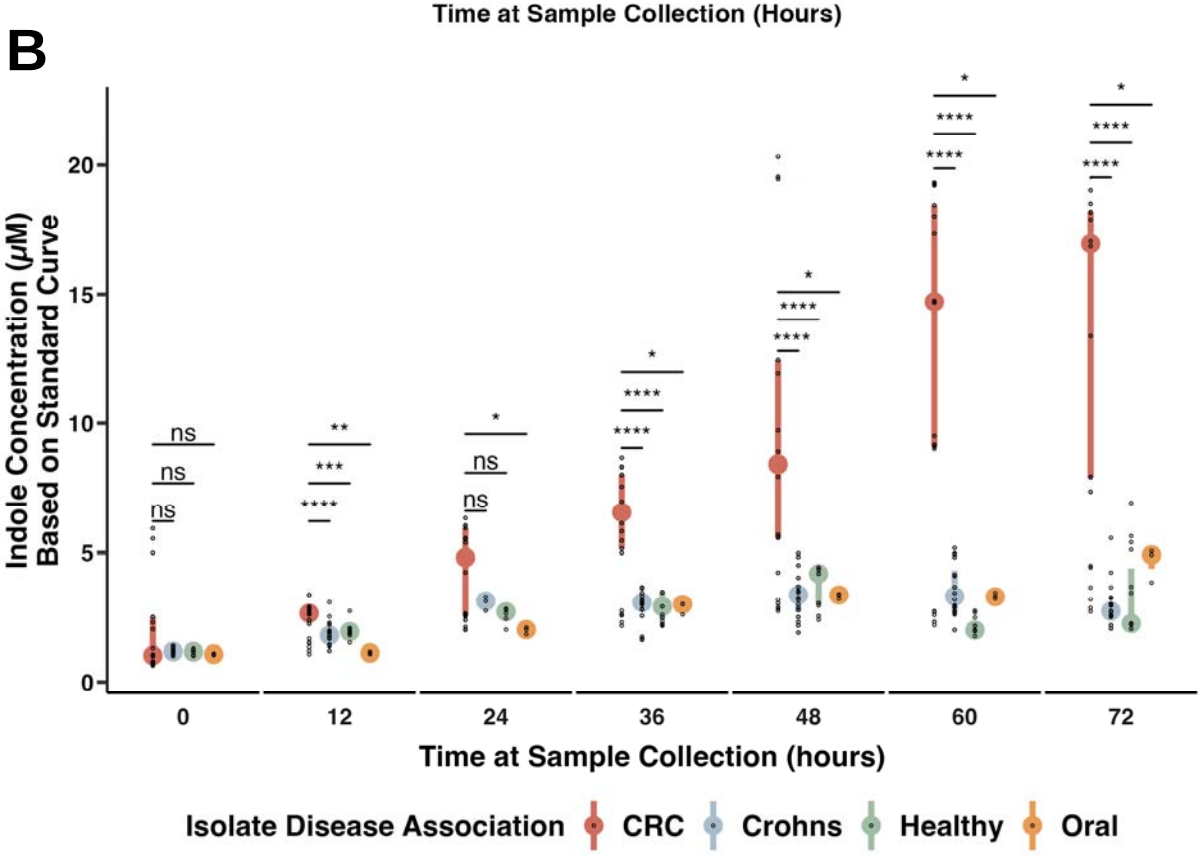
65. Choudhury A, Scano C, Barton A, Greathouse KL. Targeted therapy of oncomicrobe F. nucleatum with bioengineered probiotic expressing guided antimicrobial peptide (gAMP) [Internet]. bioRxiv; 2024 [cited 2025 Apr 16]. p. 2024.10.29.620919. Available from: https://www.biorxiv.org/content/10.1101/2024.10.29.620919v1

66. Pan S, Underhill SAM, Hamm CW, Stover MA, Butler DR, Shults CA, et al. Glycerol metabolism impacts biofilm phenotypes and virulence in Pseudomonas aeruginosa via the Entner-Doudoroff pathway. mSphere. 9(4):e00786-23.

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


A
CRC Isolates
Crohns Isolates
Healthy Isolates
Oral Isolate
Total Indole Produced (μM)
Time at Sample Collection (Hours)
SB-CTX3Tcol3
SB-CTX47T
CC7_3JVN3C1
CC2_3FMU1
CC2_6JVN3
CC53
7-1 (EAVG002)
EAVG_029
EAVG_027
EAVG_003
EAVG_012
ATCC 25586

838


B
Indole Concentration (μM) Based on Standard Curve
Time at Sample Collection (hours)
ns
**
***
****
*
Isolate Disease Association
CRC
Crohns
Healthy
Oral

839

Figure 1: Elevated indole production is a distinctive feature of colorectal cancer-associated *F. nucleatum* isolates. (1A) Total indole production by each *F. nucleatum* isolate over time quantified using Kovac's reagent and calibrated against a standard curve. The Kovacs reagent (p-dimethylaminobenzaldehyde in acidic solution) reacts with the indole ring at position 3, forming a conjugated imine that produces a characteristic pink/red color, with the intensity proportional to the total indole concentration. Data is presented as the mean total indole production (µM concentration), with error bars indicating the standard error of the mean (SEM) for the individual isolate. **(1B)** Total indole production of *F. nucleatum* isolates over time grouped by isolate disease association. An adjusted ANOVA was performed using the Benjamini-Hochberg method to compare total indole production between CRC-associated *F. nucleatum* isolates and non-CRC isolates. The Tukey's HSD test was then applied to identify significant pairwise differences. Data is represented as the median, with the interquartile range (IQR) capturing the central 50% of the data distribution. p values are indicated as follows: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$. Isolates grouped by disease association (CRC: n=6, Crohn's: n=6, healthy: n=3, oral lesion: n=1).

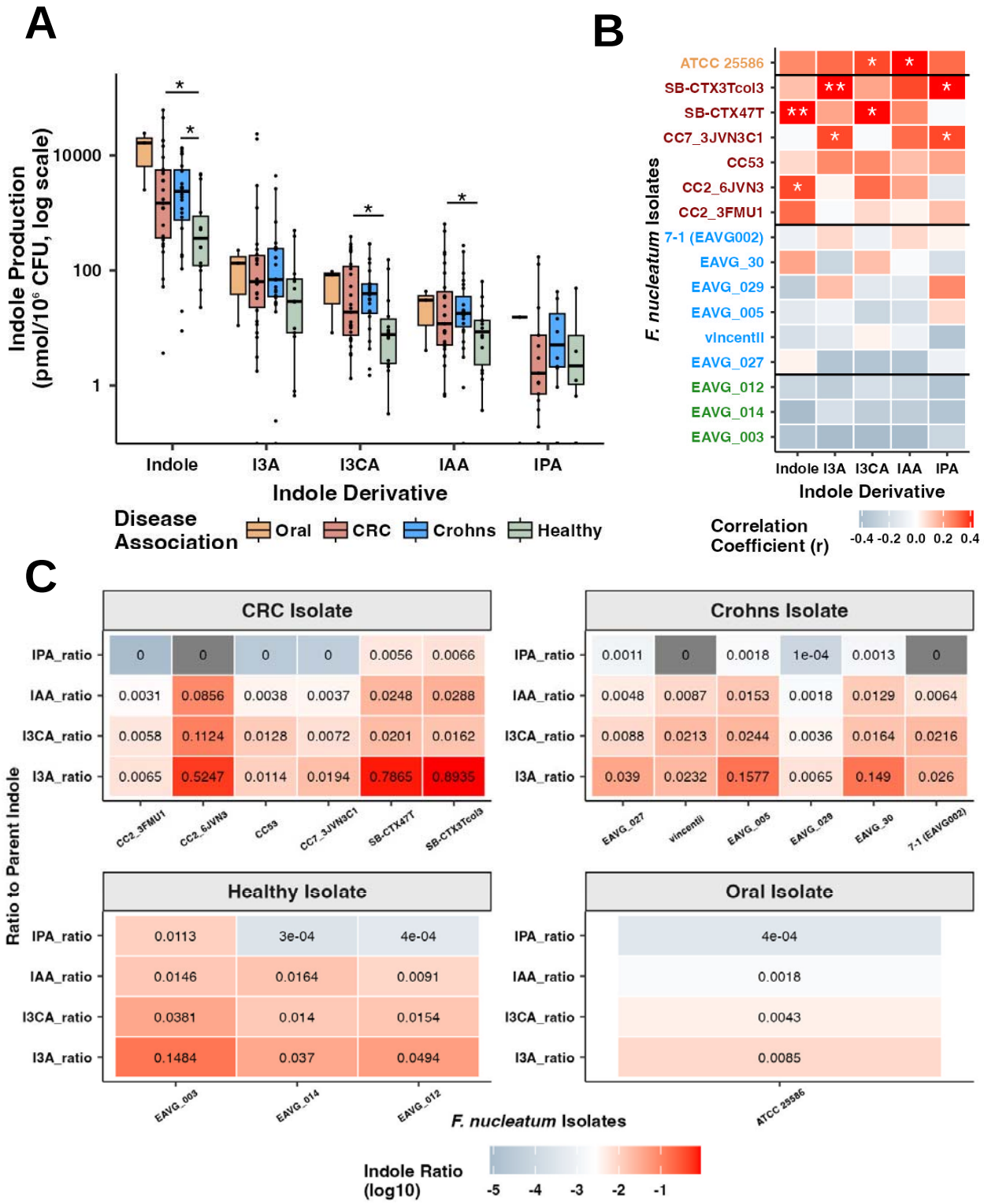

A
Indole Production
(pmol/10⁶ CFU, log scale)
10000
100
1
Indole
I3A
I3CA
IAA
IPA
Indole Derivative
Disease Association
Oral
CRC
Crohns
Healthy
B
ATCC 25586
SB-CTX3Tcol3
SB-CTX47T
CC7_3JVN3C1
CC53
CC2_6JVN3
CC2_3FMU1
7-1 (EAVG002)
EAVG_30
EAVG_029
EAVG_005
vincentII
EAVG_027
EAVG_012
EAVG_014
EAVG_003
F. nucleatum Isolates
Indole I3A I3CA IAA IPA
Indole Derivative
Correlation Coefficient (r)
-0.4 -0.2 0.0 0.2 0.4
C
CRC Isolate
IPA_ratio 0 0 0 0 0.0056 0.0066
IAA_ratio 0.0031 0.0856 0.0038 0.0037 0.0248 0.0288
I3CA_ratio 0.0058 0.1124 0.0128 0.0072 0.0201 0.0162
I3A_ratio 0.0065 0.5247 0.0114 0.0194 0.7865 0.8935
CC2_3FMU1 CC2_6JVN3 CC53 CC7_3JVN3C1 SB-CTX47T SB-CTX3Tcol3
Crohns Isolate
IPA_ratio 0.0011 0 0.0018 1e-04 0.0013 0
IAA_ratio 0.0048 0.0087 0.0153 0.0018 0.0129 0.0064
I3CA_ratio 0.0088 0.0213 0.0244 0.0036 0.0164 0.0216
I3A_ratio 0.039 0.0232 0.1577 0.0065 0.149 0.026
EAVG_027 vincentII EAVG_005 EAVG_029 EAVG_30 7-1 (EAVG002)
Healthy Isolate
IPA_ratio 0.0113 3e-04 4e-04
IAA_ratio 0.0146 0.0164 0.0091
I3CA_ratio 0.0381 0.014 0.0154
I3A_ratio 0.1484 0.037 0.0494
EAVG_003 EAVG_014 EAVG_012
Oral Isolate
IPA_ratio 4e-04
IAA_ratio 0.0018
I3CA_ratio 0.0043
I3A_ratio 0.0085
ATCC 25586
Ratio to Parent Indole
F. nucleatum Isolates
Indole Ratio (log10)
-5 -4 -3 -2 -1

Figure 2: Disease-associated isolates exhibit elevated production of specific indole derivatives and have distinct indole production profiles. (2A) 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 compounds - 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. **(2B)** *F. nucleatum* isolates were transformed into binary valures, and a Point-Biserial correlation analysis was performed. The resulting heatmap depicts Pearson correlation coefficients (ρ) between *F. nucleatum* isolates and the production 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).

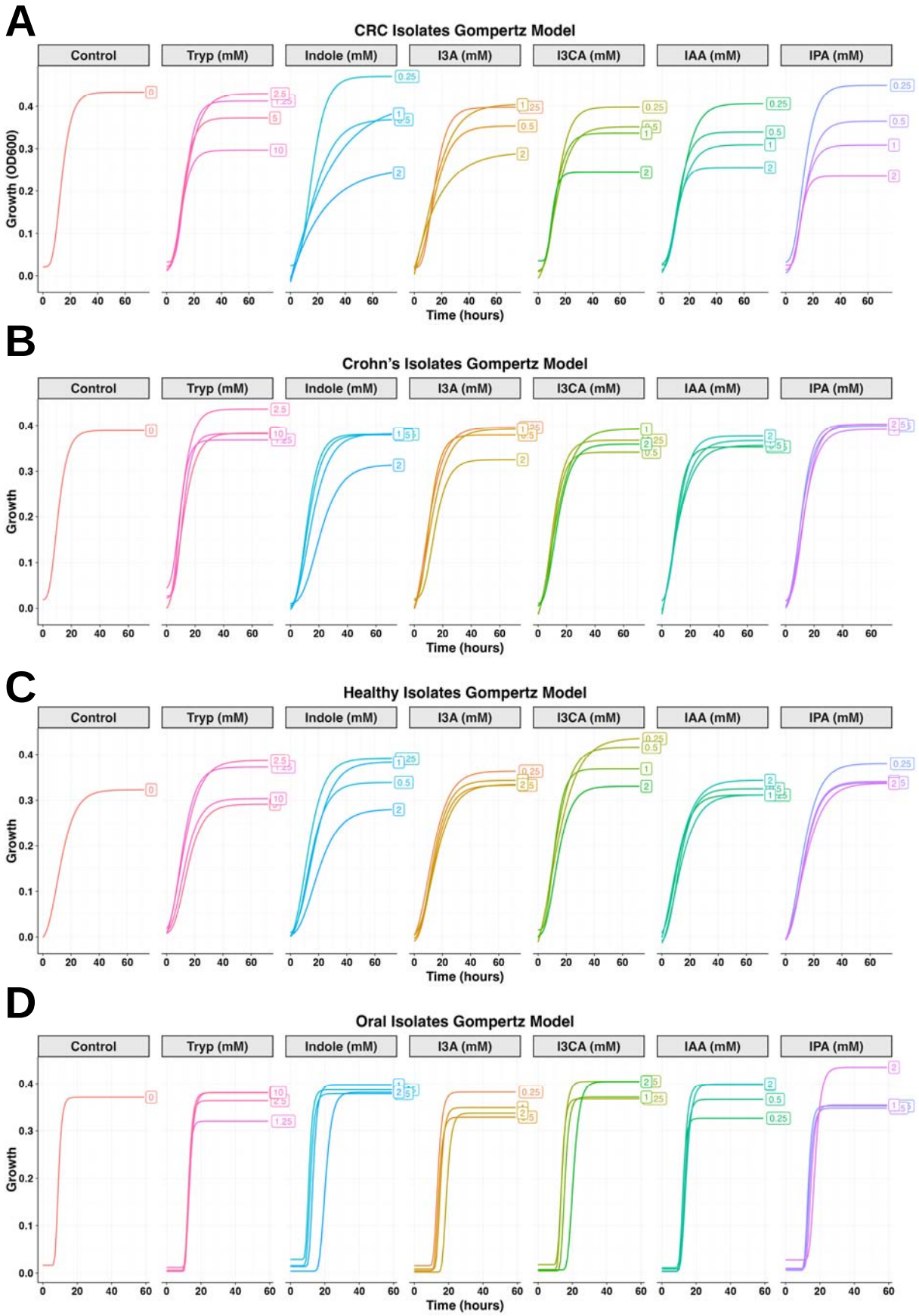
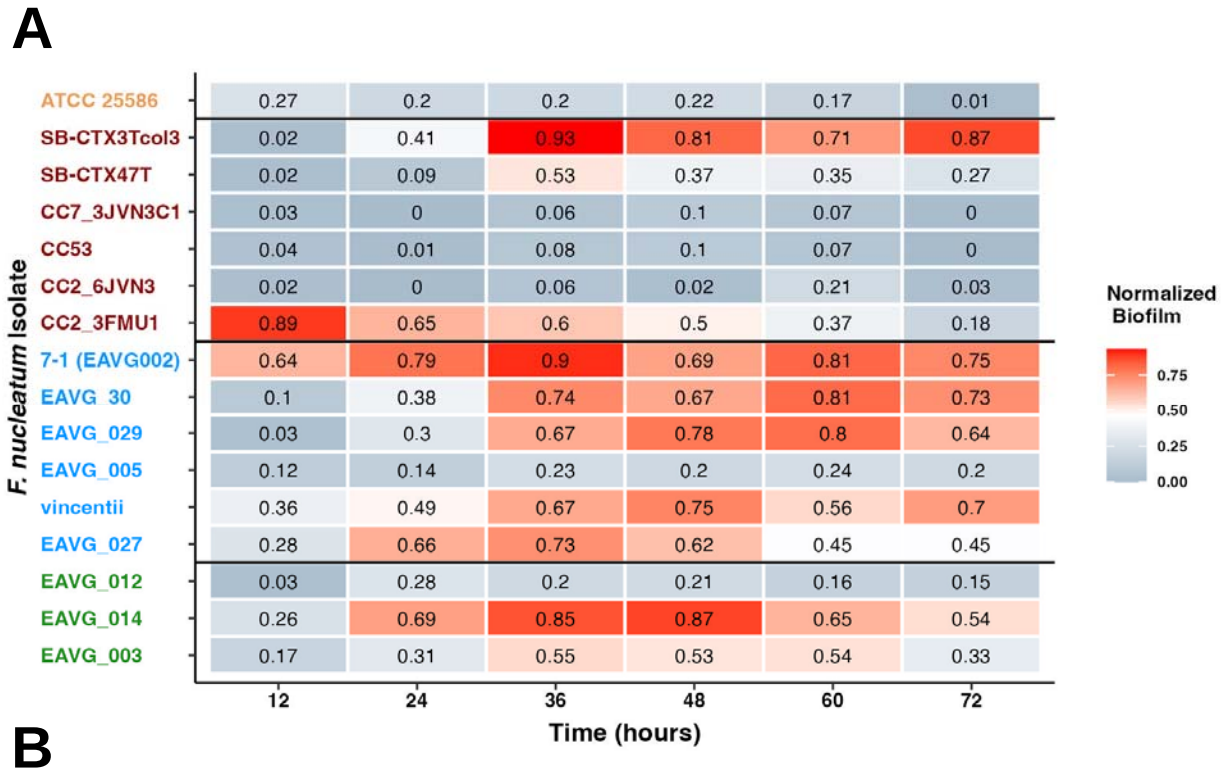
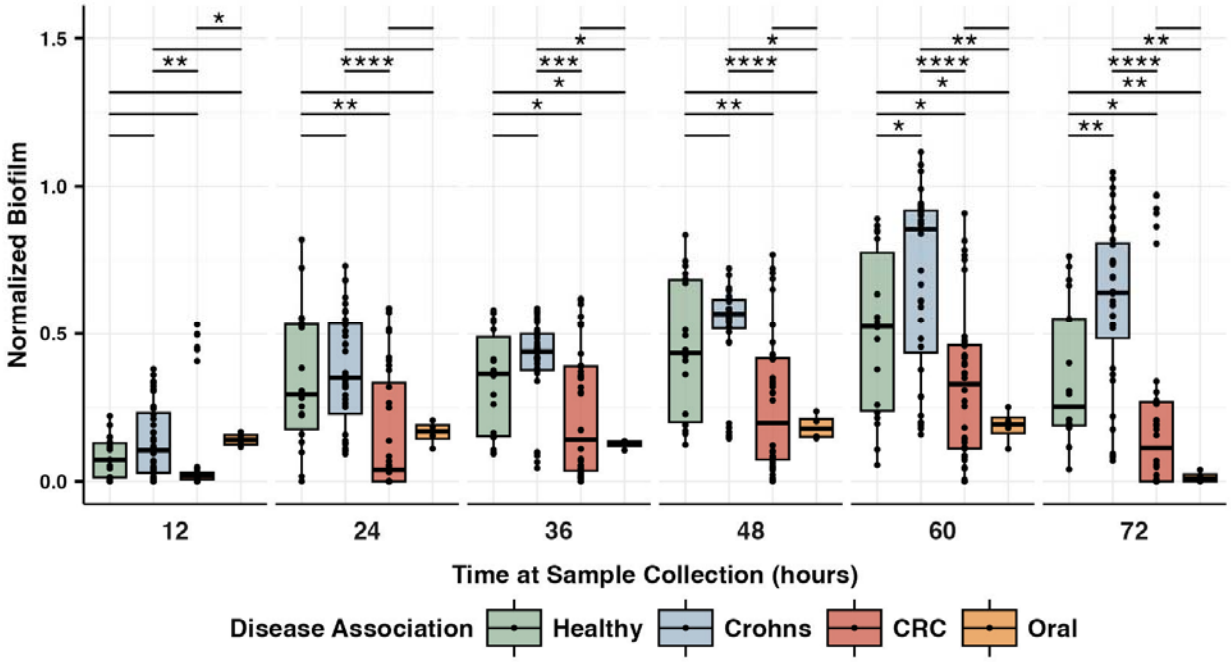

A
CRC Isolates Gompertz Model
Control
Tryp (mM)
Indole (mM)
I3A (mM)
I3CA (mM)
IAA (mM)
IPA (mM)
Growth (OD600)
Time (hours)
B
Crohn's Isolates Gompertz Model
Growth
Time (hours)
C
Healthy Isolates Gompertz Model
Growth
Time (hours)
D
Oral Isolates Gompertz Model
Growth
Time (hours)

Figure 3 Exposure to indole derivatives alter *F. nucleatum* growth kinetics in a dose-dependent manner without bactericidal effects. Gompertz growth models comparing bacterial growth responses to indole derivatives across CRC **(3A)**, Crohn's **(3B)**, Healthy **(3C)**, and Oral isolates **(3D)**. Growth (OD) versus time curves show responses to Indole, I3A, I3CA, IAA, and IPA (all in mM). Chloramphenicol (µg/mL) are used as negative controls to display strong inhibition of growth (**Supplemental Figure 3**). Tryptophan was included in these growth experiments to determine whether the bacteria's ability to convert tryptophan to indole compounds 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).



934



935
936
937

938

939

Figure 4: Biofilm Formation Profiles Distinguish Disease-Associated *F. nucleatum* Isolates. We evaluated the temporal biofilm formation patterns of bacterial isolates to identify high biofilm-producing isolates. **(4A)** Biofilm formation was quantified at six time points (12, 24, 36, 48, 60, and 72 hours) across four study groups (Healthy, Crohn's, CRC, and Oral), with measurements normalized to the maximum biofilm value at each time point. An adjusted ANOVA was performed using the Benjamini-Hochberg method to compare total indole production between grouped *F. nucleatum* isolate. The Tukey's HSD test was then applied to identify significant pairwise differences. Asterisks indicate statistical significance (* $p<0.05$, ** $p<0.01$, *** $p<0.001$, **** $p<0.0001$). **(4B)** Heatmap visualizing temporal biofilm formation patterns across Fusobacterium isolates (Fuso1-25), with normalized values represented by a white-to-red color gradient. The intensity of red coloration corresponds to biofilm strength, and numerical values indicate the proportion of maximum biofilm formation observed at that time point. Isolates grouped by disease association (CRC: n=6, Crohn's: n=6, healthy: n=3, oral lesion: n=1).

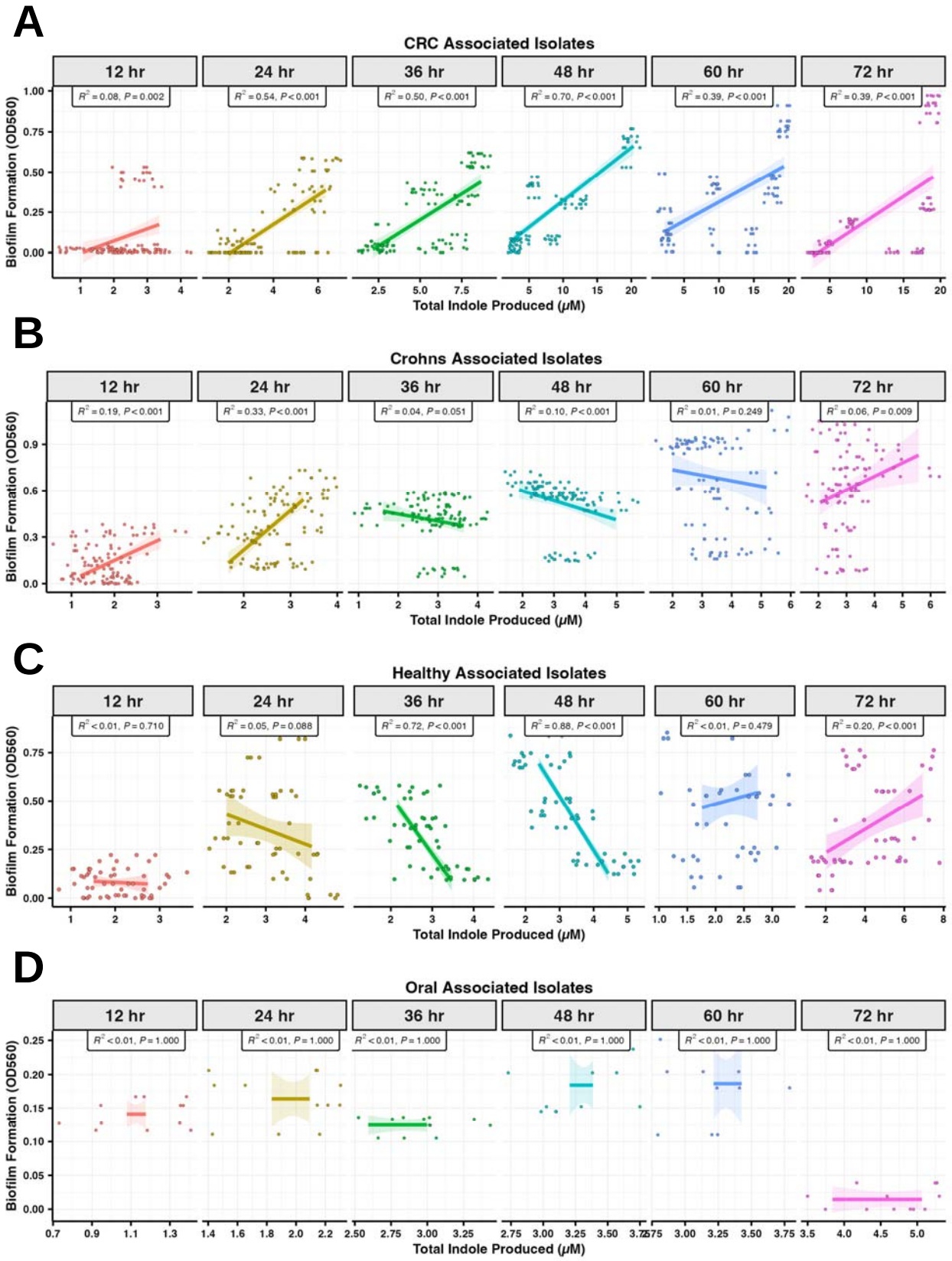

A
CRC Associated Isolates
12 hr
24 hr
36 hr
48 hr
60 hr
72 hr
$R^2 = 0.08$, $P = 0.002$
$R^2 = 0.54$, $P < 0.001$
$R^2 = 0.50$, $P < 0.001$
$R^2 = 0.70$, $P < 0.001$
$R^2 = 0.39$, $P < 0.001$
$R^2 = 0.39$, $P < 0.001$
Biofilm Formation (OD560)
Total Indole Produced (μM)
B
Crohns Associated Isolates
$R^2 = 0.19$, $P < 0.001$
$R^2 = 0.33$, $P < 0.001$
$R^2 = 0.04$, $P = 0.051$
$R^2 = 0.10$, $P < 0.001$
$R^2 = 0.01$, $P = 0.249$
$R^2 = 0.06$, $P = 0.009$
Biofilm Formation (OD560)
Total Indole Produced (μM)
C
Healthy Associated Isolates
$R^2 < 0.01$, $P = 0.710$
$R^2 = 0.05$, $P = 0.088$
$R^2 = 0.72$, $P < 0.001$
$R^2 = 0.88$, $P < 0.001$
$R^2 < 0.01$, $P = 0.479$
$R^2 = 0.20$, $P < 0.001$
Biofilm Formation (OD560)
Total Indole Produced (μM)
D
Oral Associated Isolates
$R^2 < 0.01$, $P = 1.000$
$R^2 < 0.01$, $P = 1.000$
$R^2 < 0.01$, $P = 1.000$
$R^2 < 0.01$, $P = 1.000$
$R^2 < 0.01$, $P = 1.000$
$R^2 < 0.01$, $P = 1.000$
Biofilm Formation (OD560)
Total Indole Produced (μM)

Figure 5: Biofilm formation by CRC-associated *F. nucleatum* isolates is positively correlated with total indole production. The relationship between total indole production (μM) and biofilm formation (OD_{590}) was analyzed over time in bacterial isolates from different sources: **(5A)** CRC, **(5B)** Crohn's, **(5C)** healthy, and **(5D)** 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).

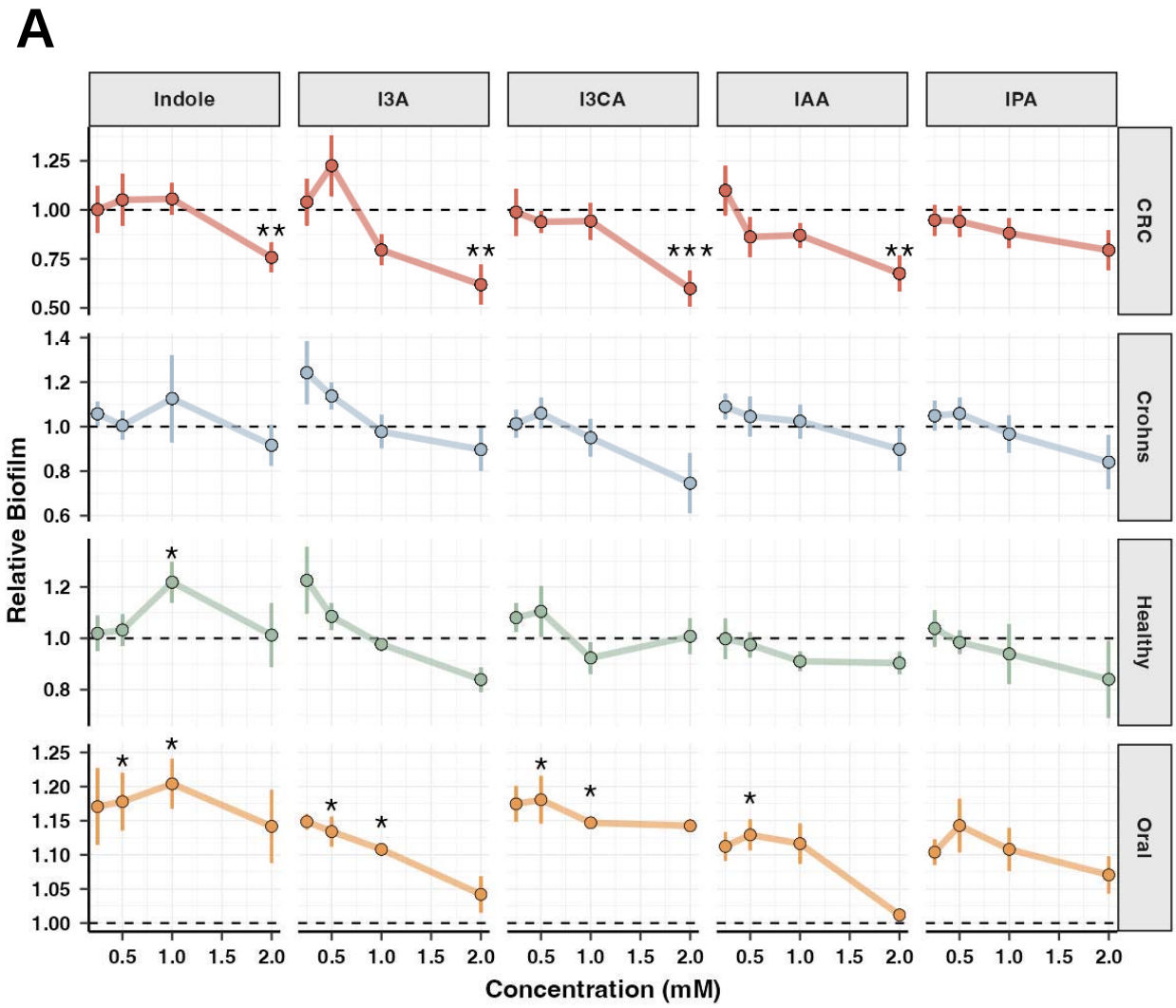
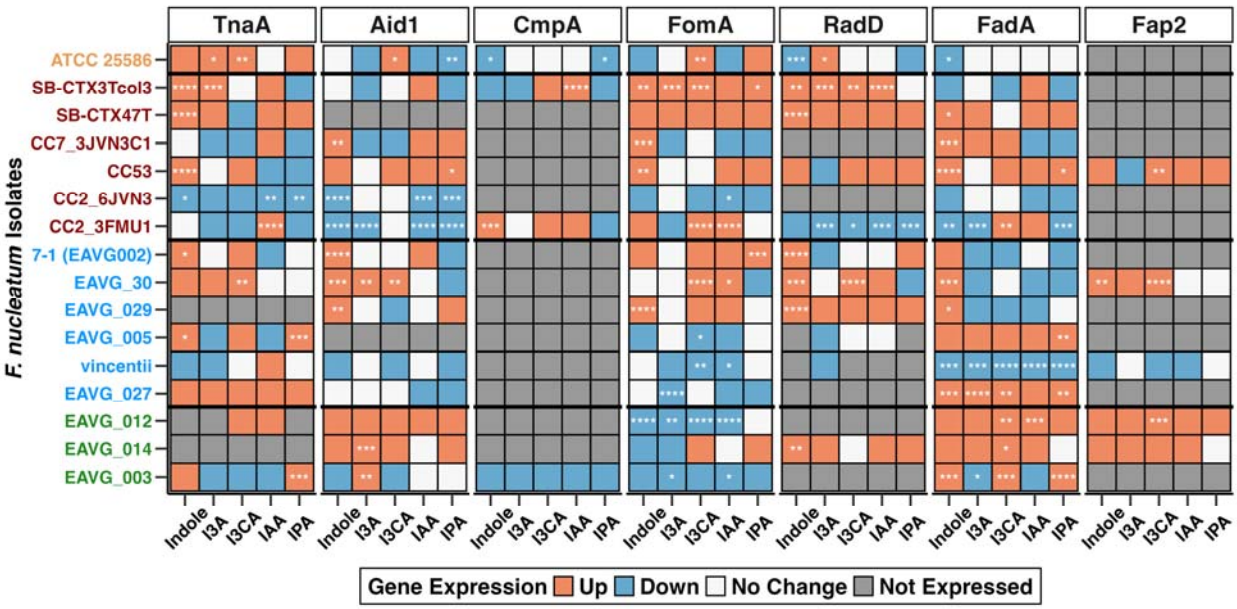
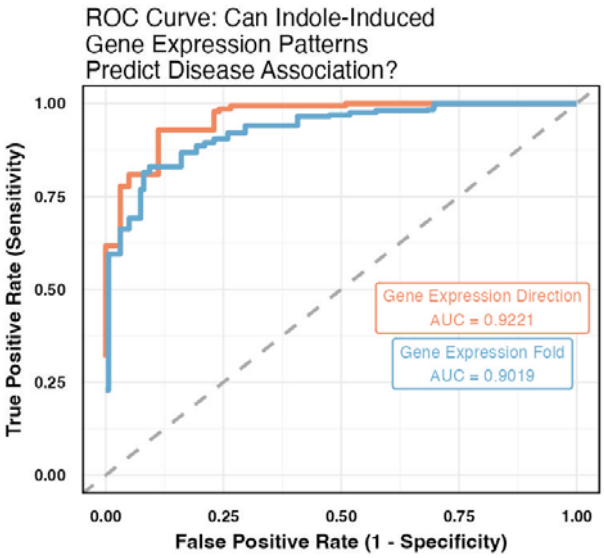


Figure 6: Exposure to Indole Derivatives Significantly Diminish Biofilm Formation in CRC-Associated *F. nucleatum* Isolates: Indole-induced changes in biofilm formation were assessed at concentrations ranging from 0.5 to 2.0 mM. Baseline biofilm formation is indicated by dotted lines, and relative changes from baseline biofilm (relative) across treatment concentration was 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).

A



B



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. (7A) Heatmap shows expression changes in virulence genes after exposure to indole derivatives for each isoalte. 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: (* $p<0.05$, ** $p<0.01$, *** $p<0.001$, **** $p<0.0001$). **(7B)** 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 classifer. 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).

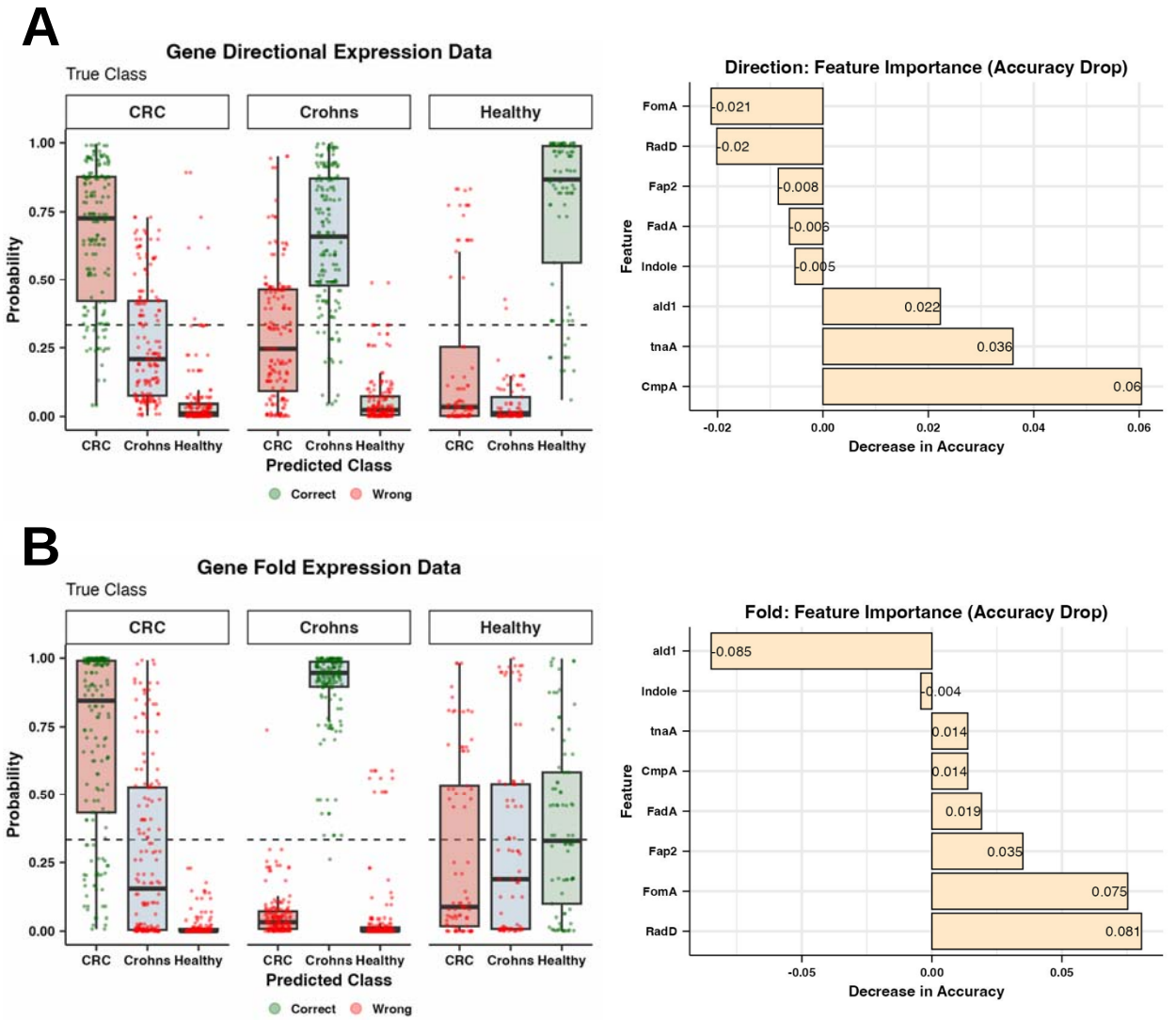


Figure 8: Naive Bayes posterior probabilities and feature importance reveal class-specific predictive confidence based on gene expression direction and fold change. Posterior class probabilities from a Naive Bayes classifier trained on gene expression direction (left) or fold change values (right) across three disease states: CRC, Crohn's disease, and healthy controls **(8A)**. Each boxplot represents the predicted probability distribution for each class, stratified by the true class (facet labels). Dots represent individual predictions, colored by classification accuracy (green = correct, red = wrong). The dashed horizontal line indicates the probability of random classification (1/3). Feature importance plots showing the decrease in model accuracy when each gene was removed **(8B)**. Left: Direction-based model. Right: Fold change-based model. Bars represent the mean change in classification accuracy across cross-validation folds. Isolates grouped by disease association (CRC: n=6, Crohn's: n=6, healthy: n=3).

1111

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

STRAIN	ORIGIN	DISEASE	SUBSPECIES	ISOLATE SUPPLIERS
F. nucleatum animalis EAVG_003	Nonspecific colon biopsy	**Healthy**	*animalis*	Dr. Emma Allen-Vercoe, University of Guelph
F. nucleatum EAVG_014	Rectum biopsy	**Healthy**	*vincentii*	Dr. Emma Allen-Vercoe, University of Guelph
F. nucleatum EAVG_012	Ascending colon biopsy	**Healthy**	*vincentii*	Dr. Emma Allen-Vercoe, University of Guelph
F. nucleatum EAVG_027	Distal colon biopsy	**Crohn's (remission)**	*animalis*	Dr. Emma Allen-Vercoe, University of Guelph
F. nucleatum subsp. vincentii	Distal colon biopsy	**Crohn's (remission)**	*vincentii*	Dr. Emma Allen-Vercoe, University of Guelph

1112

Table 2: *F. nucleatum* Virulence Factor and Subspecies Primers

PRIMER	SEQUENCE			TARGET	REFERENCES
F. nucleatum EAVG_005		Sigmoid colon biopsy	Crohn's (remission)	*nucleatum*	Dr. Emma Allen-Vercoe, University of Guelph
FadA FWD	GCTAGACAAGCACTAGCACAA			Fusobacterium adhesin A (FadA)	
FadA REV	CTTGTGTTAAGGTTCTGCTTGAAGT				
F. nucleatum EAVG_029		Ascending colon biopsy	Crohn's (active)	*vincentii*	Dr. Emma Allen-Vercoe, University of Guelph
Fap2 FWD	GGAACAAGGAATCAAGTGGTAAA			Fap2	
Fap2 REV	GCTTTGTCTCCTCTTTTATATTC				
F. nucleatum EAVG_030		Ascending colon biopsy	Crohn's (active)	*vincentii*	Dr. Emma Allen-Vercoe, In house University of Guelph
tnaA FWD	CGTGCATTAAAAGCAACTCACTATAATG			Tryptophanase enzyme	
tnaA REV	CAGTAGTTCCCGAGTCTGATAAC				
F. nucleatum subsp. nucleatum CC2_3FMU1		Tumor biopsy	Colorectal cancer	*nucleatum*	Dr. Emma Allen-Vercoe, University of Guelph
aid1 FWD	TACAGGAGGTGCCGTAGCAG			aid1 (Adherence Inducing Determinant 1)	
aid1 REV	TTTTTGTTAATTCTCCAGCTCCA				
F. nucleatum subsp. animalis CC2_6JVN3		Tumor biopsy	Colorectal cancer	*animalis*	Dr. Emma Allen-Vercoe, University of Guelph [4]
16s FWD	TTGGACAATGGACCGAGAGT			16S rRNA	
16s REV	GCCGTCACTTCTTTCTGTTGG				
F. nucleatum subsp. vincentii CC53		Tumor biopsy	Colorectal cancer	*vincentii*	Dr. Emma Allen-Vercoe, University of Guelph
F. nucleatum subsp. animalis CC7_3JVN3C1		Tumor biopsy	Colorectal cancer	*animalis*	Dr. Emma Allen-Vercoe, University of Guelph
F. nucleatum (SB-CTX47T)		Tumor biopsy	Colorectal cancer	*polymorphum*	Dr. Susan Bullman, MD Anderson Cancer Center
F. nucleatum (SB-CTX3Tcol3)		Tumor biopsy	Colorectal cancer	*nucleatum*	Dr. Susan Bullman, MD Anderson Cancer Center
F. nucleatum EAVG002-7_1-HM_73		Ascending colon biopsy	Crohn's (active)	*animalis*	Dr. Emma Allen-Vercoe, University of Guelph
F. nucleatum subsp. nucleatum Knorr 25586		ATCC	Oral Lesion	*nucleatum*	ATCC

CmpA FWD	TTGGGATCAAGGAAAACATCAATTAGG	coaggregation mediating protein A (CmpA)	[30]
CmpA REV	ATAATTCCTTTATTATCTCCCATATAAGCAATACC		
RadD FWD	GGATTTATCTTTGCTAATTGGGGAAATTATAG	Fn1526 (now designated as radD)	
RadD REV	ACTATTCCATATTCTCCATAATATTTCCCATTAGA		
FomA FWD	AGAGTTTGATCCTGGCTCAG	outer membrane protein porin (FomA)	
FomA REV	GTCATCGTGCACACAGAATTGCTG		
Subsp. animalis FWD	TTTGTATTTGACAATTTAATTAGAGCA	Subspecies-specific PCR: Fn. animalis hypothetical protein gene CDS, Forwardprimer	[31]
Subsp. animalis REV	TGGGATAATAATCCAAATATYGCAC		
Subsp. nucleatum FWD	CTTCATTGTTAGGATTACAAGAATAGGC	Subspecies-specific PCR: Fn. nucleatum Ankyrin repeat protein gene CDS, Reverse primer:	
Subsp. nucleatum REV	TTTCAGTAACATTTATTGGGAGGTTT		
Subsp. polymorphum FWD	AAGATTAAATAAAGATGGAAATAACCATRAA	Subspecies-specific PCR: Fn. polymorphum hypothetical protein gene CDS, Reverse primer:	
Subsp. polymorphum REV	CCATCTCCAAAGCCTTTCC		
Subsp. vincentii FWD	CTTTGTCTTATGATTTATTGCRTATCTTC	Subspecies-specific PCR: Fn. vincentii hypothetical protein gene CDS, Reverse primer:	
Subsp. vincentii REV	GAAGAATCAGAAGCACTACCAGCA		

Table 3: 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 4: 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 5: 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 6: 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