Materials and Methods

Ethics Statement

This study was approved by the Johns Hopkins Institutional Review Board and the Medical Institutional Review Board and UMMC Medical Ethics Committee at the University of Malaya. All samples were obtained in accordance with the Health Insurance Portability and Accountability Act (HIPAA).

Patient Selection and Sample Acquisition

Colon tumors (adenomas and cancers) and paired normal tissues were collected from patients undergoing surgery at Johns Hopkins Hospital or the University of Malaya Medical Centre in Kuala Lumpur, Malaysia. All tissue not needed for pathologic diagnosis was rapidly preserved in formalin, Carnoy's solution and/or RNAlater (Qiagen Inc., Germantown, MD) for analysis. Patients who received pre-operative radiation and/or chemotherapy or with a personal history of CRC were excluded. For patients in this study, two mechanical bowel preparations were routinely used and recorded (polyethylene glycol [Miralax™ or Fortrans™] or Fleet Phospho-soda™ enema [PE]). The proportion of individuals who received polyethylene glycol (PEG) vs. PE use prior to surgery was identical in the biofilm positive and negative groups. Pre-operative intravenous antibiotics were administered in all cases (cefotetan, clindamycin/gentamicin or cefoperazone/metronidazole). No patient received pre-operative oral antibiotics. Dietary information was not available.

Healthy control patients undergoing screening colonoscopy or colonoscopy for diagnostic work-up (eg, anemia) at Johns Hopkins Hospital were enrolled. All patients underwent a standard mechanical bowel preparation. Mucosal biopsies from grossly normal colon were taken from the right (cecum or ascending) and left (descending or sigmoid) colon during the colonoscopy. All tissue was rapidly preserved in formalin, Carnoy's solution and/or RNAlater for analysis. Patients who had a personal history of CRC, inflammatory bowel disease or were treated with antibiotics within the past three months were excluded.

Analysis of Johns Hopkins Hospital and University of Malaya Medical Centre Samples

Fluorescent in situ hybridization (FISH)

Fixed, paraffin-embedded tissues were sectioned to 5 μm thickness and de-waxed following standard procedures. Sections were stained with Periodic acid Schiff (PAS) to confirm mucus presence and preservation and successive sections were hybridized with the Eub338 universal bacterial probe and with a nonsense probe to test for nonspecific binding of probes. Slides were imaged using a Nikon E800 microscope with NIS elements software or Zeiss LSM 510 META laser scanning microscope with LSM imaging software (for confocal imaging). Paired images are presented at identical exposure intensities.

Oligonucleotide probes were synthesized and conjugated at the 5' end to the fluorophores listed in Table S4 (Life Technologies).

Probes were applied to slides at a concentration of 2 pmol/ul of each probe in prewarmed hybridization buffer (900 mM NaCl, 20 mM Tris pH 7.5, 0.01% SDS, 20% formamide). Slides were incubated at 46°C in a humid chamber for 2 hours, and washed at 48°C for 15 minutes in wash buffer (215 mM NaCl, 20 mM Tris pH 7.5, 5 mM EDTA). Slides were dipped in water, then in 100% ethanol, air-dried, and coverslips were mounted using ProLongGold antifade reagent (Life Technologies).

Analysis of Johns Hopkins Hospital Samples

Biofilm Bacterial Quantification

Biofilm bacterial density and depth were measured using slides hybridized with the universal bacterial probe, Eub338, and imaged at 1000x magnification with a Nikon E800 microscope and Nikon NIS elements viewing software.

Measures of bacterial density were based on the following model. A 10x10 μm square placed over a region of a 5 μm thick tissue section (500 μm^3) constitutes a volume of $5x10^{-10}$ ml. One bacterium in this volume is equivalent to 2×10^9 bacteria/ml. The visual distinction of a single bacterium is lost but spaces can still be seen between the bacteria when 250 bacteria occupy a 10x10 μm space; these cases were

assigned a concentration of 10^{11} bacteria/ml. A solid mat of bacteria with no discernible spaces between the bacteria constitutes an increase to 2500 bacteria in a 10x10 μm space; these cases were assigned a concentration of 10^{12} bacteria/ml (1) . The mean of five (10x10 μm) fields was used to determine bacterial density.

The biofilm depth was measured using ImageJ software calibrated with an image of a stage micrometer from the same microscope and magnification used in the images being quantified. Biofilm depth was calculated as the mean of five measurements taken along a 200 μm span of the biofilm.

Scanning Electron Microscopy (SEM)

Tissue samples were fixed in 2% glutaraldehyde, 2% paraformaldehyde in 0.1 M sodium cacodylate (NaCaco), 3 mM CaCl, 1% sucrose pH 7.4 overnight with gentle rocking. Samples were rinsed three times in washing buffer (0.1 M NaCaco, 3 mM CaCl, 3% sucrose), and placed in 1% osmium tetroxide in 1 M NaCaco for 1 hour in the dark. Samples were rinsed twice in distilled water followed by dehydration in an ethanol series. Samples were next placed in a 1:1 mixture of 100% ethanol to hexamethyldisilazane (HMDS) for two washes of 10 minutes each. This was followed by three washes with 100% HMDS for five minutes each. Samples were then removed and placed in a vacuum desiccant overnight followed by gold palladium coating before viewing under a Leo Zeiss Field emission SEM. Samples were scored by two independent observers (CMD, CLS) for biofilm presence and morphologies.

Fluorescence Spectral Imaging and Unmixing

Samples that were determined to have a bacterial presence by universal probe were next analyzed by fluorescence spectral imaging as described above (see Fluorescent in situ hybridization Methods) using 9 probes simultaneously, targeting broad phylogenetic groups and subgroups (Table S4) (2-12) .

Spectral images were acquired with a Zeiss LSM 780 laser scanning confocal microscope with a 32-channel GaAsP detector and Zeiss ZEN software. All images were acquired with a Zeiss Plan-Apochromat 40x/1.4 NA(420762-9900) objective; 2x line averaging, 2048x2048px frame size, 1.58 μs

pixel dwell time; and 8.7 nm spectral resolution. Five fields of view were selected per sample. Spectral images of each field of view were acquired sequentially with six different lasers proceeding from long to short excitation wavelength: HeNe633 (633nm), HeNe594 (594nm), DPSS561-10 (561nm), Ar514 (514nm), Ar488 (488nm), and Diode 405-30 (405nm).

FISH probe reference spectra were measured from spectral images of pure populations of cultured bacterial cells singly labeled with the appropriate taxon-specific FISH probe. Tissue autofluorescence reference spectra were measured from spectral images of tissue subjected to the FISH procedure but without probe, and imaged under experimental imaging conditions.

Linear unmixing was performed with a custom Mathematica script using a least squares method. Each spectral image was unmixed independently using the appropriate reference spectra for the excitation wavelength. For each field of view, unmixed channels for each FISH probe were extracted from the unmixing results corresponding to the appropriate excitation wavelength. Extracted unmixed channels were compiled and colorized in ImageJ using the Image5D plugin.

Sample Preparation for Sequencing

Mucosal samples from surgically-removed tumors, paired surgical normal tissues and colonoscopy biopsies were collected in the pathology or endoscopy suites at Johns Hopkins Hospital and immediately placed in RNAlater (Qiagen Inc. Germantown, MD) and stored at -80^{o}C. Tissue samples (100-500 mg) were placed in a 15 ml conical tube with 2.5 ml Qiagen buffer ASL. Samples were incubated at 95^{o}C for 15 minutes with frequent vortexing to remove bacteria from the epithelial surface. Following the dislodging of mucosal associated bacteria, 1.4 ml of supernatant was removed and cells were thoroughly lysed using a Barocycler NEP2320 (Pressure Biosciences, Inc. South Easton, MA), by cycling between atmospheric pressure, 0 psi to 25,000 psi while maintaining a temperature of 60^{o}C. Following pressure lysis, DNA was extracted using the QIAamp DNA Stool Kit (Qiagen). Recovered genomic DNAs were quantitated using a Nanodrop spectrophotometer (Bio-Rad Life Science Research, Hercules, CA). The V3-V5 region of bacterial 16S rDNA was amplified and sequenced following the procedures described by

the Human Microbiome Project standard protocol (http://www.hmpdacc.org/doc/16S_Sequencing_SOP_4.2.2.pdf). Briefly, the V3-V5 region of 16S rDNA was amplified with PCR primers (357F 5' CCTACGGGAGGCAGCAG 3' and 926R 5' CCGTCAATTCMTTTRAGT 3') that were appended with Roche 454 Titanium FLX library adapter sequences. All B-adapter primers were identical, while A-adapter primers also contained a unique barcode of 5-10 nucleotides to allow indexing of individual samples. Each sample was PCR amplified for 30 cycles with Phusion HF DNA polymerase (New England Biolabs Inc. Ipswich, MA). PCR products were purified by gel electrophoresis. Bands of the appropriate size were excised from the gel and purified using the QIAquick Gel Extraction Kit (Qiagen Inc. Germantown, MD). Purified DNAs were quantified using the 454 FLX Library Quantification Kit (KAPA Biosystems Inc. Woburn, MA) and pooled for sequencing in equal molar quantity.

Sequence data analysis

Raw sequence reads were initially assigned to samples based on multiplex identifier barcodes, trimmed of forward and reverse primer sequences, and filtered for quality and length (minimum 150 bp) using the QIIME package (v1.6.0) (13-16) . High quality reads were then organized by sample and error-corrected using the Acacia tool (v1.52) (17) , and subsequently screened for chimeras utilizing *de novo* UCHIME (v4.2.40) (18) . Chloroplast DNA was identified and removed using the RDP Bayesian classifier (v2.5) (19) .

The final high-quality contaminant-free dataset was then submitted to the CloVR-16S pipeline (v1.1) (20) for diversity estimation, taxonomic characterization and comparative analysis of sample groups of interest. Sequences were clustered *de novo* into species-level OTUs using UCLUST (21) with a 95% identity threshold (22, 23). Taxonomic assignment of OTU representatives was performed using the RDP classifier with a minimum threshold of 0.5. There was no exclusion removal of low frequency OTUs (e.g., singletons). Pipeline runs were executed using CloVR (v2012.11.16) on the DIAG academic cloud (http://diagcomputing.org).

Immunohistochemistry and Immunofluorescence

Immunohistochemistry was performed on tissue fixed in 10% formalin and paraffin-embedded following standard procedures. Sections were de-paraffinized and rehydrated through a xylene, ethanol-water gradient. Ki67 staining was performed on an automated immunostainer Benchmark ST Staining System using detection reagents from the iView DAB detection kit (Roche) with Ki67 monoclonal antibody (clone 30-9, Roche). Antibody for phospho-Stat3 (clone Tyr705 D3A7, Cell Signaling Technology) was applied at 1:400 dilution overnight following antigen retrieval. The TUNEL assay was performed per the manufacturer's instructions using the fluorescein *in situ* cell death detection kit (Roche). Immunofluorescent staining was performed following standard procedures. Antibodies for Il-6 (ab6672, Abcam), E-cadherin (Mouse IgG2a, BD), and smooth muscle antigen (SMA) (clone 1A4, mouse monoclonal, Sigma) were used at 1:400, 1:100, 1:500 dilutions respectively. Enzymatic antigen retrieval was performed prior to IL-6 antibody application (15 minutes proteinase K treatment at 37°C); citrate-based antigen retrieval was performed prior to E-cadherin and SMA (microwave boiling in antigen unmasking solution (Vector Laboratories)). All antibodies were applied overnight at 4°C, followed by secondary antibody application for one hour at room temperature. All slides were treated with DAPI for 10 minutes and mounted with glass coverslips using Prolong Gold antifade reagent (Life Technologies).

Quantification of Stat3, Proliferation and Apoptosis

Phosphorylated Stat3 (pStat3) was semi-quantitatively assessed using a 0-+3 grading scale in which epithelial cell and immune cell pStat3 are individually assessed (24) .

A total of 5 well-oriented crypts were selected from each sample to be scored for Ki67+ cells by two blinded individuals (CMD,ECW). Positive cells were counted on both sides of each crypt starting at the base and ending at the luminal surface in increments of 15 cells. Each interval was scored as cells positive per 15 cells. The mean number of proliferating cells within each interval was calculated for each analyzed sample, groups were compared using the nonparametric Mann Whitney U test.

Apoptosis scoring was performed by two independent observers (CMD, ECW). TUNEL positive cells were counted per 1000 epithelial cells in 10 randomly selected fields. Results were graphed as percent positive and groups were compared using the nonparametric Mann Whitney U test.

Immunofluorescence quantification of IL-6 and E cadherin

Pixel intensity per area was measured from selected cell populations (differentiated surface epithelium, crypt epithelium, lamina propria or whole tissue) and background fluorescence was subtracted using imageJ. For differentiated surface epithelial quantification, five measurements were taken from 400x images along 100μm linear distance of surface epithelium (for a total of 500 linear μm measured per sample). Crypt epithelial measurements were collected from the base of five crypts. Total lamina propria fluorescence was quantified in three distinct, representative 400x fields per specimen. Total tissue fluorescence was measured as total fluorescence intensity upon selection of the entire tissue specimen at 200x on at least two tissue sections per specimen. Individual specimen fluorescence intensity values were calculated as the mean of the individual measurements made for each specimen.

Isolation of colonic epithelial cells

Colonic epithelial cells were isolated using a modified rapid low-temperature method (25). Briefly, approximately 500 mg of epithelial tissue was washed with ice-cold PBS and divided into 2-3 mm fragments before transferring to chelating buffer (27 mM trisodium citrate, 5mM Na_2PO4, 8mM KH_2PO4, 1.5mM KCl, 0.5 mM DTT, 55mM D-sorbitol, 44 mM sucrose, 6mM EDTA, 5mM EGTA, pH 7.3) for 45 minutes at 4°C. Cells were dissociated by repeated vigorous shaking. After removal of debris, using a 100 μm cell strainer, epithelial cells were collected by centrifugation at 150 g for 10 minutes and stored at -80°C until protein extraction.

IL-6 ELISA

Colonic epithelial cell pellets were lysed in cell extraction buffer (Life Technologies) for 30 minutes on ice with vortexing at 10 minute intervals. Cellular debris was removed by centrifugation at 13,000 rpm for 10 minutes. One microgram of total protein from the clear lysate was used in the IL-6 ELISA (Sigma) according to the vendor's protocol.

Statistical analysis

Prior to downstream statistical analysis, sequence data were subsampled to equivalent depths (2500 sequences per sample) (15, 26) . Unweighted UniFrac distances (27) , and principal coordinate analysis plots were computed in QIIME. Additional statistical analyses were performed in R (v2.15.1). The nonparametric Mann Whitney U test was used for all presented analyses except for single instances where the Fisher's exact test and Spearman correlation were used as as appropriate and noted in the text. Pairwise beta-diversity comparisons utilized the nonparametric Mann Whitney U test.

Supplemental References

1. Swidsinski A, Weber J, Loening-Baucke V, Hale LP & Lochs H (2005) Spatial organization and composition of the mucosal flora in patients with inflammatory bowel disease. *J Clin Microbiol* 43(7): 3380-3389.

2. Amann RI, *et al* (1990) Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. *Appl Environ Microbiol* 56(6): 1919-1925.

3. Valm AM, *et al* (2011) Systems-level analysis of microbial community organization through combinatorial labeling and spectral imaging. *Proc Natl Acad Sci U S A* 108(10): 4152-4157.

4. Diaz PI, *et al* (2006) Molecular characterization of subject-specific oral microflora during initial colonization of enamel. *Appl Environ Microbiol* 72(4): 2837-2848.

5. Weller R, Glockner FO & Amann R (2000) 16S rRNA-targeted oligonucleotide probes for the in situ detection of members of the phylum cytophaga-flavobacterium-bacteroides. *Syst Appl Microbiol* 23(1): 107-114.

6. Kong Y, He M, McAlister T, Seviour R & Forster R (2010) Quantitative fluorescence in situ hybridization of microbial communities in the rumens of cattle fed different diets. *Appl Environ Microbiol* 76(20): 6933-6938.

7. Kempf VA, Trebesius K & Autenrieth IB (2000) Fluorescent in situ hybridization allows rapid identification of microorganisms in blood cultures. *J Clin Microbiol* 38(2): 830-838.

8. Neef A, Amann R & Schleifer K (1995) Detection of microbial cells in aerosols using nucleic acid probes. *Syst Appl Microbiol* 18(1): 113-122.

9. Wallner G, Amann R & Beisker W (1993) Optimizing fluorescent in situ hybridization with rRNA-targeted oligonucleotide probes for flow cytometric identification of microorganisms. *Cytometry* 14(2): 136-143.

10. Rigottier-Gois L, Rochet V, Garrec N, Suau A & Dore J (2003) Enumeration of bacteroides species in human faeces by fluorescent in situ hybridisation combined with flow cytometry using 16S rRNA probes. *Syst Appl Microbiol* 26(1): 110-118.

11. Paster BJ, Bartoszyk IM, Dewhirst FE (1998) Identification of oral streptococci using PCR-based, reverse-capture, checkerboard hybridization. *Methods Cell Sci* 20: 223.

12. Manz W, Amann R, Ludwig W, Wagner M & Schleifer K (1992) Phylogenetic oligodeoxynucleotide probes for the major subclasses of proteobacteria: Problems and solutions. *Syst Appl Microbiol* 15(4): 593-600.

13. Kuczynski J, *et al* (2012) Using QIIME to analyze 16S rRNA gene sequences from microbial communities. *Curr Protoc Microbiol* Chapter 1: Unit 1E.5.

14. Caporaso JG, *et al* (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 7(5): 335-336.

15. Muegge BD, *et al* (2011) Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans. *Science* 332(6032): 970-974.

16. Koren O, *et al* (2012) Host remodeling of the gut microbiome and metabolic changes during pregnancy. *Cell* 150(3): 470-480.

17. Bragg L, Stone G, Imelfort M, Hugenholtz P & Tyson GW (2012) Fast, accurate error-correction of amplicon pyrosequences using acacia. *Nat Methods* 9(5): 425-426.

18. Edgar RC, Haas BJ, Clemente JC, Quince C & Knight R (2011) UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 27(16): 2194-2200.

19. Wang Q, Garrity GM, Tiedje JM & Cole JR (2007) Naive bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* 73(16): 5261-5267.

20. Angiuoli SV, *et al* (2011) CloVR: A virtual machine for automated and portable sequence analysis from the desktop using cloud computing. *BMC Bioinformatics* 12: 356-2105-12-356.

21. Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26(19): 2460-2461.

22. Chen W, Liu F, Ling Z, Tong X & Xiang C (2012) Human intestinal lumen and mucosa-associated microbiota in patients with colorectal cancer. *PLoS One* 7(6): e39743.

23. White JR, *et al* (2010) Alignment and clustering of phylogenetic markers--implications for microbial diversity studies. *BMC Bioinformatics* 11: 152-2105-11-152.

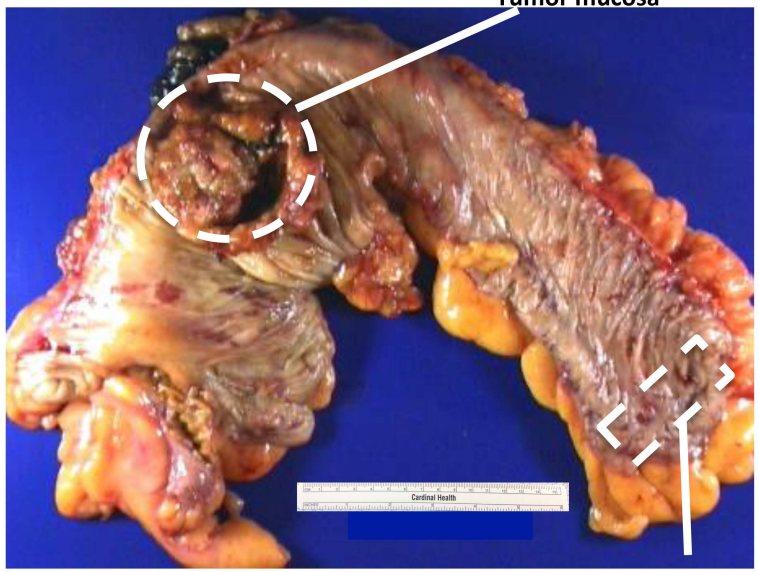
24. Wick EC, *et al* (2012) Shift from pStat6 to pStat3 predominance is associated with inflammatory bowel disease-associated dysplasia. *Inflamm Bowel Dis* 18(7): 1267-1274.

25. Flint N, Cove FL & Evans GS (1991) A low-temperature method for the isolation of small-intestinal epithelium along the crypt-villus axis. *Biochem J* 280 (Pt 2)(Pt 2): 331-334.

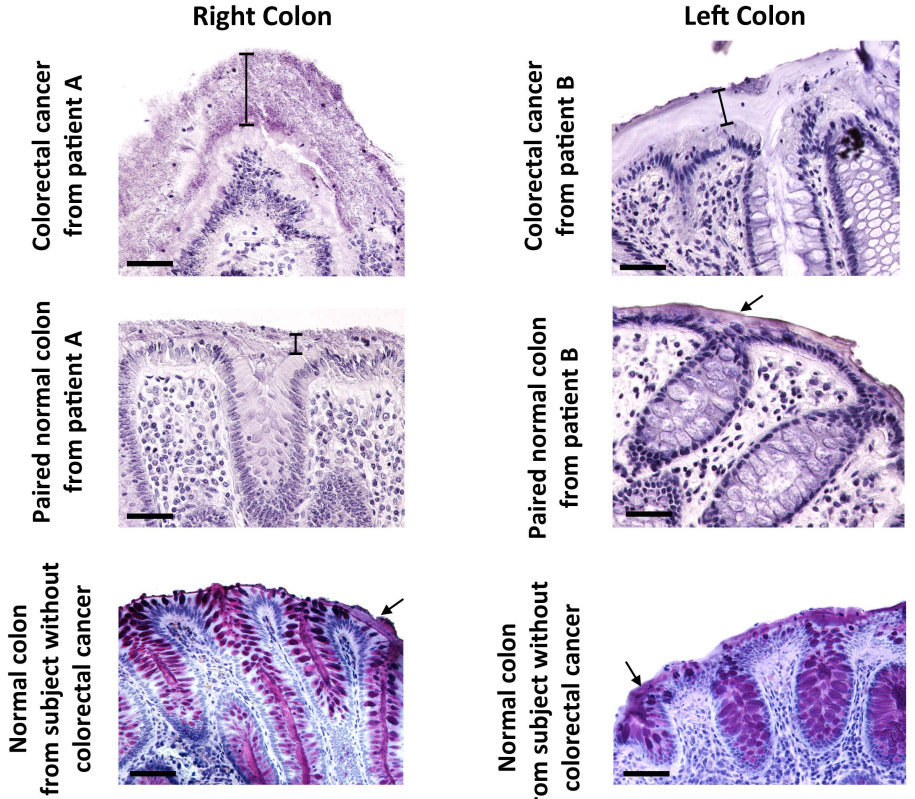
26. Lozupone CA, *et al* (2013) Meta-analyses of studies of the human microbiota. *Genome Res* 23(10): 1704-1714.

27. Lozupone C, Lladser ME, Knights D, Stombaugh J & Knight R (2011) UniFrac: An effective distance metric for microbial community comparison. *ISME J* 5(2): 169-172.

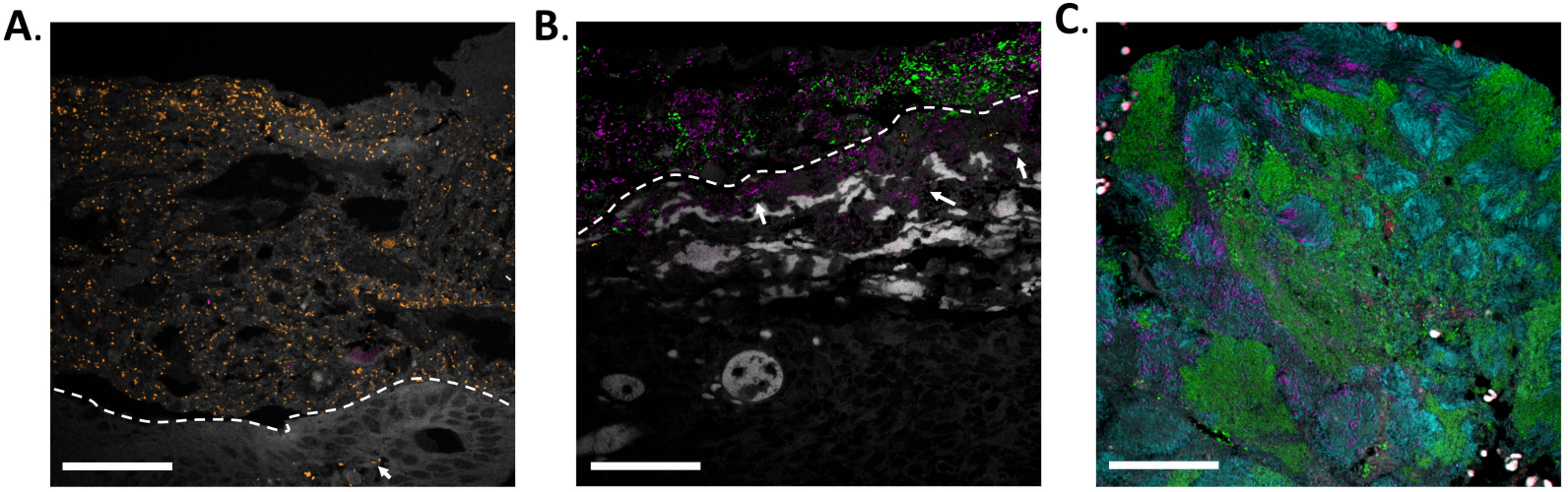
Supplementary Figure 1.



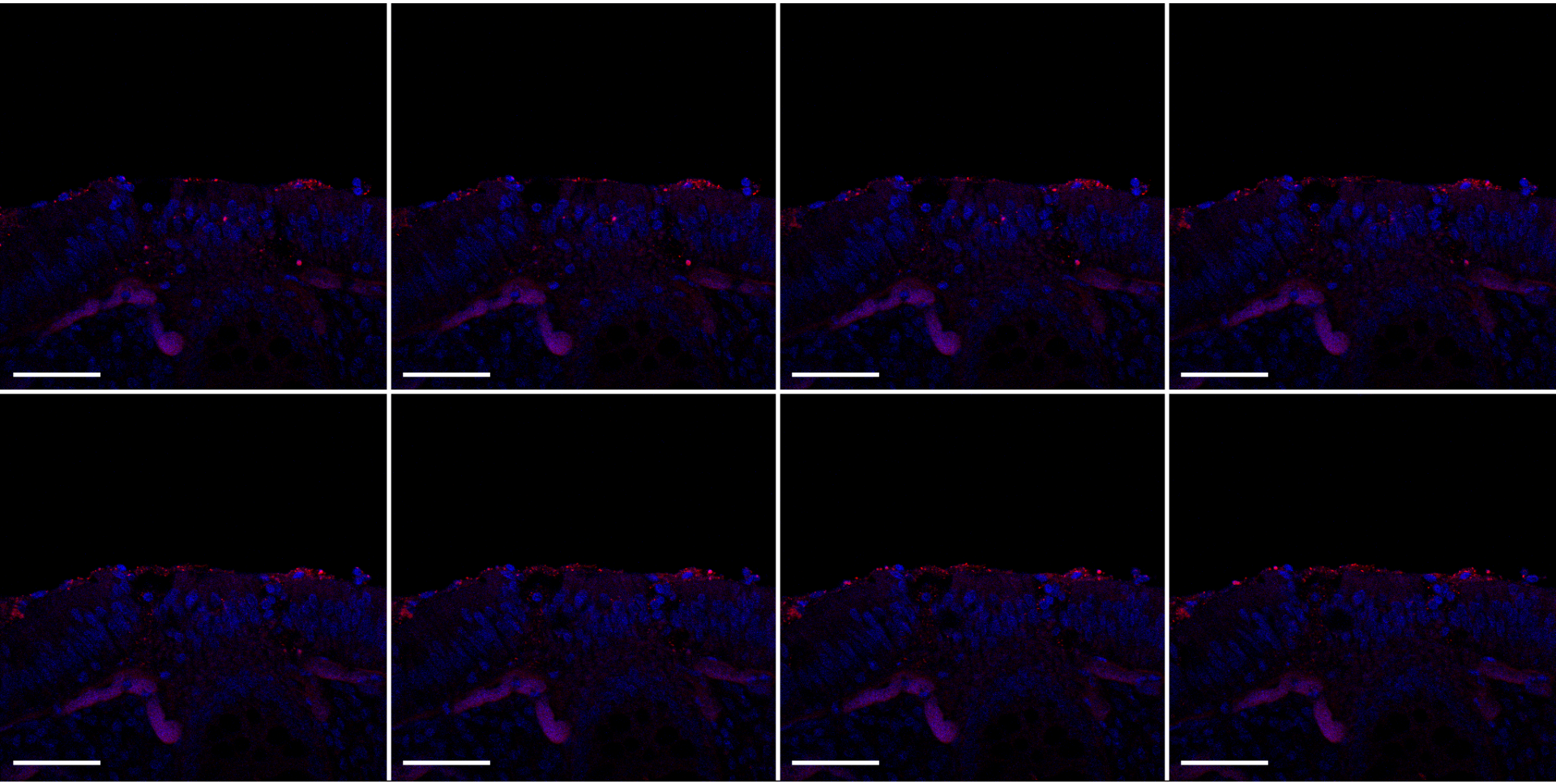
Supplementary Figure 2.



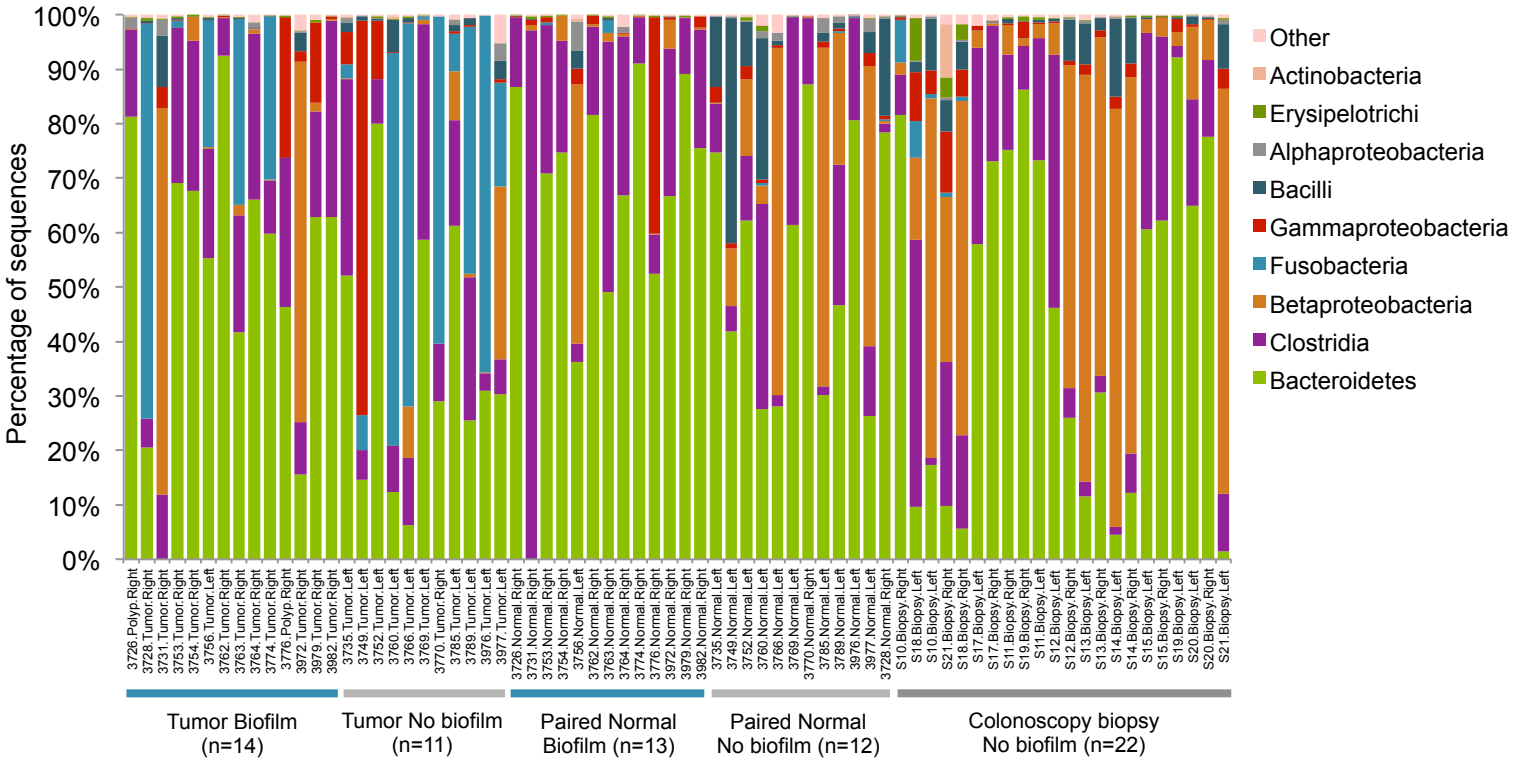
Supplementary Figure 3.



Supplementary Figure 4.

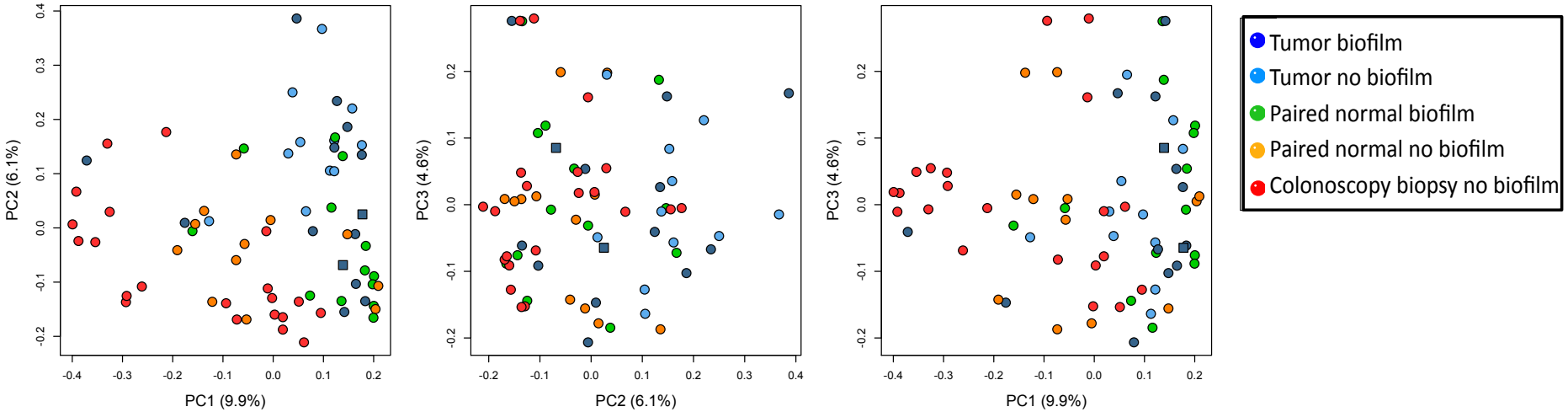


Supplementary Figure 5.

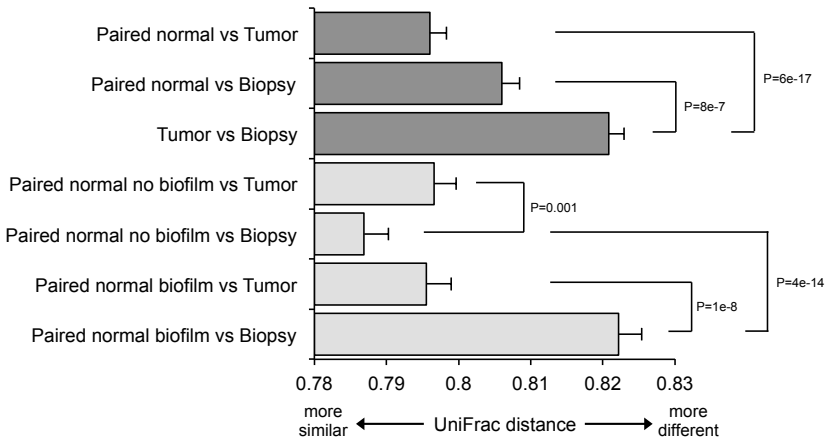


Supplementary Figure 6.

A.

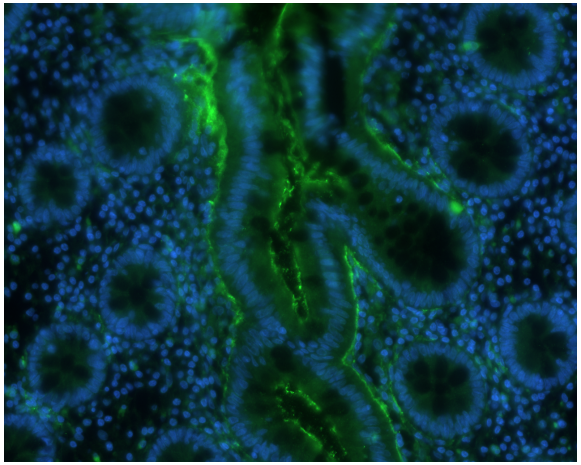


B.

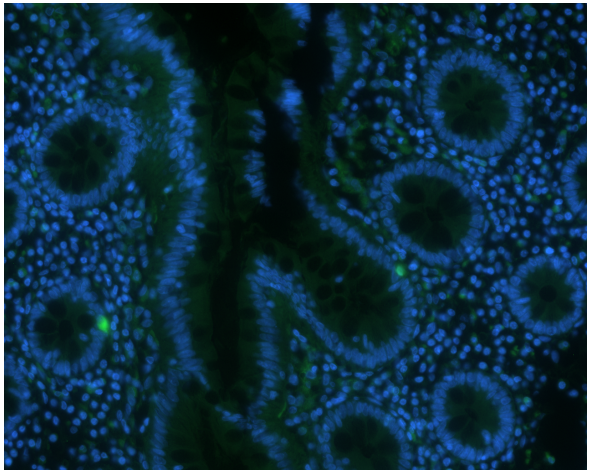


Supplementary Figure 7.

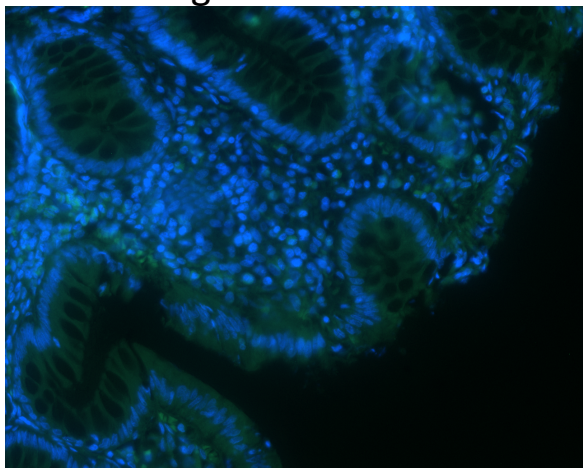
A. IL-6 staining



B. IgG negative control

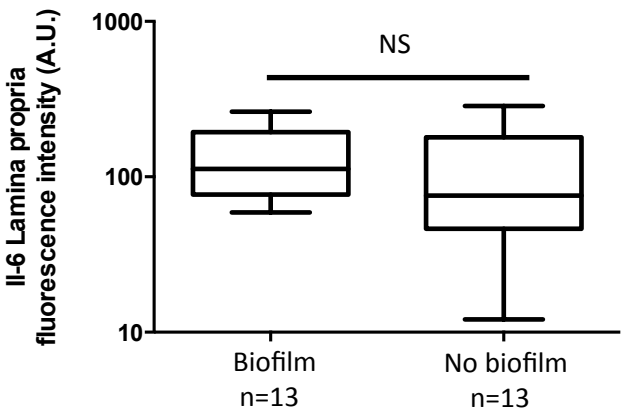


C. Secondary antibody only negative control



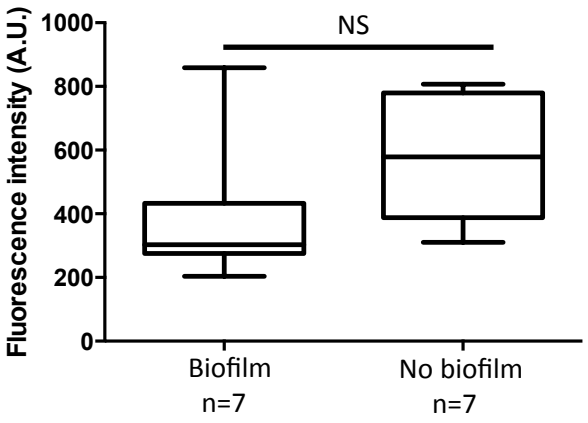
Supplementary Figure 8.

Lamina Propria IL-6
normal tissue from patients with CRC



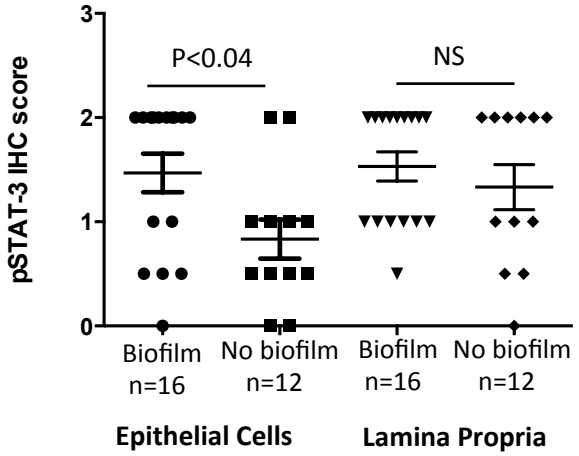
Supplementary Figure 9.

Differentiated epithelial cell E-cadherin
normal tissue from patients with CRC

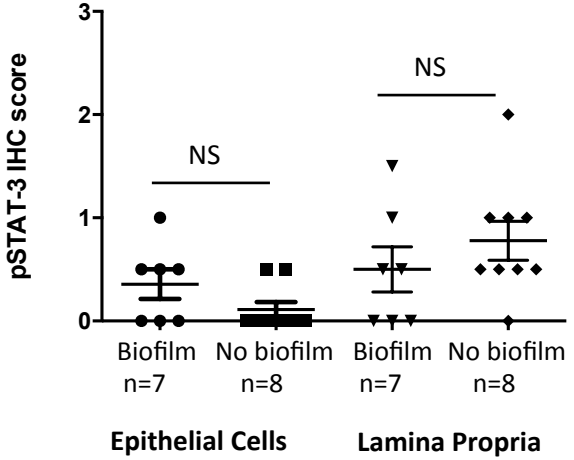


Supplementary Figure 10.

A. Paired normal tissue from patients with CRC



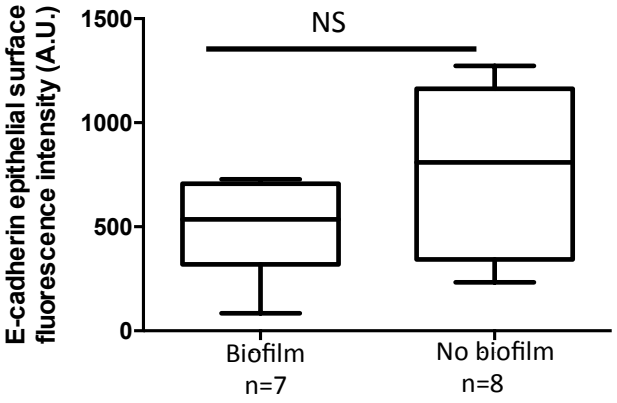
B. Colonoscopy biopsies from subjects without CRC



Supplementary Figure 11.

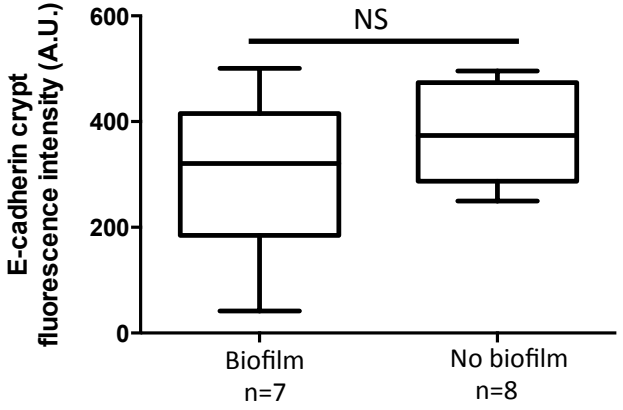
A. Differentiated Epithelial Cell E-cadherin

Colonoscopy biopsies from subjects without CRC



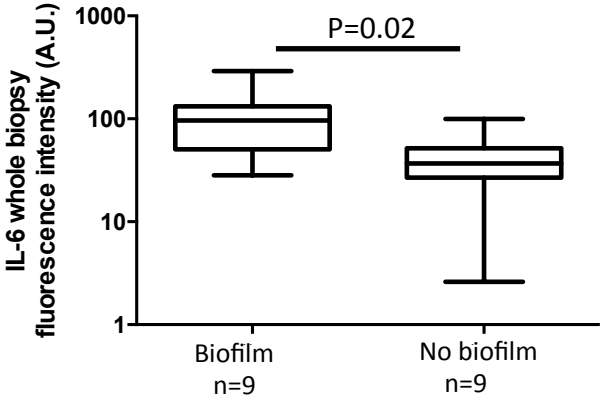
B. Crypt E-cadherin

Colonoscopy biopsies from subjects without CRC

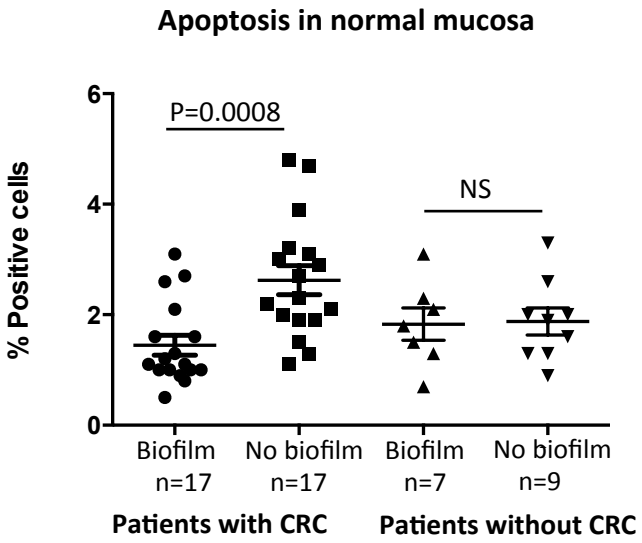


C. Whole biopsy IL-6

Colonoscopy biopsies from subjects without CRC



Supplementary Figure 12.



Supplementary Table 1.

Surgical CRC and Polyp metadata

Patient ID	Patient Type	Age	Sex	Race	Tumor Site	Biofilm	Stage	Size (mm)	Histology
3972	Surgical CRC	78	M	Caucasian	Cecum	Yes	3	40.0	Adenocarcinoma
3979	Surgical CRC	77	F	African American	Cecum	Yes	3	35.0	Adenocarcinoma
3726	Surgical Polyp	50	M	Caucasian	Ascending	Yes	NA	40.0	Tubular adenoma-no dysplasia
3728	Surgical CRC	69	M	Caucasian	Ascending	Yes	1	87.0	Adenocarcinoma
3731	Surgical CRC	74	M	Caucasian	Ascending	Yes	2	30.0	Adenocarcinoma
3741	Surgical CRC	64	M	Caucasian	Ascending	Yes	1	8.0	Adenocarcinoma
3753	Surgical CRC	49	F	African American	Ascending	Yes	4	47.0	Mucinous Adenocarcinoma
3754	Surgical CRC	67	F	African American	Ascending	Yes	2	30.0	Adenocarcinoma
3762	Surgical CRC	73	M	Caucasian	Ascending	Yes	4	54.0	Adenocarcinoma
3763	Surgical CRC	66	F	Caucasian	Ascending	Yes	2	30.0	Adenocarcinoma
3764	Surgical CRC	59	F	Caucasian	Ascending	Yes	4	22.0	Adenocarcinoma
3776	Surgical Polyp	84	F	Caucasian	Ascending	Yes	NA	35.0	Tubular adenoma-no dysplasia
3982	Surgical CRC	62	M	Caucasian	Ascending	Yes	2	45.0	Mucinous Adenocarcinoma
3984 A/B*	Surgical Polyp/ Surgical Polyp	47	M	Caucasian	Ascending/ Ascending	Yes/Yes	NA	33.0/8.0	Tubular adenoma-no dysplasia/ Tubular adenoma-no dysplasia
3987	Surgical CRC	66	F	Caucasian	Ascending	Yes	2	50.0	Mucinous Adenocarcinoma
3986	Surgical CRC	59	M	Caucasian	Hepatic Flexure	No	2	50.0	Adenocarcinoma
3770	Surgical CRC	71	M	Caucasian	Hepatic Flexure	No	1	35.0	Adenocarcinoma
3774	Surgical CRC	45	M	Asian	Hepatic Flexure	Yes	2	45.0	Adenocarcinoma
3752	Surgical CRC	73	F	Caucasian	Transverse	No	2	25.0	Adenocarcinoma
3976	Surgical CRC	52	F	Caucasian	Transverse	No	1	20.0	Mucinous Adenocarcinoma
4017 A/B*	Surgical Polyp/ Surgical Polyp	64	F	African American	Transverse/ Rectosigmoid	No/No	NA	30.0/60.0	Tubulovillous adenoma-no dysplasia/ Tubulovillus adenoma-dysplasia
3769	Surgical CRC	78	F	African American	Splenic Flexure	No	3	60.0	Adenocarcinoma
3992	Surgical CRC	91	F	Caucasian	Splenic Flexure	Yes	2	45.0	Adenocarcinoma
3789	Surgical CRC	55	M	Hispanic	Descending	No	3	50.0	Adenocarcinoma
3988	Surgical CRC	48	M	Caucasian	Descending	No	1	35.0	Mucinous Adenocarcinoma
3749	Surgical CRC	39	M	Caucasian	Sigmoid	No	3	50.0	Adenocarcinoma
3756	Surgical CRC	54	M	Caucasian	Sigmoid	Yes	4	45.0	Adenocarcinoma
3766	Surgical CRC	56	F	Caucasian	Sigmoid	No	4	55.0	Adenocarcinoma
3977	Surgical CRC	38	F	Caucasian	Sigmoid	No	1	50.0	Adenocarcinoma
3760	Surgical CRC	29	F	Caucasian	Rectosigmoid	No	2	80.0	Adenocarcinoma
3785	Surgical CRC	54	F	Caucasian	Rectosigmoid	No	3	40.0	Adenocarcinoma
4009	Surgical CRC	53	M	Caucasian	Rectosigmoid	No	3	86.0	Adenocarcinoma
3735	Surgical CRC	64	M	Caucasian	Rectum	No	3	70.0	Adenocarcinoma
3978	Surgical CRC	90	F	Caucasian	Rectum	No	1	27.0	Mucinous Adenocarcinoma

*A/B-two adenomas were analyzed from a single individual

Supplementary Table 2.

Surgical CRC and Polyp metadata (Malaysia)

Patient ID	Patient Type	Age	Sex	Race	Tumor Site	Biofilm	Stage	Size (mm)	Histology
S005	Surgical CRC	60	F	Malay	Cecum	Yes	3	ND	Adenocarcinoma
S016	Surgical CRC	78	M	Chinese	Cecum	Yes	3	50	Adenocarcinoma
S003	Surgical CRC	68	F	Malay	Ascending	Yes	2	70	Adenocarcinoma
S018	Surgical CRC	73	F	Chinese	Hepatic Flexure	Yes	2	70	Adenocarcinoma
S025	Surgical CRC	54	F	Malay	Transverse	No	2	60	Adenocarcinoma
S019	Surgical CRC	54	F	Other	Splenic Flexure	Yes	4	80	Adenocarcinoma
S021	Surgical CRC	57	F	Malay	Splenic Flexure	No	3	ND	Adenocarcinoma
S009	Surgical CRC	58	F	Chinese	Descending	No	3	60	Adenocarcinoma
S002	Surgical CRC	77	F	Chinese	Sigmoid	No	3	40	Adenocarcinoma
S006	Surgical CRC	79	M	Indian	Sigmoid	No	2	25	Adenocarcinoma
S010	Surgical CRC	58	F	Malay	Sigmoid	No	2	55	Adenocarcinoma
S011	Surgical CRC	67	M	Chinese	Sigmoid	Yes	2	30	Adenocarcinoma
S013	Surgical CRC	55	M	Malay	Sigmoid	No	3	65	Adenocarcinoma
S014	Surgical CRC	70	F	Chinese	Sigmoid	No	3	20	Adenocarcinoma
S015	Surgical Polyp	72	F	Indian	Sigmoid	No	NA	25	Tubulovillous Adenoma
S026	Surgical CRC	78	F	Chinese	Sigmoid	No	2	80	Adenocarcinoma
S020	Surgical CRC	76	M	Chinese	Rectosigmoid	Yes	2	100	Adenocarcinoma
S023	Surgical CRC	85	F	Chinese	Rectosigmoid	No	2	30	Adenocarcinoma
S024	Surgical CRC	77	M	Malay	Rectosigmoid	No	3	80	Adenocarcinoma
S007	Surgical CRC	71	M	Chinese	Rectum	No	2	22	Adenocarcinoma
S012	Surgical CRC	77	M	Indian	Rectum	No	2	50	Adenocarcinoma
S008	Surgical CRC	72	M	Malay	Rectum	Yes	4	42	Adenocarcinoma

Supplementary Table 3.

Patient ID	Age	Sex	Race	Bowel Prep	Biopsy Site	Biofilm
1	51	F	African American	GoLytely Prep	Descending	Yes
					Ascending	No
2*	66	F	Caucasian	GoLytely Prep	Descending	No
					Ascending	No
3	66	M	African American	GoLytely Prep	Descending	Yes
					Ascending	Yes
4*	66	M	African American	GoLytely Prep	Descending	No
					Ascending	No
5*	44	F	Caucasian	GoLytely Prep	Descending	No
					Ascending	No
6	52	F	Caucasian	GoLytely Prep	Descending	No
					Ascending	No
7	74	M	Caucasian	Miralax Prep	Descending	No
					Ascending	No
8	51	F	African American	GoLytely & Miralax	Descending	No
					Ascending	Yes
9	73	M	African American	GoLytely Prep	Descending	No
					Ascending	No
10*	40	F	Caucasian	GoLytely Prep	Descending	No
					Ascending	No
11*	42	F	Hispanic	GoLytely Prep	Descending	No
					Ascending	No
12*	43	F	Caucasian	Miralax Prep	Descending	No
					Ascending	No
13	62	M	African American	GoLytely Prep	Descending	No
					Ascending	No
14*	75	M	Caucasian	GaviLyte Prep	Descending	No
					Ascending	No
15	44	M	Caucasian	Miralax Prep	Descending	No
					Ascending	No
17	71	F	African American	GoLytely Prep	Descending	No
					Ascending	No
18	74	F	African American	NuLytely Prep	Descending	No
					Ascending	No
19	62	M	African American	GoLytely Prep	Descending	No
					Ascending	No
20	59	M	African American	GoLytely Prep	Descending	No
					Ascending	No
21*	49	F	Asian	NuLytely Prep	Descending	No
					Ascending	No
22*	47	M	African American	GoLytely Prep	Descending	No
					Ascending	No
23*	78	F	Caucasian	Mag Citrate & Dulcolax Prep	Descending	No
					Ascending	No
24*	47	M	African American	Miralax Prep	Descending	No
					Ascending	No
25	63	M	African American	GoLytely Prep	Descending	No
					Ascending	No
26*	39	F	African American	Miralax Prep	Descending	No
					Ascending	No
27	51	F	African American	GoLytely Prep	Descending	No
					Ascending	No
28	47	F	Caucasian	Miralax Prep	Descending	No
					Ascending	No
29*	64	F	Caucasian	GoLytely Prep	Descending	No
					Ascending	No
30	51	M	Caucasian	Miralax Prep	Descending	No
					Ascending	No
31*	65	M	African American	Miralax Prep	Descending	No
					Ascending	No
32*	58	F	African American	Miralax Prep	Descending	No
					Ascending	No
33	36	M	Asian	Miralax Prep	Descending	No
					Ascending	No
34	66	F	Caucasian	Miralax Prep	Descending	No
					Ascending	No
35	60	F	Caucasian	Miralax Prep	Descending	No
					Ascending	No
36	79	F	African American	GoLytely Prep	Descending	No
					Ascending	No
37	66	M	Caucasian	Miralax Prep	Descending	No
					Ascending	No
38	59	M	Caucasian	Miralax Prep	Descending	Yes
					Ascending	Yes
39	72	F	African American	Miralax Prep	Descending	Yes
					Ascending	Yes
40	58	F	African American	GoLytely Prep	Descending	No
					Ascending	No
41	46	F	Asian	GoLytely Prep	Descending	Yes
					Ascending	Yes
42*	69	M	African American	Miralax Prep	Descending	Yes
					Ascending	Yes
43	62	F	African American	GoLytely Prep	Descending	No
					Ascending	No
44*	47	F	Caucasian	GoLytely Prep	Descending	No
					Ascending	No
45*	73	M	African American	GoLytely Prep	Descending	No
					Ascending	No
46	63	F	African American	GoLytely Prep	Descending	No
					Ascending	No
47*	66	M	Caucasian	Miralax Prep	Descending	No
					Ascending	No
48	56	F	African American	GaviLyte Prep	Descending	No
					Ascending	No
49	63	F	African American	GoLytely Prep	Descending	No
					Ascending	No
50	64	M	Other	GoLytely Prep	Descending	No
					Ascending	No
52*	43	F	Caucasian	NuLytely Prep	Descending	No
					Ascending	No
53	47	M	Caucasian	GoLytely Prep	Descending	No
					Ascending	No
54	55	M	Caucasian	NuLytely Prep	Descending	No
					Ascending	No
55	60	F	African American	GoLytely Prep	Descending	No
					Ascending	No
56	52	F	African American	CoLyte Prep	Descending	Yes
					Ascending	Yes
57*	52	M	Caucasian	TriLyte Prep	Descending	No
					Ascending	No
58	57	F	African American	Miralax Prep	Descending	No
					Ascending	No
59	77	M	African American	NuLytely & Miralax Prep	Descending	No
					Ascending	No
60*	45	F	Caucasian	TriLyte Prep	Descending	No
					Ascending	No
61	58	M	Caucasian	Miralax Prep	Descending	No
					Ascending	No
62	62	M	African American	GoLytely Prep	Descending	Yes
					Ascending	No

*patient underwent diagnostic workup

Supplementary Table 4.

Probe Target(s)	Probe Name	Fluorophore	Probe Sequence (5'-3')	Reference
Kingdom Bacteria except Planctomycetales and Verrucomicrobia	Eub338	Cy3, Alexa 405	GCTGCCTCCCGTAGGAGT	14
Fusobacteria	Fus714	Alexa 488	GGCTTCCCCATCGGCATT	15
Prevotella, Bacteroides	PRV392	Rhodamine Red X	GCACGCTACTTGGCTGG	16
Bacteroidetes (Bacteroides, Parabacteroides, Prevotella)	CFB286	Alexa 514	TCCTCTCAGAACCCCTAC	17
Betaproteobacteria	Bet42a	Alexa 647	GCCTTCCCACTTCGTTT	24
Gammaproteobacteria	Gam42a	Alexa 647	GCCTTCCCACATCGTTT	24
Lachnospiraceae	Lac435	Texas Red X	TCTTCCCTGCTGATAGA	18
Enterobacteriaceae except *Proteus* spp	Ent186	Alexa 555	CCCCCWCTTTGGTCTTGC	19
Bacteroides fragilis	S-S-Bfrag-998-a-A-20	Alexa 633	GTTTCCACATCATTCCACTG	22
Escherichia coli, *Shigella*	Eco1167	Oregon Green 514	GCATAAGCGTCGCTGCCG	20
Streptococcus	Str405	Rhodamine Red X	TAGCCGTCCCTTTCTGGT	23
nonsense probe	non338	Cy3	ACTCCTACGGGAGGCAGC	21

Supplementary Figure Legends

Figure S1. Example of mucosal tumor and normal tissue sites selected for analyses of surgically resected colons from CRC or adenoma patients.

Figure S2. PAS-stained histopathology images of cancer and normal tissue pairs from Patient A and Patient B as well as the right and left normal colonoscopy biopsies from healthy individuals shown in Figure 1A. The mucus layer of the epithelium of each image, stained by PAS, is delineated by brackets or arrows.

Figure S3. Bacterial biofilms detected on CRCs and adenomas have variable compositions. (**A**) Right adenoma biofilm comprised solely of *Enterobacteriaceae* (orange) and *Lachnospiraceae* (magenta). (**B**) Right CRC biofilm composed of *Bacteroidetes* (green) and *Lachnospiraceae* (magenta). (**C**) Right CRC biofilm composed of *Fusobacteria* (cyan), *Bacteroidetes* (green) and *Lachnospiraceae* (magenta). (Scale bar: 50 μm). Dotted white lines depict margin between bacterial biofilm and tumor tissue; white arrows identify bacteria invading into tumor tissue; S3C image is comprised only of biofilm at this magnification.

Figure S4. A series of eight z-stack slices (through 4 μm) depicting bacterial invasion of normal tissue (epithelial cells and submucosa) from a patient with colorectal cancer (Scale bar: 50 μm).

Figure S5. Individual subject histograms of bacterial classes grouped by tissue type and biofilm status. Tumors comprised of 23 CRCs and 2 adenomas. Paired normal tissues indicate surgically-resected normal mucosa from tumor host. Colonoscopy biopsies indicate normal mucosa biopsies from individuals undergoing screening colonoscopy.

Figure S6. (A) Pairwise display of the first three principal coordinate axes of the PCoA. (B) Unweighted Unifrac distance analysis. Dark-shaded bars display all tissues from tumor hosts (surgical paired normal or tumor) whether biofilm positive (N=13) or negative (N=12) and all colonoscopy biopsies (N=21) evaluated by sequence analysis. Light-shaded bars display a similar analysis subsetted by biofilm status.

Figure S7. Colon mucosal tissue samples showing IL-6 immunofluorescence staining and controls. (A) IL-6 immunofluorescence staining; (B) Parallel section treated with IgG negative control antibody; (C) Sample treated only with secondary antibody

Figure S8. IL-6 quantification by immunofluorescence in lamina propria from biofilm positive or biofilm negative normal surgical colon tissues from patients with CRC. Data displayed as bar and whisker graphs where line designates the median, boxes the 25/75th percentile and whiskers the 95th percentile. A.U., Arbitrary Units; NS, nonsignificant.

Figure S9. Measurement of E-cadherin in differentiated epithelial cells in biofilm positive and biofilm negative normal surgical tissues from patients with CRC. Data displayed as bar and whisker graphs where line designates the median, boxes the 25/75th percentile and whiskers the 95th percentile. A.U., Arbitrary Units; NS, nonsignificant.

Figure S10. Quantification of immunohistochemistry (IHC) for activated Stat3 (pStat3) in biofilm positive or biofilm negative normal colon tissues from patients with CRC (**A**) and biofilm positive and biofilm negative colonoscopy biopsies from subjects without CRC (**B**). Epithelial cell pStat3 was significantly increased in biofilm positive normal tissues from CRC patients. Data are displayed as mean+/-SD. NS, nonsignificant.

Figure S11. Quantification of E-cadherin or IL-6 fluorescence intensity in biofilm positive and biofilm negative colonoscopy biopsies from subjects without CRC. E-cadherin fluorescence

intensity was quantified separately in differentiated surface epithelial cells (**A**) and crypt cells (**B**). Total IL-6 fluorescence intensity was quantified in each biopsy specimen (**C)**. Total IL-6 fluorescence was significantly higher in biofilm positive compared to biofilm negative biopsy specimens. Data displayed as bar and whisker graphs where line designates the median, boxes the 25/75th percentile and whiskers the 95th percentile. A.U., Arbitrary Units; NS, nonsignificant.

Figure S12. Percent of apoptotic cells scored per 1000 epithelial cells counted. Normal surgical tissue from patients with CRC with and without a biofilm, along with normal mucosa from colonoscopy biopsies from healthy individuals with and without a biofilm (subjects without CRC). NS, nonsignificant.