

RESEARCH ARTICLE

Quantitative Profiling of Colorectal Cancer-Associated Bacteria Reveals Associations between *Fusobacterium* spp., Enterotoxigenic *Bacteroides fragilis* (ETBF) and Clinicopathological Features of Colorectal Cancer

Katie S. Viljoen¹, Amirtha Dakshinamurthy¹, Paul Goldberg², Jonathan M. Blackburn^{1*}

1 Institute of Infectious Disease & Molecular Medicine, Division of Medical Biochemistry, Faculty of Health Sciences, University of Cape Town, Cape Town, South Africa, **2** Surgical Gastroenterology Unit, Department of Surgery, Groote Schuur Hospital, Cape Town, South Africa

* jonathan.blackburn@uct.ac.za



OPEN ACCESS

Citation: Viljoen KS, Dakshinamurthy A, Goldberg P, Blackburn JM (2015) Quantitative Profiling of Colorectal Cancer-Associated Bacteria Reveals Associations between *Fusobacterium* spp., Enterotoxigenic *Bacteroides fragilis* (ETBF) and Clinicopathological Features of Colorectal Cancer. PLoS ONE 10(3): e0119462. doi:10.1371/journal.pone.0119462

Academic Editor: Andrew McDowell, University of Ulster, UNITED KINGDOM

Received: August 28, 2014

Accepted: January 23, 2015

Published: March 9, 2015

Copyright: © 2015 Viljoen et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This study was supported by the Cancer Association of South Africa. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Abstract

Various studies have presented clinical or *in vitro* evidence linking bacteria to colorectal cancer, but these bacteria have not previously been concurrently quantified by qPCR in a single cohort. We quantify these bacteria (*Fusobacterium* spp., *Streptococcus gallolyticus*, *Enterococcus faecalis*, Enterotoxigenic *Bacteroides fragilis* (ETBF), Enteropathogenic *Escherichia coli* (EPEC), and *afaC*- or *pks*-positive *E. coli*) in paired tumour and normal tissue samples from 55 colorectal cancer patients. We further investigate the relationship between a) the presence and b) the level of colonisation of each bacterial species with site and stage of disease, age, gender, ethnicity and MSI-status. With the exception of *S. gallolyticus*, we detected all bacteria profiled here in both tumour and normal samples at varying frequencies. ETBF (FDR = 0.001 and 0.002 for normal and tumour samples) and *afaC*-positive *E. coli* (FDR = 0.03, normal samples) were significantly enriched in the colon compared to the rectum. ETBF (FDR = 0.04 and 0.002 for normal and tumour samples, respectively) and *Fusobacterium* spp. (FDR = 0.03 tumour samples) levels were significantly higher in late stage (III/IV) colorectal cancers. *Fusobacterium* was by far the most common bacteria detected, occurring in 82% and 81% of paired tumour and normal samples. *Fusobacterium* was also the only bacterium that was significantly higher in tumour compared to normal samples ($p = 6e-5$). We also identified significant associations between high-level colonisation by *Fusobacterium* and MSI-H (FDR = 0.05), age (FDR = 0.03) or *pks*-positive *E. coli* (FDR = 0.01). Furthermore, we exclusively identified *atypical* EPEC in our cohort, which has not been previously reported in association with colorectal cancer. By quantifying colorectal cancer-associated bacteria across a single cohort, we uncovered inter- and intra-individual patterns of colonization not previously recognized, as well as important associations with clinicopathological features, especially in the case of *Fusobacterium* and ETBF.

Introduction

A causal link between specific pathogens and numerous cancers has now been firmly established. Clear evidence exists for example that the vast majority of cervical cancers are directly caused by infection with human papillomavirus (HPV) [1]. Similarly, *Helicobacter pylori* is a known risk factor for the development of gastric cancer and is considered a class I carcinogen by the WHO [2,3].

The possibility of oncogenic bacteria in the colon was already evident in the 1950s when a clinical association between *Streptococcus bovis* bacteraemia/endocarditis and CRC was discovered [4]. Subsequently, multiple studies have demonstrated enrichment with specific bacterial pathogens in faecal or tissue samples of CRC patients, including, *Fusobacterium* spp. [5–7], *S. gallolyticus* [8–10], *E. faecalis* [11] and Enterotoxigenic *Bacteroides fragilis* (ETBF) [12].

Previously, 16S rRNA profiling of colorectal cancer (CRC) paired tumour and normal biopsies has revealed that while only 3% of biopsy specimens from healthy controls contained any type of bacteria, ~90% of patients with adenomas or carcinomas had 10^3 – 10^5 bacteria in both malignant and macroscopically normal samples [13]. This clearly demonstrates the susceptibility of these patients to colonisation of the normally sterile colonic epithelium—not only in existing tumour tissue, but also in the surrounding macroscopically normal tissue, which may suggest a pre-existing risk to colonisation/infection.

Based on both *in vitro* and *in vivo* observations, bacterially-driven oncogenic mechanisms in CRC have been proposed to include activation of Wnt signaling (ETBF [14], Enteropathogenic *Escherichia coli* (EPEC) [15], *Fusobacterium* [16]), pro-inflammatory signaling (*E. faecalis* [17,18], *S. gallolyticus* [19,20]) and genotoxicity (EPEC [21], AIEC [22–24]).

The oncogenic potential of these bacteria, as well as suspected bacterial components implicated in the aetiopathogenesis of CRC, are summarized in Table 1.

To date, however, the presence and levels of multiple CRC-associated bacteria have not been examined across a single cohort. Further, to our knowledge, ETBF and *E. faecalis* have only been quantified by quantitative PCR (qPCR) in faecal samples of CRC patients, and EPEC has only been quantified in a small CRC cohort with archival FFPE samples [25].

Here we use qPCR to measure the presence of six pathogens, previously reported in association with CRC, in paired adenocarcinoma and adjacent normal mucosal samples; these include *Fusobacterium* spp., *Streptococcus gallolyticus*, *Enterococcus faecalis*, ETBF, EPEC and *afaC*- or *pks*-positive *E. coli*.

Materials and Methods

Cohort selection

This study consists of two cohorts: Firstly 55 paired colorectal patient samples (adenocarcinoma tissue and adjacent normal mucosa) were collected during surgical resection at the Groote Schuur Hospital, with no pre-selected conditions. Ethical consent was obtained (UCT HREC REF 366/2010) and each patient provided written informed consent to donate samples from the tissues left over after surgical resection to subsequent molecular studies. The second cohort was sourced in order to obtain more patients with sporadic microsatellite instability (MSI). For this cohort, 18 adenocarcinoma samples were selected from archival FFPE specimens that had previously been screened for MSI by immunohistochemistry of the mismatch repair genes *MLH1*, *MSH2* and *MSH6*; these patients were referred for MSI testing because CRC was diagnosed under the age of 50 and/or in two or more first or second degree relatives with an HNPCC-related tumor, regardless of age. For our purpose, we selected patients with MSI (absence of staining of one or more MMR proteins), whom also had mutational screening data available so that we

Table 1. Summary of the putative oncogenic mechanisms and the bacterial components implicated in CRC pathogenesis for the six bacterial species quantified in this study.

Bacterial species	Support for putative oncogenic mechanism	Suspected bacterial components implicated
Enteropathogenic <i>Escherichia coli</i> (EPEC)	Downregulates mismatch repair proteins <i>in vitro</i> [21,25]; increases mutational frequency <i>in vitro</i> [21].	<i>espF</i> [21]
<i>Escherichia coli</i> with adherent/and or invasive properties.	Enriched in CRC patients [13,26,27]; CRC-associated strains commonly have genes related to M-cell translocation (<i>lpfA</i>) [22]; genotoxicity (<i>pks</i>) [22–24,28], or cell cycle modulation (<i>cnf1</i>) [23].	<i>pks</i> [22–24], <i>afaC</i> [22], <i>lpfA</i> [22] <i>cnf1</i> [23] and <i>cdt</i> [23].
<i>Fusobacterium</i> spp.	Multiple independent metagenomic studies identify <i>Fusobacterium</i> spp. as overrepresented in CRC tissue compared to matched normal mucosa and healthy controls [5–7]. <i>F. nucleatum</i> increases tumour multiplicity in an APC Min/+ mouse model [29]; Triggers β -catenin nuclear signaling [16].	<i>FadA</i> [16]
Enterotoxigenic <i>Bacteroides fragilis</i> (ETBF)	Enriched in faecal samples from CRC patients [12]; Triggers β -catenin nuclear signaling; induces c-Myc expression and cellular proliferation [14]; increases colitis and tumour in a Min/+ mice model [30].	<i>B. fragilis</i> toxin (<i>Bft</i>) [14]
<i>Streptococcus gallolyticus</i>	Enriched in CRC patients with [8–10,31] and without bacteremia [9]. <i>S. infantarius</i> or its wall extracted antigens promote progression of preneoplastic lesions in rats and promotes pro-inflammatory COX-2 signaling [19,20].	Cell wall extracted antigens
<i>Enterococcus faecalis</i>	Enriched in faecal samples from CRC patients [11]; Produces extracellular superoxide [32], promotes inflammation and CRC in IL-10 knockout mice [17,18], and promotes COX-2 related chromosomal instability [33].	Reactive oxygen species (superoxide, hydrogen peroxide)

doi:10.1371/journal.pone.0119462.t001

could distinguish between sporadic and HNPCC-based MSI. Of the 18 patients selected two had confirmed mutations in the *MLH1* mismatch repair gene, and were therefore classified as HNPCC. Detailed participant-level characteristics are presented in [S1 Table](#).

DNA extraction

DNA was isolated from paired patient samples using a Dounce homogenizer and the AllPrep DNA/RNA/Protein kit (Qiagen) according to the manufacturer’s instructions; for the detection of gram-positive bacteria DNA was extracted using the QIAamp DNA Mini Kit (Qiagen). DNA was extracted from \pm 25mg of tissue, using the following protocol: each sample was incubated in 180 μ l lysozyme @ 20mg/ml for 40 min at 37°C; after adding 20 μ l proteinase K, samples were incubated at 56°C until the tissue was completely lysed (at least 4 hours, or overnight if tissue was still visible after 4 hours); samples were next incubated for 30 min in Buffer AL, thereafter, DNA was isolated according to the manufacturer’s instructions. DNA integrity was confirmed in each case by gel analysis (Agilent BioAnalyser 2100; data not shown). For FFPE samples, DNA was extracted from one or two FFPE slides (depending on availability) using the RecoverAll Total Nucleic Acid Isolation Kit (Ambion). After deparaffinization, samples were incubated in 180 μ l lysozyme (20mg/ml) for 40 min at 37°C, followed by incubation for 42 hours at 50°C in Proteinase K. For the remainder of the protocol DNA was isolated according to the manufacturer’s instructions.

MSI testing

For the cohort of fresh-frozen samples, MSI analysis was conducted on DNA extracted from paired tissue samples as well as the corresponding blood samples for each patient, using allelic profiling of the Bethesda panel of microsatellite markers (BAT25, BAT26, D2S123, D5S346, and D17S250), using primers specified by Loukola *et al.* [34]. Samples were classified as microsatellite stable (MSS), microsatellite instable-low (MSI-L) or microsatellite instable-high (MSI-H) if they had 0, 1 or at least 2 of the 5 markers showing instability, respectively [35].

For the FFPE cohort, immunohistochemistry (IHC) had previously been performed for MSH2, MSH6 and MLH1 by the Division of Anatomical Pathology, University of Cape Town. Samples that displayed absence of staining for any of the mismatch repair proteins evaluated were considered MSI-H since IHC has been shown to have high sensitivity (92.7%) and specificity (100%) in detecting MSI [36]. Originally, patients were referred for IHC based on the following criteria: 1) Colorectal cancer diagnosed under the age of 50 years of age and 2) Colorectal cancer diagnosed in two or more first or second degree relatives with an HNPCC-related tumor, regardless of age.

Primers and control DNA

Primers for the detection of each bacterial species were sourced from the literature or designed in-house, and their specificity was confirmed using Primer BLAST [37]. All primers, along with their limits of detection (LODs) and qPCR efficiencies, are listed in S2 Table.

The following reagents were obtained through the NIH Biodefense and Emerging Infections Research Resources Repository, (NIAID, NIH) as part of the Human Microbiome Project: *Streptococcus gallolyticus* subsp. *gallolyticus*, Strain TX20005, HM-272D; Genomic DNA from *Bacteroides fragilis*, Strain 3_1_12, HM-20D, Genomic DNA from *Clostridium difficile*, Strain NAP07 (CDC#2007054), HM-88D; Genomic DNA from *Enterococcus faecalis*, Strain HH22, HM-200D; Genomic DNA from *Escherichia coli*, Strain B171, NR-9297; and Genomic DNA from *Fusobacterium nucleatum* subsp. *polymorphum*, Strain F0401. ETBF genomic DNA (ATCC43858) was kindly provided by Dr Annalisa Pantosti from the Istituto Superiore di Sanità, Italy. DNA from enterohemorrhagic *E. coli* (to confirm EPEC identity) was kindly supplied by Dr. Anthony Smith at the National Institute for Communicable Diseases, South Africa. DNA from AIEC (strains HM358, HM229 and HM334) was kindly provided by Dr. Barry Campbell from the University of Liverpool, UK.

qPCR amplification conditions

Experiments were performed in triplicate on a Roche LightCycler 480 Real-Time PCR System in 96-well format, using 50 ng patient DNA per well. Separate assays were performed for each bacterial gene detected; the cycling conditions are specified in S3 Table. EPEC (*eaeA*, *bfpA* and *stx1* and *stx2*), ETBF and *S. gallolyticus*, were each detected in 20 μ l reactions using SensiFAST SYBR No-ROX Kit (Bioline); AIEC, *Fusobacterium* spp. and *E. faecalis* were each detected in 25 μ l reactions using Maxima SYBR green qPCR Master Mix (Thermo Scientific). In order to increase specificity, it was necessary in some cases to perform touchdown PCR, whereby the annealing temperature is lowered in a stepwise manner to discourage amplification of off-targets during the first 10 cycles of PCR [38]; Touchdown qPCR was performed for detection of EPEC (*bfpA* and *eaeA*), *S. gallolyticus*, ETBF, EHEC (*stx1* and *stx2*) and AIEC (*afaC*).

qPCR quantification

For each qPCR assay, absolute quantification was performed using a standard curve, which was constructed using serially diluted genomic DNA from the relevant positive control strain. The concentration of bacterial DNA found was expressed in terms of genome copies by calculating the weight of one genome copy for each species as used by Dolezel *et al.* [39]: DNA content (pg) = genome size (bp)/(0.978 x 10⁹).

For example, *Fusobacterium* spp. have an estimated genome size of 2.2 Mb and since one molecule of double stranded DNA of length 978 Mb weighs approximately one picogram, a single *Fusobacterium* spp. genome weighs approximately 2.25 fg (2.2Mb/978Mb = 0.00225 pg)

and therefore 1 ng of DNA from *Fusobacterium* spp. equates to 444,545 copies (1000pg/ (2.2Mb/978Mb)) of the bacterium (Table 2).

In the case of AIEC strains, genome size may vary substantially between strains, since these strains are classified according to phenotypic traits and not sequence similarity. We opted to use the prototypical LF82 AIEC strain, which has a genome size of 4.88 for quantification.

In all cases, data were normalised to total genomic DNA and represented as number of bacterial genomes per 50 ng human DNA, thereby effectively normalising the qPCR data to tissue size.

Positive control standards were spiked with the same amount of human genomic DNA (extracted from uninfected human cell cultures) as used in the patient sample reactions (50ng). The limit of detection (LOD) was defined as the lowest concentration at which a positive result (correct meltcurve) could be obtained in at least 50% of replicates (see S2 Table for details). For all assays except *ClB* and *afaC* at least 70% of replicates were positive at the relevant LOD. In cases where results were inconsistent (1/3 replicates positive), samples were retested and taken as positive if a positive meltcurve was obtained in both runs (the results were then averaged across the two runs to obtain quantitative data).

qPCR quantification in FFPE samples

We first evaluated the quality of DNA extracted from archival FFPE slides (which had been stored between 2 and 23 years) using three primer pairs designed to amplify 100bp, 200bp and 300bp amplicons of the *GAPDH* gene [44]. For most samples we detected either a very faint or no visible band at 100bp; whilst a 200bp amplicon could only be amplified in a few samples. On testing a shorter amplicon (69bp) of the *COX1* gene (which we found to be stably expressed in our cohort and is therefore assumed to have no significant differences in copy number between samples), all samples could be amplified by qPCR; the difference in cycle threshold (Ct) between the highest and lowest quality sample was 9.3. We therefore redesigned the reverse primers for bacterial detection to shorten the resulting amplicons to 60–70bp, and used the *cox1* results to account for degradation in our bacterial quantification. In the case of *eeA*, two reverse primers were designed, one that detects intimin subtypes epsilon, gamma, zeta, alpha, pi, rho, beta, lambda, iota, kappa, eta, delta, xi, mu, kappa and jota; while the second was designed to specifically detect intimin theta (which was found in both EPEC-positive MSI-H samples from the fresh-frozen cohort). The efficiencies for the *COX1* qPCR was calculated using 5-fold serial dilutions constructed using for a high- and low quality patient sample, as 1.96 and 2, respectively. A ‘fold change’ value was then calculated for *COX1* in each sample, using the $\Delta\Delta C_t$ method and the mean Ct across 6 randomly selected DNA samples from the fresh-frozen (high-quality DNA) sample cohort was used as reference (the maximum ΔC_t between fresh-frozen samples was 1.8). These sample-specific ‘fold change’ values for *COX1* between FFPE samples to be tested and the reference set of fresh-frozen samples were used as a correction factor to adjust for DNA sample quality. A theoretical limit of detection was also calculated for each sample by multiplying the correction factor for each sample with the LOD that had previously determined for high quality DNA. After performing absolute quantification, the result was multiplied by the correction factor for each sample. The validity of this method was assessed by comparing *Fusobacterium* quantitation obtained from DNA extracted from fresh frozen samples to that of the matched FFPE samples (which we had available for four patients); after removing a single outlier sample, the Pearson’s correlation coefficient was 0.94, and the median fold change between matched fresh-frozen and FFPE samples was 1. FFPE samples that tested negative for *Fusobacterium* were set to ‘NA’ for downstream analysis, since the negative results could be due either to sample quality or to absence of the bacterium.

Table 2. Estimates of bacterial genome copies per nanogram of bacterial DNA.

	Strain, genome size	Number of target gene copies/genome assumed.	Estimated bacterial copies/ng bacterial DNA.
EPEC (<i>eaeA/bfp</i>)	E2348/69, 4.97 Mb	1	2 x 10 ⁵
ETBF (<i>bft</i>)	3_1_12, 5.49 Mb	1	1.8 x 10 ⁵
<i>E. faecalis</i> (16s rRNA)	V583, 3.34 Mb	4[40]	1.2 x 10 ⁶
<i>Fusobacterium</i> spp. (16s rRNA)	NA, 2.8 Mb[41]	5[41]	1.8 x 10 ⁶
AIEC (<i>afaC</i>)	LF82, 4.88 Mb[42]	1	2 x 10 ⁵
AIEC (<i>CIB</i>)	LF82, 4.88 Mb	1	2 x 10 ⁵
<i>S. gallolyticus</i> (<i>sodA</i>)	UCN34, 2.35[43]	1	4 x 10 ⁵
EHEC (<i>stx1</i>)	O157:H7, 5.6	1	1.8 x 10 ⁵
EHEC (<i>stx2</i>)	O157:H7, 5.6	1	1.8 x 10 ⁵

doi:10.1371/journal.pone.0119462.t002

Statistical analyses

In order to assess quantitative differences between paired tumour and normal samples for each bacterium, we used the Wilcoxon signed rank test applied to the subset of samples, which had at least one positive sample in a pair (tumour or normal).

To assess the association between each bacterium and clinicopathological features, we compared a) samples with vs. without colonisation by a particular bacterium and b) samples with high vs. low/no-colonisation by a particular bacterium. Except for *Fusobacterium*, all other bacterial quantitative data suffered from zero-inflation due to the large number of colonisation-negative patients, which lead to unequal variances between groups. We therefore converted the quantitative data to categorical data where for each bacterium, samples were categorised as 'no-colonisation', 'low-colonisation' or 'high-colonisation'. Quantitative data (copies/50ng) were log₂ transformed and samples with no-colonisation were arbitrarily set to 1 before log₂ transformation; the third quartile (calculated across colonisation-positive cases only) was used to discriminate low- and high-colonisation cases (see Fig. 1 for categories). Associations with clinicopathological features were examined using Fisher's exact test. Meanwhile, in the case of *Fusobacterium* (where the data was normally distributed), we used the Kruskal-Wallis test to evaluate differences between groups stratified by the clinicopathological parameter of interest. Results with an FDR ≤ 0.05 after applying multiple-testing-correction (Benjamini-Hochberg method) over all clinicopathological comparisons made, for each species, were considered significant.

Results

Clinicopathological characterization

The clinicopathological characteristics of the patients in cohort 1 (fresh-frozen samples) are summarized in Table 3. Briefly, the mean age of patients was 59 (SD±15.3), while gender was divided equally. The samples for which MSI-testing was performed (N = 23) included 7 MSI-H (including 4 HNPCC patients) and 3 MSI-L patients. The majority of cases were stage II or III cancers (81.6%), while stage I and IV cancers accounted for 12.2% and 6.1% of cases, respectively. The cohort consists of 60% rectal and 40% colon cancers, with proximal cancers accounting for 45% of the colon cancers. Further, the majority of our cohort is mixed-ancestry (70.4%), while patients of caucasian (14.8%), black (11.1%) and indian ethnicities (3.7%) made up the rest of the cohort.

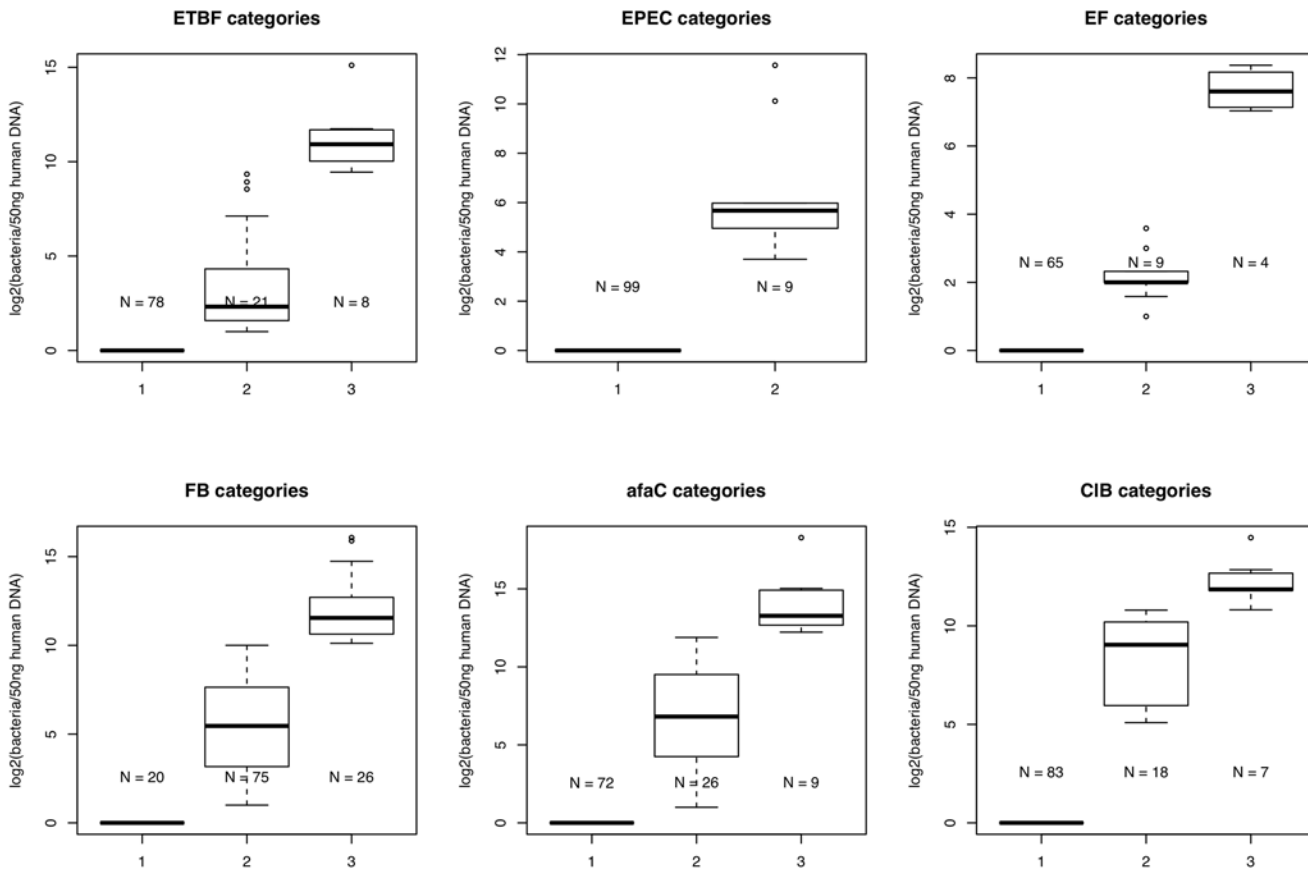


Fig 1. Levels of colonization by each bacterium/gene were categorized using the third quartile (taken across colonisation-positive samples) as a cutoff for high- or low-level colonisation. Categories: 1 (No colonisation), 2 (low colonisation), 3 (high colonisation). In the case of EPEC, because there were so few EPEC-positive patients (N = 6), samples were analysed as positive or negative only. EF: *E. faecalis*; FB: *Fusobacterium*.

doi:10.1371/journal.pone.0119462.g001

Given the association between EPEC and MSI demonstrated by Maddocks et al. [25] and here, as well as the reported association between the levels of colonization by *Fusobacterium* spp. and MSI status, since we had limited numbers of MSI-H samples in the original fresh-frozen cohort, we sourced additional MSI-H samples in order to increase statistical power to measure the relationship between *Fusobacterium* or EPEC and MSI status. To do this, we leveraged archival FFPE samples from Groote Schuur Hospital, Cape Town, South Africa, for which immunohistochemistry of the mismatch repair genes MLH1, MSH2 and MSH6 had been conducted previously.

Bacterial quantification

We quantitated CRC-associated bacteria in adenocarcinoma and matched normal mucosal samples by qPCR, using a serial dilution of genomic DNA from each bacterium as standards, and found varying levels of colonisation in tumour and/or adjacent normal mucosa for all bacteria measured, except *S. gallolyticus*, for which no positive samples were found. While the association between *S. gallolyticus* bacteremia or infective endocarditis and CRC is well established [10], we found only one study, by Abdulmir et al. [9], where *S. gallolyticus* was measured in CRC patients without a history of bacteremia or infective endocarditis. That study found that 4% of healthy controls but 48.7% and 32.7% of CRC patients with or without bacteremia were infected with *S. gallolyticus* in the relevant colonic tissue [9]. In contrast, we did

Table 3. Clinicopathological characteristics of the cohort of fresh-frozen tissues (N = 55).

Feature (patients with missing data)	Number of patients N = 55
Mean age (2)	59 (SD±15.3)
BMI (4)	26.8 (SD±4.7)
Gender (1)	
Female	27 (50%)
Male	27 (50%)
MSI status (23)	
MSS	22 (68.8%)
MSI-H	7 (21.9%)
MSI-L	3 (9.4%)
CRC Type	
HNPCC	6 (10.9%)
Sporadic	49 (89.1%)
Tumour stage (6)	
I	6 (12.2%)
II	18 (36.7%)
III	22 (44.9%)
IV	3 (6.1%)
Tumour site (5)	
Ceacum	4 (8%)
Ascending colon	1 (2%)
Hepatic flexure	1 (2%)
Transverse colon	3 (6%)
Splenic flexure	1 (2%)
Descending colon	3 (6%)
Sigmoid colon	3 (6%)
Rectosigmoid junction (RSJ)	4 (8%)
Rectum	30 (60%)
Radiation/Chemo received before resection (2)	
Yes	22 (41.5%)
No	31 (58.5%)
Ethnicity (1)	
Black	6 (11.1%)
Caucasian	8 (14.8%)
Indian	2 (3.7%)
Mixed-Ancestry	38 (70.4%)

In the case of age and BMI mean values and their standard deviations (SD) are reported. The numbers in column 1 in brackets represent the number of patients with missing data in each category.

doi:10.1371/journal.pone.0119462.t003

not detect *S. gallolyticus* in any of our adenocarcinoma or matched normal mucosa samples using the same primers used by Abdulmir *et al.* [9] for conventional PCR and qPCR. We note that the levels reported in that study were typically very low and we also note that none of our cohort had any reported history of bacteremia/bacterial-endocarditis. It is important in this regard that our qPCR assay was very sensitive (LOD = 5 copies/50ng DNA) and allowed for the detection of gram-positive bacteria such as *S. gallolyticus* by the addition of lysozyme to the homogenized human tissue prior to DNA extraction; this suggests that the

discrepancy between our results and those of Abdulmir *et al.* might be explained by a) differences in sample preparation, b) ethnic differences in the susceptibility to colonization by *S. gallolyticus* or c) geographical differences in *S. gallolyticus* strains found in Southern Africa that may have precluded detection of the bacterium in our cohort. Further investigation is therefore required to clarify this discrepancy.

Of the bacteria that we detected, *Fusobacterium* was by far the most common, occurring in 82% and 81% of paired tumour and normal samples, respectively, with 80% concurrent colonisation in paired samples. *Fusobacterium* was also the only bacterium that was significantly higher in tumour compared to normal samples ($p = 6e-5$, Wilcoxon signed rank test), which is in agreement with previous studies [5,7,45]. The qPCR results are summarized in Fig. 2 and Table 4.

We sequenced *Fusobacterium* amplicons from 10 of the samples in our cohort to confirm melt curve matches, but could not in all cases determine strain-level identity from these amplicons (data not shown). *F. nucleatum* is the only *Fusobacterium* species that has been associated with CRC in the literature to date so it seems reasonable to suppose that in our study the *Fusobacterium* spp is *F. nucleatum*. However, to confirm this would have required a metagenomic sequencing approach to explore *Fusobacterium* spp representation in more detail, since it is not yet known which other *Fusobacterium* spp. might be relevant to the disease; this was beyond the scope of the present study so we refer hereafter exclusively to *Fusobacterium* spp.

In our cohort, we detected ETBF in 14/54 (26%) of colorectal adenocarcinomas and 15/53 (28%) of adjacent normal mucosa samples and 71% of ETBF+ patients were infected in both adenocarcinoma and matched adjacent normal samples. This is consistent with previous studies on faecal samples, which have reported ETBF in $\pm 12\%$ of healthy controls [12,46], 27% of patients with diarrhea [46], and 38% of patients with CRC [12] with colonisation rates appearing to vary widely by geographical location [30].

Although Balamurugan *et al.* demonstrated significantly higher levels of faecal *E. faecalis* in CRC patients compared to healthy controls [11], to our knowledge, ours is the first study to quantitatively measure *E. faecalis* in paired adenocarcinoma (28% *E. faecalis*-positive) and normal mucosa samples (18% *E. faecalis*-positive) with 50% of infected patients being infected in both adenocarcinoma and matched normal mucosa samples. We did not however find any significant clinical associations with *E. faecalis* colonisation.

To investigate the presence of *E. coli* genes that are commonly found in AIEC and which might be relevant to oncogenesis, we quantified the presence of *ClB* (part of the *pks* genomic island) and *afaC* (present in all operons of the afimbrial adhesin family) in paired CRC samples; *pks+* *E. coli* has previously been detected in 55–67% of CRC patients [22,47], compared to 8% of healthy controls [22]. By contrast, in our cohort, 22% of adenocarcinomas and 24% of adjacent normal mucosa samples were *pks+*, and 56% of *pks+* patients were infected in both adenocarcinoma and matched normal mucosa samples. We also detected *afaC* in 36% and 31% of adenocarcinoma and normal mucosa samples, respectively, and found that 80% of *afaC+* patients were infected in both adenocarcinoma and matched normal mucosa samples; this is much lower than that found by Prorok-Hamon *et al.*, who found 67% of CRC patients to be *afaC+* compared to 17% of controls [22]. This discrepancy could be explained by our relatively high LOD for *afaC* and *pks*. Lastly, in contrast to Buc *et al.* [47], who found *pks* to be more common in distal compared to the proximal colon, we did not find a significant association between the presence of *pks* and site of disease. It should be noted that while we examined the presence of *pks* and *afaC*, many other AIEC-related genes that might be relevant to oncogenesis exist, including cyclomodulins [47] and *lpfA* [22].

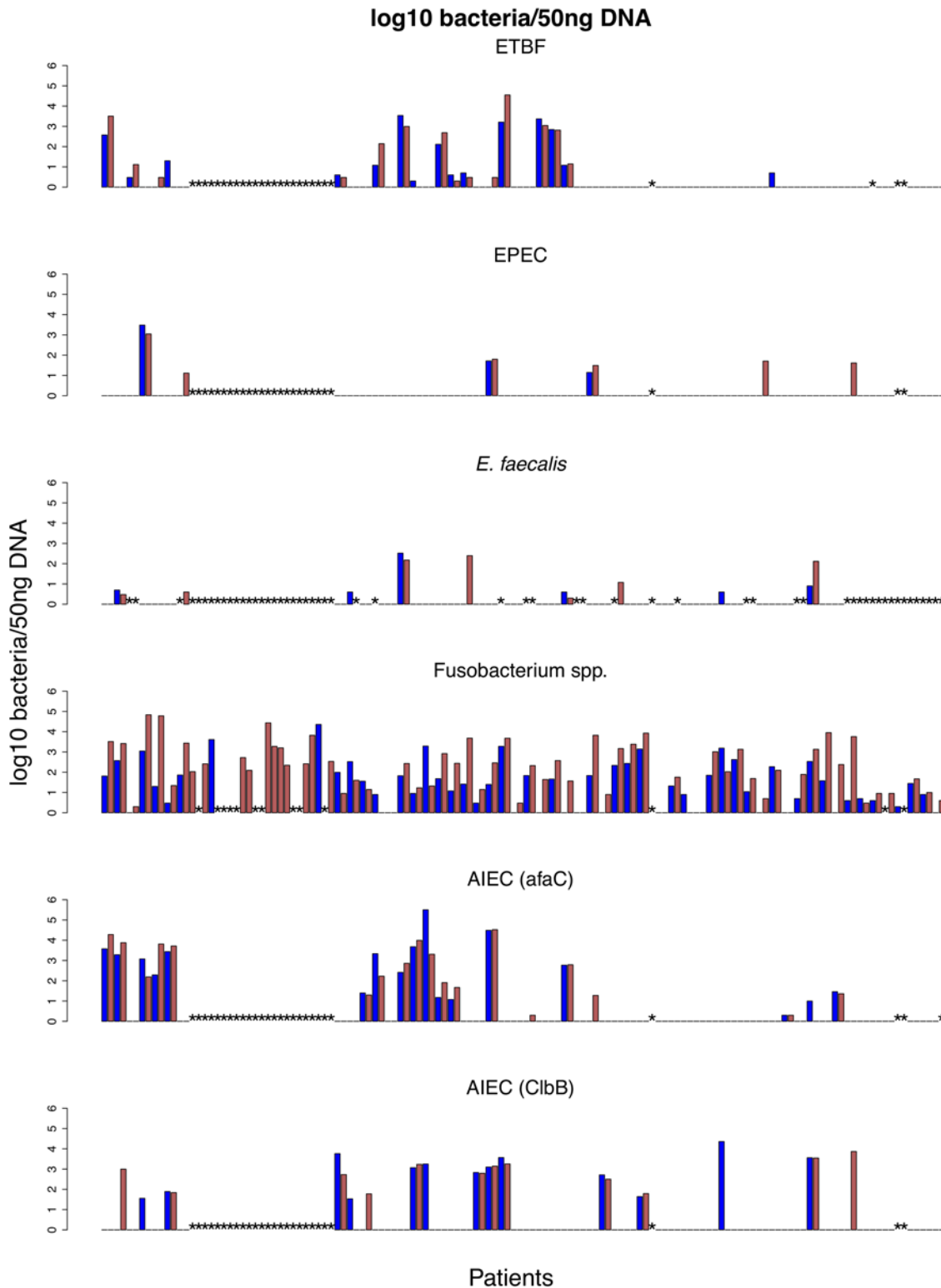


Fig 2. qPCR quantification of bacteria in paired patient samples, expressed as log₁₀ bacteria/50ng of patient DNA. Each bar represents one samples (either tumour or normal) and the order of samples are the same for each bacterium. Red (tumour); blue (normal); * (Not determined).

doi:10.1371/journal.pone.0119462.g002

Table 4. Quantification of bacteria in colorectal cancer and adjacent normal tissues.

Pathogen	Colonisation rate T (%)	Colonisation rate N (%)	Concurrent colonisation in T & N (%)
<i>Fusobacterium</i>	58/71 (82%)	48/59 (81%)	43/54 (80%)
AIEC (<i>afaC</i>)	19/53 (36%)	17/54 (31%)	16/20 (80%)
AIEC (<i>pks</i>)	12/54 (22%)	13/55 (24%)	9/16 (56%)
<i>E. faecalis</i>	11/40 (28%)	7/38 (18%)	5/10 (50%)
ETBF	14/54 (26%)	15/53 (28%)	12/17 (71%)
EPEC	6/54 (11%)	3/54 (6%)	3/6 (50%)
<i>S. gallolyticus</i>	0/45 (0%)	0/45 (0%)	0/45 (0%)

T and N denote adenocarcinoma and adjacent normal mucosa, respectively. Rates of concurrent colonisation in T and N samples were calculated as a fraction of the number of patients who were infected in T and/or N with a particular bacterium.

doi:10.1371/journal.pone.0119462.t004

ETBF and *afaC*-positive *E. coli* are significantly enriched in the colon compared to the rectum of CRC patients.

As shown in Fig. 3, the presence of ETBF and *afaC*-positive strains were significantly associated with the colon compared to the rectum in normal samples (FDR = 0.001 and 0.03, respectively), as well as in tumour samples in the case of ETBF (FDR = 0.002). We did not find any significant differences in colonisation between the proximal and distal colon for any of the bacteria in this study.

Colonisation by ETBF and high-level colonisation by *Fusobacterium* are associated with late-stage CRC.

As shown in Fig. 4, the presence of ETBF was significantly associated with stage of disease (Fisher’s exact, FDR = 0.04 and 0.002 for normal and tumour samples, respectively). Similarly, in the case of *Fusobacterium*, late stage (III/IV) tumour samples were significantly associated with high-level colonisation by *Fusobacterium* (Kruskal-Wallis, FDR = 0.03).

In order to investigate which CRC stages were significantly different in terms of the level of colonisation by *Fusobacterium* we used Dunn’s test to compare individual stages in a pairwise manner. *Fusobacterium* levels were significantly higher in stage III CRCs compared to stage I

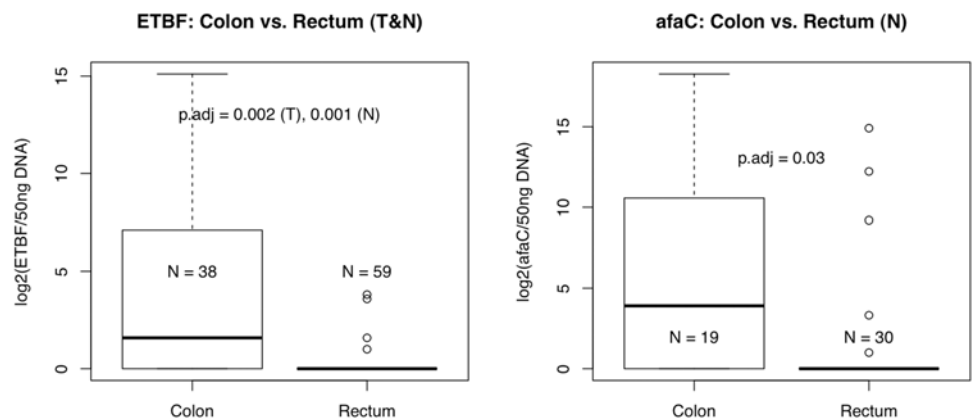


Fig 3. ETBF and *afaC*+ *E. coli* are significantly more prevalent in colon vs. rectal cancers. This applies to both tumour and normal tissue in the case of ETBF (FDR = 0.002, 0.001, respectively) and normal tissue only in the case of *afaC* (FDR = 0.03).

doi:10.1371/journal.pone.0119462.g003

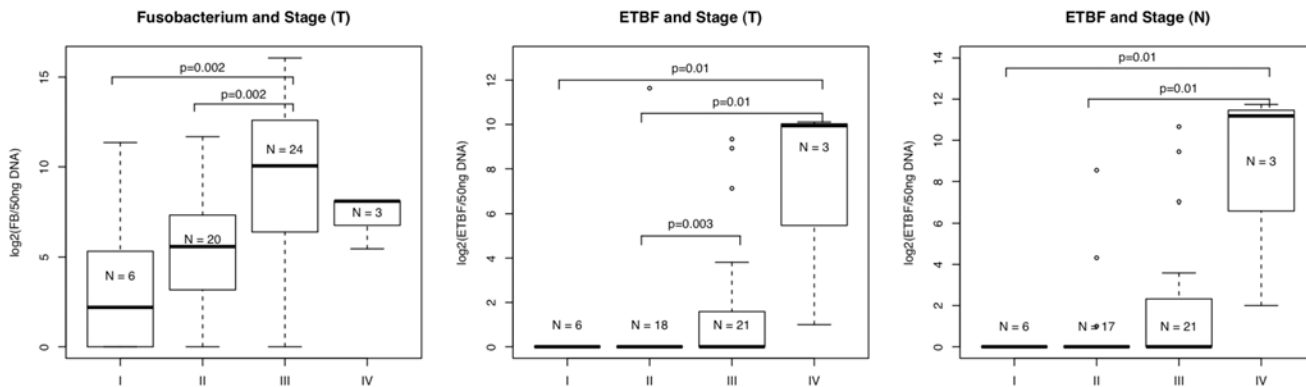


Fig 4. ETBF and *Fusobacterium* are found at significantly higher levels in late stage (III/IV) cancers. For *Fusobacterium*, individual stages were compared in a pairwise manner using Dunn’s test. For ETBF, individual stages were compared in a pairwise manner using Fisher’s exact test. *Fusobacterium* is found at significantly higher levels in stage III CRCs compared to stage I or II CRCs. ETBF is found more frequently in stage III or IV CRCs compared to stage I or II CRCs; and in the corresponding normal mucosa of stage IV CRCs compared to stage I CRCs.

doi:10.1371/journal.pone.0119462.g004

or II CRCs ($p = 0.002$ for both comparisons). For ETBF, for which we found a difference in the presence or absence of ETBF between stages, individual stages were compared in a pairwise manner using Fisher’s exact test. ETBF was found more frequently in stage III or IV CRCs compared to stage I or II CRCs (stage I vs. IV $p = 0.01$; stage II vs. stage IV $p = 0.01$; stage II vs. stage III $p = 0.003$) as well as in the normal mucosa of stage IV CRCs compared to stage I CRCs (stage I vs. IV $p = 0.01$; stage II vs. IV $p = 0.01$).

High-level colonisation by *Fusobacterium* also seems to correlate with chronic inflammation in CRC. For example, McCoy *et al.* found a significant positive correlation between *Fusobacterium* species abundance and local inflammation in adenoma cases [6] whilst we found that there is a trend towards high-level colonisation by *Fusobacterium* in patients with noted inflammation in normal tissue (Kruskal-Wallis test, $p = 0.01$, FDR = 0.07) (Fig. 5) and tumour tissue (Kruskal-Wallis, $p = 0.18$, FDR = 0.2). We also found a positive association between high levels of colonisation by *Fusobacterium* and *pks*-positive *E. coli* in normal tissue (Fisher’s exact, FDR = 0.007) or EPEC in tumour tissue (Fisher’s exact, $p = 0.08$, FDR = 0.2). These data suggest that certain individuals may be more susceptible to bacterial colonisation and inflammation of the normally sterile colonic epithelium.

Further clinical associations with high-level colonisation by *Fusobacterium*

We found a significant relationship between high-level colonisation by *Fusobacterium* and MSI-H, compared to samples that were MSS or MSI-L (Kruskal-Wallis, FDR = 0.05), Fig. 5. We also found a significant increase in *Fusobacterium* levels in CRCs of younger patients (< 60 years), (Kruskal-Wallis, FDR = 0.03) with 31% vs. 11% of patients under or over the age of 60 falling into the *Fusobacterium*-high group of colonisation (Fig. 5). In normal samples, we also noted a trend towards high-level colonisation in males compared to females (Kruskal-Wallis, $p = 0.05$), but this was not significant after multiple testing correction, Fig. 5.

In order to objectively assess the levels of colonisation by *Fusobacterium* between different ethnic groups we used subsets of the data to account for the significant difference in patient age by ethnicity (ANOVA $p = 7.2e-6$); across all patients, the mean age of black patients was 36, that of mixed ancestry patients was 58, and that of white patients was 77. We therefore performed two age- and gender-matched comparisons: a) black patients (mean age = 35, $N = 6$) vs. mixed ancestry patients under the age of 50 (mean age = 42, $N = 10$) and b) caucasian

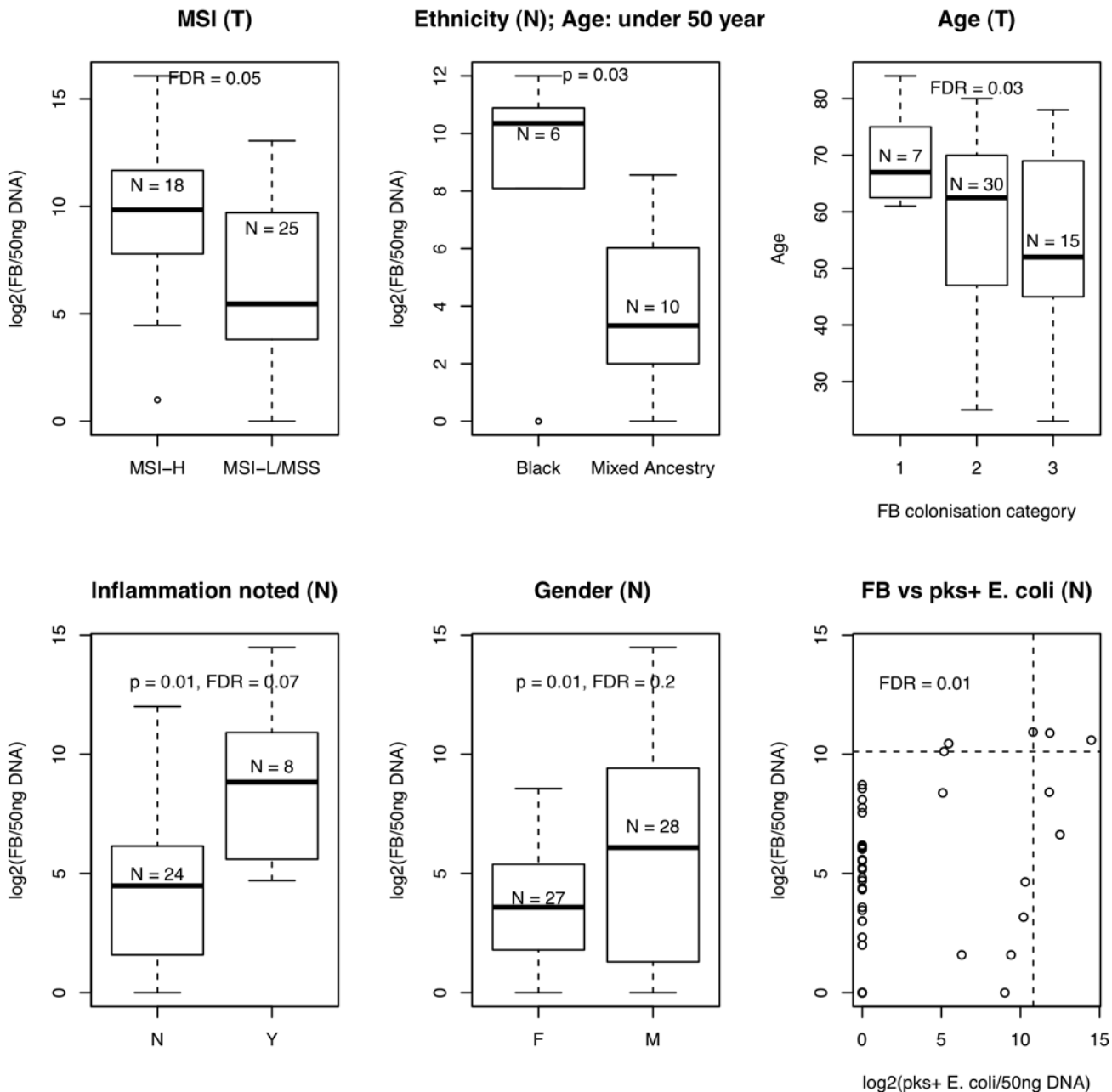


Fig 5. *Fusobacterium* clinicopathological associations. High-level colonisation by *Fusobacterium* is significantly more prevalent in younger patients, males and patients of Black ethnicity. Due to the disproportionately high number of young, black patients seen in our cohort the relationship between ethnicity and levels of colonisation by *Fusobacterium* was assessed using the subset of patients ≤ 50 years. A borderline significant relationship was seen between high-level colonisation by *Fusobacterium* and MSI-H compared to MSS/MSI-L (In our cohort three MSI-L cases were included with the MSS cohort). The vertical and horizontal dotted lines in the bottom right Fig. represent the cutoff for high-level colonisation by *pks+* *E. coli* and *Fusobacterium*, respectively (see [methods](#) for further detail). FB: *Fusobacterium*; B: Black; C: Caucasian; I: Indian; M: Mixed Ancestry; N: normal tissue; F: Female; M: Male.

doi:10.1371/journal.pone.0119462.g005

patients (mean age = 77, N = 8) vs. mixed ancestry patients over the age of 70 (mean age = 72, N = 19). *Fusobacterium* was found at significantly higher levels in black patients compared to their age-matched mixed ancestry counterparts in adjacent normal samples, (Kruskal-Wallis, $p = 0.03$, [Fig. 5](#)), but not in tumour samples (Kruskal-Wallis, $p = 0.6$). No significant differences

were found between age-matched caucasian and mixed ancestry patients in terms of *Fusobacterium* colonisation levels.

Finally, the *Fusobacterium*-high group was also significantly associated with the presence of *pks*-positive *E. coli* in normal samples (Fisher's exact, FDR = 0.01) and two of the three EPEC+ tumours were also infected with *Fusobacterium*-high (Fisher's exact, $p = 0.08$, FDR = 0.2), [Fig. 5](#).

EPEC detection and characterisation

In the fresh-frozen cohort, we detected EPEC via the intimin gene (*eaeA*) in 11% and 6% of tumour and normal samples, respectively; with 50% concurrent colonisation in paired samples. Colonisation levels varied from ± 13 –3037 bacteria/50ng of DNA extracted. Notably, all EPEC-positive cases ($N = 6$) were identified as atypical EPEC (aEPEC) by screening for *bfpA* (present in typical EPEC) and *stx1* (present in shiga-toxin producing *E. coli* (STEC)), neither of which were present in any of our samples. aEPEC has not been previously reported in association with CRC, although the EPEC detected in FFPE CRC samples by Maddocks *et al.* [25] could be aEPEC since they only profiled *eaeA*, and not *bfpA* or *stx1*.

No significant clinical associations were found for EPEC--this is not surprising given the small number of EPEC positive patients ($N = 6$). However, of the six patients infected with EPEC, 67% (2/3) of sporadic MSI-H cases (fresh-frozen cohort), and only 9% (2/22) of MSS were EPEC-positive; the remaining two EPEC-infected patients were of unknown MSI status. Therefore, similar to *Fusobacterium*, there seems to be a trend towards colonisation by EPEC in sporadic MSI-H patients. Furthermore, in light of the effect of intimin subtype (of which there are currently 27 known variants [48]) on tissue tropism [49–51] we sequenced the 150bp amplicon amplified during intimin detection, which is located in the variable region of intimin and identified intimin theta exclusively in the two EPEC-positive MSI-H cases and in one case with unknown MSI-status. In the remaining EPEC-positive samples, intimin subtype could not be conclusively identified based on the 150bp product, but produced equal BLAST scores for the intimin subtypes: zeta 2&3, alpha 2, pi, iota 1, delta, beta 2, epsilon 2&8, jota and lambda in all of the remaining samples. In samples with concurrent colonisation in paired samples, the 150bp product sequences were identical. Intimin sequencing results are summarized in [Table 5](#).

Our finding that intimin theta was exclusively identified in MSI-H EPEC positive cases (both located in the caecum), and in one case of unknown MSI status (located in the rectum) is interesting. Moreover, the two MSI-H patients infected with intimin-theta aEPEC were also the only two patients (with available MSI data) where *MLH1* was hypermethylated, as determined by methylation-specific qPCR. Both these patients were also infected with high levels of *Fusobacterium* (2730 and 68700 copies/50ng in tumour samples). Our data contrast with the work of Maddocks *et al.* who recently demonstrated *in vitro* EPEC-induced depletion of the mismatch repair proteins occurring at the protein level, despite an apparent increase in *MLH1* and *MSH2* mRNA following infection of HT29 cells with EPEC (strain E2348/69) [21]. Maddocks *et al.* concluded that EPEC-induced depletion of MLH1 and MSH2 proteins was dependent on mitochondrial targeting of the EPEC effector protein EspF and that this depletion significantly increased the mutational frequency of infected cells [21]. Further work therefore seems needed to reconcile the apparently differing molecular origins of MLH1 protein depletion suggested by the cell-line-based studies of Maddocks *et al.* and our studies on clinical samples.

Lastly, although we did not sequence the entire intimin gene, the 150bp amplified sequences were consistently identical within but not between patients, suggesting that strains isolated

Table 5. Summary of BLAST search query to identify intimin subtypes.

Sample	Highest scoring BLAST hits	% identity	Patient MSI status (T)	<i>MLH1</i> hypermethylated	Paired T&N sequences identical
44T	theta	100	MSI-H	Y	Y
44N	theta	99	MSI-H	N	Y
63T	theta	100	MSI-H	Y	NA
34N	zeta 2&3, alpha 2, pi, iota 1, delta, beta 2, epsilon 2&8, jota, lambda	100	MSS	N	Y
34T	zeta 2&3, alpha 2, pi, iota 1, delta, beta 2, epsilon 2&8, jota, lambda	100	MSS	N	Y
22T	theta, gamma	98	ND	ND	NA
45T	zeta 2&3, alpha 2, pi, iota 1, delta, beta 2, epsilon 2& 8, jota, lambda	97	ND	ND	NA
29N	zeta, alpha 2, pi, iota 1, delta, beta 2, epsilon, jota, lambda	100	MSS	ND	Y
29T	zeta 2&3, alpha 2, pi, iota 1, delta, beta 2, epsilon2& 8, jota, lambda	100	MSS	ND	Y

Samples highlighted in bold were used to determine the effect of FFPE fixation on the ability to detect EPEC by qPCR. ND: not determined.

doi:10.1371/journal.pone.0119462.t005

from tumour or normal biopsies from a given patient are identical, in agreement with the findings by Martin *et al.* concerning *E. coli* strains in paired CRC samples [26].

Next, given the reported relationship between EPEC and MSI *in vitro* [21,25] as well as the relationship between intimin theta+ aEPEC and MSI seen here, we sourced 18 additional MSI-H samples from archival FFPE samples. However, none of which tested positive for EPEC, but because the median level of EPEC colonisation across EPEC-positive samples from the fresh-frozen cohort was relatively low (51 copies/50ng DNA), we investigated whether the level of degradation in the FFPE samples precluded detection in these samples. To this extent we compared the qPCR results from fresh-frozen (150bp amplicon) and matched archival FFPE samples (70 bp amplicon) for three EPEC-positive patients (5 EPEC-positive T or N samples). EPEC could only be detected in one of the five matched FFPE samples—the sample that displayed the highest level of colonisation (3037 copies/50ng) in the fresh frozen tissue. Further, the median estimated LOD for the FFPE samples was 191 copies/50ng DNA (see [Methods](#) for further details), which is higher than the median level detected in fresh frozen samples (51 copies/50ng DNA). We therefore conclude that if EPEC were present in the MSI-H FFPE samples at levels similar to that seen in fresh-frozen samples, the level of degradation in the FFPE samples would have precluded detection of EPEC, even when attempting to amplify a 70bp amplicon.

Discussion

There are an increasing number of reports in the literature of specific bacteria enriched in CRC patients compared to healthy controls. Here, our goal was to simultaneously characterise these bacteria across a single cohort in both tumour and histologically normal (as identified by a qualified anatomical pathologist) adjacent tissue in order to gain a better understanding of colonisation patterns in CRC patients. By quantifying multiple CRC-associated bacteria in one cohort, we have been able to uncover inter- and intra-individual patterns of colonisation not previously recognised. We further identified significant associations with clinicopathological features including MSI-H (*Fusobacterium*), stage of disease (ETBF and *Fusobacterium*), tumour location (ETBF and *afaC*-positive *E. coli*), age (*Fusobacterium*), as well as a positive association between *Fusobacterium* and *pks*-positive strains.

Notably, our finding that late stage (III) tumour samples were significantly associated with high-level colonisation by *Fusobacterium* is consistent with previous studies demonstrating a positive association between high-level colonisation by *Fusobacterium* and regional lymph node metastases [5,45]. Bonnet *et al.* found a similar trend between cyclomodulin-positive *E. coli*, and stage III/IV colon cancers, which we however did not observe here [52]. Tumour tissue provides a nutrient-rich surface that is not protected by an intact mucosal layer, and the tumour-homing activity of certain bacteria is well documented [53], but this does not necessarily imply oncogenic potential. However, in addition to the enrichment of *Fusobacterium* in tumour vs. normal tissues and in late stage CRCs, *Fusobacterium* spp. are also enriched in irritable bowel disease (IBD) patients (who have a 2–3 fold increased risk of developing CRC) [54] compared to healthy controls. Interestingly, *Fusobacterium* spp. isolated from inflamed tissue in IBD patients were significantly more invasive in a subsequent *in vitro* assay compared to non-inflamed tissue from IBD patients or healthy controls [55], possibly suggesting an active role for *Fusobacterium* in gastrointestinal diseases.

Our finding that *Fusobacterium* levels positively correlate with MSI-H, younger age and black ethnicity may be particularly relevant in the South African setting where a disproportionately high number of young black CRC patients—41–57% of black CRC patients [56–58] compared to only 10% of white CRC patients under the age of 50 [56]—has been reported. CRCs in young black patients do not appear to originate from colonic polyps [57] and have not been linked to IBD or diverticulosis [57]; the majority of these cancers are located in the proximal colon [56–58] (often in the caecum [57]), and often display mucinous histology [58,59] and a higher rate of microsatellite instability (MSI) compared to older patients [56,59]. Our findings may therefore suggest a genetic aspect to susceptibility to high-level colonisation by *Fusobacterium*, which warrants further investigation in a larger cohort.

Tahara *et al.* have previously observed an association between high-level colonisation by *Fusobacterium* and MSI-H, *MLH1* methylation as well as the CpG island methylator phenotype (CIMP) [45] suggesting that *Fusobacterium* might promote MSI by inducing *MLH1* hypermethylation. Importantly, the association between MSI and *Fusobacterium* observed here was independent of the origin of MSI in our cohort, with 4/8 HNPCC adenocarcinoma samples falling into the *Fusobacterium*-high group of colonisation. HNPCC requires inactivation of both alleles of the affected mismatch-repair gene and it is possible that *Fusobacterium* precipitates loss of the wild-type allele through methylation. However, the role of aberrant methylation in the aetiopathogenesis of HNPCC remains questionable: Kaz *et al.* found promoter methylation of *MLH1* in 53% of HNPCC adenomas [60], compared to only 4% of sporadic adenomas, whilst Speake *et al.* found 40% and 25% of hyperplastic polyps of sporadic or HNPCC origin to be CIMP-H [61]. However, LOH or gene conversion are the most frequent mechanism of inactivation of the wild type *MLH1* allele in HNPCC tumours [62–66]. Further, HNPCC and sporadic MSI-H cancers have distinct histological and molecular features: While, both cancer types display lymphocytic infiltration, mucin secretion and poor differentiation [67], HNPCCs tends to originate from classical adenomas compared to sessile-serrated adenomas in the case of MSI-H CRCs [68]; on a molecular level, HNPCCs are strongly associated with mutations in *APC* or β -catenin and/or *KRAS* [67,68], while MSI-H sporadic CRCs instead exhibit *BRAF* mutations, which are present in CRC precursor lesions [68]. Therefore, while it is possible that *Fusobacterium* might cause MSI (and thereby CRC), it seems more likely that *Fusobacterium* preferentially flourishes in MSI-H compared to MSS cancers, perhaps due to the altered glycosylation profile in MSI-H cancers [69], that could facilitate adherence of certain bacteria [70]. Additionally, *F. nucleatum* infection has been shown to stimulate proliferation in CRC- but not in non-neoplastic-cell lines [16] and, *Fusobacterium* spp. have been demonstrated to stimulate cellular proliferation following an initial oncogenic hit (affecting a

component of the WNT signaling pathway) in mice [71] and in CRC cell lines [16]. Taken together, it therefore seems most likely that *Fusobacterium* is not oncogenic itself, but may contribute to tumorigenesis by promoting inflammation and cancer cell proliferation.

It has long been appreciated that certain individuals are more susceptible to aberrant pathogenic colonization of the gut epithelium, which may be accompanied by chronic inflammation, for example in patients with IBD. However, our finding that colonisation by certain bacteria are significantly associated with clinicopathological features in CRC—including the stage and site of disease—is new and might be linked to differential susceptibilities in relation to clinical features, such as age and ethnicity; these association do not necessarily imply oncogenicity since many of the CRC-associated bacteria investigated in this study are asymptotically present in a significant percentage of the population [12,46]. One might therefore expect a pathogenic trend similar to that of *H. pylori* where genetic, environmental and strain-specific risk modifiers govern susceptibility to bacterially-mediated oncogenesis in the colon and where only a small fraction of individuals infected with the bacterium will eventually develop cancer. Evaluating the distribution of bacteria in relation to ethnicity, lifestyle- and clinicopathological factors is the first step in evaluating host-susceptibility to infection and putative bacterially-mediated oncogenic mechanisms. Furthermore, bacterial abundance is not the only factor that may be correlated with clinicopathological features since low-abundant bacteria may exert a significant effect on the host through the secretion of toxins at high levels. For example, Duttilh *et al.* showed that Enterobacterial toxins were among the most highly expressed in metatranscriptomic sequencing data from CRC paired tumour and normal tissues [72], including toxins from *E. coli*, *Salmonella enterica* and *Shigella flexneri* [72]. Evaluating the presence of bacterial toxins with oncogenic potential at the transcriptional or proteomic level will thus provide an additional layer of information to unravel complex host-pathogen interactions with relevance to CRC in the future. Future studies should also be aimed at validating our findings in a larger cohort (particularly in MSI-H CRCs in the case of EPEC); and at profiling *Fusobacterium* at the species level, as well as other AIEC toxins implicated in CRC not examined here, such as *lpfA*.

Establishing causality for any of the bacteria examined here remains a challenge and would require rigorous investigation in animal models as well as large scale epidemiological data, as was used in establishing causality in the case of *H. pylori* and gastric cancer. However, by evaluating the distribution of bacteria in relation to ethnicity, lifestyle- and clinicopathological factors in a South African cohort, we have taken a first step towards this goal and we expect that our data will facilitates the development of targeted research questions for future studies.

Supporting Information

S1 Table. Participant-level characteristics. (a). FF = fresh-frozen; FFPE = formalin-fixed paraffin embedded; Tissue type: N = matched normal mucosa, T = tumour tissue; Ethnicity: MA = mixed ancestry, C = caucasian, B = black, I = indian; Gender: M = male, F = female; Stage: Dukes stage of tumour tissue. (b). Bacterial quantitation data expressed as bacteria/50ng human DNA; EPEC limit of detection (LOD) & *Fusobacterium* LOD: for FFPE tissue these are the estimated LOD's based on normalisation against COX1; MSI method: PCR = Bethesda panel of markers; MLH1 meth. = MLH1 methylation testing by methylation-specific PCR; MMR prot. = MMR protein(s) with known methylation or absence of staining by immunohistochemistry (IHC) of MLH1, MSH2 and MSH6; FB = *Fusobacterium*; EF = *E. faecalis*. (DOCX)

S2 Table. Primers and their limits of detection (LODs) and qPCR efficiencies. (DOCX)

S3 Table. qPCR conditions used. (DOCX)

Acknowledgments

KV thanks the University of Cape Town, the Harry Crossley Foundation, the National Research Foundation (NRF) and the German Academic Exchange Service (DAAD) for PhD bursaries. JMB thanks the NRF for a South African Research Chair Initiative grant. We thank Ms Alvera Vorster, Mr. Ryan Goosen and Henry Adeola for their technical assistance, and the Centre for Proteomic and Genomic Research (CPGR) for the use of equipment for qPCR-based detection. We thank Dr. Annalisa Pantosti for the provision of ETBF genomic DNA, Dr. Anthony Smith for the provision of EHEC genomic DNA and Dr. Barry Campbell for the provision of genomic DNA from AIEC strains. We also thank Dr. Campbell for the helpful discussion regarding AIEC.

Author Contributions

Conceived and designed the experiments: KV JMB. Performed the experiments: KV. Analyzed the data: KV JMB. Contributed reagents/materials/analysis tools: AD PG. Wrote the paper: KV JMB.

References

1. Schiff M, Castle PE, Jeronimo J, Rodriguez AC, Wacholder S. Human papillomavirus and cervical cancer. *Lancet*. 2007; 370: 890–907. PMID: [17826171](#)
2. Manuel A, Machado D, Figueiredo C, Seruca R, Juel L. Helicobacter pylori infection generates genetic instability in gastric cells. *Biochim Biophys Acta*. Elsevier B.V.; 2010; doi: [10.1016/j.bbcan.2010.01.007](#)
3. Vogelmann R, Amieva MR. The role of bacterial pathogens in cancer. *Curr Opin Microbiol*. 2007; 10: 76–81. doi: [10.1016/j.mib.2006.12.004](#) PMID: [17208515](#)
4. McCoy WC, JMM 3rd. Enterococcal endocarditis associated with carcinoma of the sigmoid; report of a case. *Journal Med Assoc State Alabama*. 1951;21.
5. Warren L, Freeman JD, Dreolini L, Castellarin M, Krzywinski M, Strauss J, et al. Fusobacterium nucleatum infection is prevalent in human colorectal carcinoma. *Genome Res*. 2012; 22: 299–306. doi: [10.1101/gr.126516.111](#) PMID: [22009989](#)
6. McCoy AN, Araújo-Pérez F, Azcárate-Peril A, Yeh JJ, Sandler RS, Keku TO. Fusobacterium is associated with colorectal adenomas. *PLoS One*. 2013; 8: e53653. doi: [10.1371/journal.pone.0053653](#) PMID: [23335968](#)
7. 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; 22: 292–298. doi: [10.1101/gr.126573.111](#) PMID: [23132912](#)
8. Gupta A, Madani R, Mukhtar H. Streptococcus bovis endocarditis, a silent sign for colonic tumour. *Color Dis*. 2010; 12: 164–171. doi: [10.1111/j.1463-1318.2009.01814.x](#) PMID: [19226366](#)
9. Abdulmir AS, Hafidh RR, Bakar FA. Molecular detection, quantification, and isolation of Streptococcus gallolyticus bacteria colonizing colorectal tumors: inflammation-driven potential. *Mol Cancer*. 2010; 9: 1–18. doi: [10.1186/1476-4598-9-1](#) PMID: [20051109](#)
10. Boleij A, van Gelder MMHJ, Swinkels DW, Tjalsma H. Clinical Importance of Streptococcus gallolyticus infection among colorectal cancer patients: systematic review and meta-analysis. *Clin Infect Dis*. 2011; 53: 870–8. doi: [10.1093/cid/cir609](#) PMID: [21960713](#)
11. Balamurugan R, Rajendiran E, George S, Samuel GV, Ramakrishna BS. Real-time polymerase chain reaction quantification of specific butyrate-producing bacteria, Desulfovibrio and Enterococcus faecalis in the feces of patients with colorectal cancer. *J Gastroenterol Hepatol*. 2008; 23: 1298–303. doi: [10.1111/j.1440-1746.2008.05490.x](#) PMID: [18624900](#)
12. Toprak NU, Yagci A, Gulluoglu BM, Akin ML, Demirkalem P, Celenk T, et al. A possible role of Bacteroides fragilis enterotoxin in the aetiology of colorectal cancer. *Clin Microbiol Infect*. 2006; 12: 782–6. doi: [10.1111/j.1469-0691.2006.01494.x](#) PMID: [16842574](#)

13. Swidsinski A, Khilkin M, Kerjaschki D, Schreiber S, Ortner M, Weber J, et al. Association Between Intraepithelial *Escherichia coli* and Colorectal Cancer. *Gastroenterology*. 1998; 115: 281–286. PMID: [9679033](#)
14. Wu S, Morin PJ, Maouyo D, Sears CL. *Bacteroides fragilis* enterotoxin induces c-Myc expression and cellular proliferation. *Gastroenterology*. 2003; 124: 392–400. doi: [10.1053/gast.2003.50047](#) PMID: [12557145](#)
15. Umar S, Wang Y, Morris AP, Sellin JH. Dual alterations in casein kinase I-E and GSK-3 β modulate β -catenin stability in hyperproliferating colonic epithelia. *Am J Physiol Gastrointest Liver Physiol*. 2007; 292: 599–607. doi: [10.1152/ajpgi.00343.2006](#)
16. 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*. Elsevier Inc.; 2013; 14: 195–206. doi: [10.1016/j.chom.2013.07.012](#) PMID: [23954158](#)
17. Balish E, Warner T. *Enterococcus faecalis* induces inflammatory bowel disease in interleukin-10 knockout mice. *Am J Pathol. American Society for Investigative Pathology*; 2002; 160: 2253–7. doi: [10.1016/S0002-9440\(10\)61172-8](#) PMID: [12057927](#)
18. Kim SC, Tonkonogy SL, Albright CA, Tsang J, Balish EJ, Braun J, et al. Variable phenotypes of enterocolitis in interleukin 10-deficient mice monoassociated with two different commensal bacteria. *Gastroenterology*. 2005; 128: 891–906. doi: [10.1053/j.gastro.2005.02.009](#) PMID: [15825073](#)
19. Ellmerich S, Schöller M, Duranton B, Gossé F, Galluser M, Klein JP, et al. Promotion of intestinal carcinogenesis by *Streptococcus bovis*. *Carcinogenesis*. 2000; 21: 753–6. PMID: [10753212](#)
20. Biarc J, Nguyen IS, Pini A, Gossé F, Richert S, Thiersé D, et al. Carcinogenic properties of proteins with pro-inflammatory activity from *Streptococcus infantarius* (formerly *S.bovis*). *Carcinogenesis*. 2004; 25: 1477–84. doi: [10.1093/carcin/bgh091](#) PMID: [14742316](#)
21. David O, Maddocks K, Scanlon KM, Sonnenberg MS. An *Escherichia coli* Effector Protein Promotes Host Mutation via Depletion of DNA Mismatch Repair Proteins. *MBio*. 2013; 4. doi: [10.1128/mBio.00152-13](#)
22. Prorok-Hamon M, Friswell MK, Alswied A, Roberts CL, Song F, Flanagan PK, et al. Colonic mucosa-associated diffusely adherent afaC+ *Escherichia coli* expressing lpfA and pks are increased in inflammatory bowel disease and colon cancer. *Gut*. 2014; 63: 761–70. doi: [10.1136/gutjnl-2013-304739](#) PMID: [23846483](#)
23. Buc E, Dubois D, Sauvanet P, Raisch J, Delmas J. High Prevalence of Mucosa-Associated *E. coli* Producing Cyclomodulin and Genotoxin in Colon Cancer. *PLoS One*. 2013; 8. doi: [10.1371/journal.pone.0056964](#)
24. Cuevas-ramos G, Petit CR, Marcq I, Boury M, Oswald E. *Escherichia coli* induces DNA damage in vivo and triggers genomic instability in mammalian cells. *PNAS*. 2010; 107. doi: [10.1073/pnas.1001261107](#)
25. Maddocks ODK, Short AJ, Sonnenberg MS, Bader S, Harrison DJ. Attaching and effacing *Escherichia coli* downregulate DNA mismatch repair protein in vitro and are associated with colorectal adenocarcinomas in humans. *PLoS One*. 2009; 4: e5517. doi: [10.1371/journal.pone.0005517](#) PMID: [19436735](#)
26. Martin HM, Campbell BJ, Hart CA, Mpofu C, Nayar M, Singh R, et al. Enhanced *Escherichia coli* adherence and invasion in Crohn's disease and colon cancer. *Gastroenterology*. 2004; 127: 80–93. doi: [10.1053/j.gastro.2004.03.054](#) PMID: [15236175](#)
27. Kotlowski R, Bernstein CN, Sepeshi S, Krause DO. High prevalence of *Escherichia coli* belonging to the B2+D phylogenetic group in inflammatory bowel disease. *Gut*. 2007; 56: 669–75. doi: [10.1136/gut.2006.099796](#) PMID: [17028128](#)
28. Arthur JC, Perez-Chanona E, Mühlbauer M, Tomkovich S, Uronis JM, Fan T-J, et al. Intestinal inflammation targets cancer-inducing activity of the microbiota. *Science*. 2012; 338: 120–3. doi: [10.1126/science.1224820](#) PMID: [22903521](#)
29. 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*. Elsevier Inc.; 2013; 14: 207–15. doi: [10.1016/j.chom.2013.07.007](#) PMID: [23954159](#)
30. Wu S, Rhee KJ, Albesiano E, Rabizadeh S, Wu X, Yen HR, et al. A human colonic commensal promotes colon tumorigenesis via activation of T helper type 17 T cell responses. *Nat Med. Nature Publishing Group*; 2009; 15: 1016–22. doi: [10.1038/nm.2015](#) PMID: [19701202](#)
31. Corredoira-Sánchez J, García-Garrote F, Rabuñal R, López-Roses L, García-País MJ, Castro E, et al. Association between bacteremia due to *Streptococcus gallolyticus* subsp. *gallolyticus* (*Streptococcus bovis* I) and colorectal neoplasia: a case-control study. *Clin Infect Dis*. 2012; 55: 491–6. doi: [10.1093/cid/cis434](#) PMID: [22563018](#)

32. Huycke MM, Abrams V, Moore DR. Enterococcus faecalis produces extracellular superoxide and hydrogen peroxide that damages colonic epithelial cell DNA. *Carcinogenesis*. 2002; 23: 529–536. PMID: [11895869](#)
33. Wang X, Allen TD, May RJ, Lightfoot S, Houchen CW, Huycke MM. Enterococcus faecalis induces aneuploidy and tetraploidy in colonic epithelial cells through a bystander effect. *Cancer Res*. 2008; 68: 9909–17. doi: [10.1158/0008-5472.CAN-08-1551](#) PMID: [19047172](#)
34. Loukola A, Eklin K, Laiho P, Salovaara R, Kristo P, Järvinen H, et al. Microsatellite marker analysis in screening for hereditary nonpolyposis colorectal cancer (HNPCC). *Cancer Res*. 2001; 61: 4545–9. PMID: [11389088](#)
35. Imai K, Yamamoto H. Carcinogenesis and microsatellite instability: the interrelationship between genetics and epigenetics. *Carcinogenesis*. 2008; 29: 673–80. doi: [10.1093/carcin/bgm228](#) PMID: [17942460](#)
36. Lindor BNM, Burgart LJ, Leontovich O, Goldberg RM, Cunningham JM, Sargent DJ, et al. Immunohistochemistry Versus Microsatellite Instability Testing in Phenotyping Colorectal Tumors. 2002; 20: 1043–1048. PMID: [11844828](#)
37. Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden TL. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics*. 2012; 13: 134. doi: [10.1186/1471-2105-13-134](#) PMID: [22708584](#)
38. Don RH, Cox PT, Wainwright BJ, Baker K, Mattick JS. “Touchdown” PCR to circumvent spurious priming during gene amplification. *Nucleic Acids Res*. 1991; 19: 4008. PMID: [1861999](#)
39. Dolezel J, Bartos J, Voglmayr H, Greilhuber J. Nuclear DNA content and genome size of trout and human. *Cytometry A*. 2003; 51: 127–8. doi: [10.1002/cyto.a.10013](#) PMID: [12541287](#)
40. Sedgley CM, Nagel AC, Shelburne CE, Clewell DB, Appelbe O, Molander A. Quantitative real-time PCR detection of oral Enterococcus faecalis in humans. *Arch Oral Biol*. 2005; 50: 575–83. doi: [10.1016/j.archoralbio.2004.10.017](#) PMID: [15848151](#)
41. Větrovský T, Baldrian P. The variability of the 16S rRNA gene in bacterial genomes and its consequences for bacterial community analyses. *PLoS One*. 2013; 8: e57923. doi: [10.1371/journal.pone.0057923](#) PMID: [23460914](#)
42. Miquel S, Peyretailade E, Claret L, de Vallée A, Dossat C, Vacherie B, et al. Complete genome sequence of Crohn’s disease-associated adherent-invasive E. coli strain LF82. *PLoS One*. 2010; 5. doi: [10.1371/journal.pone.0012714](#)
43. Rusniok C, Couvé E, Da Cunha V, El Gana R, Zidane N, Bouchier C, et al. Genome sequence of *Streptococcus gallolyticus*: insights into its adaptation to the bovine rumen and its ability to cause endocarditis. *J Bacteriol*. 2010; 192: 2266–76. doi: [10.1128/JB.01659-09](#) PMID: [20139183](#)
44. Van Beers EH, Joosse SA, Ligtenberg MJ, Fles R, Hogervorst FBL, Verhoef S, et al. A multiplex PCR predictor for aCGH success of FFPE samples. *Br J Cancer*. 2006; 94: 333–7. doi: [10.1038/sj.bjc.6602889](#) PMID: [16333309](#)
45. Tahara T, Yamamoto E, Suzuki H, Maruyama R, Chung W, Garriga J, et al. Fusobacterium in colonic flora and molecular features of colorectal carcinoma. *Cancer Res*. 2014; 74: 1311–8. doi: [10.1158/0008-5472.CAN-13-1865](#) PMID: [24385213](#)
46. Zhang G, Svenungsson B, Karnell A, Weibtraub A. Prevalence of Enterotoxigenic *Bacteroides fragilis* in Adult Patients with Diarrhea and Healthy Controls. *Clin Infect Dis*. 1999; 29: 590–594. PMID: [10530453](#)
47. Buc E, Dubois D, Sauvanet P, Raisch J, Delmas J, Darfeuille-Michaud A, et al. High prevalence of mucosa-associated E. coli producing cyclomodulin and genotoxin in colon cancer. *PLoS One*. 2013; 8: e56964. doi: [10.1371/journal.pone.0056964](#) PMID: [23457644](#)
48. Xue-han Z, Qing Y, Ya-dong L, Bin L, Renata I, Kong-wang H. Development of a LAMP assay for rapid detection of different intimin variants of attaching and effacing microbial pathogens. *J Med Microbiol*. 2013; 62: 1665–1672. doi: [10.1099/jmm.0.054551-0](#) PMID: [23893919](#)
49. Phillips AD, Navabpour S, Hicks S, Dougan G, Wallis T, Frankel G, et al. Enterohaemorrhagic *Escherichia coli* O157: H7 target Peyer’s patches in humans and cause attaching/effacing lesions in both human and bovine intestine. 2000; 377–381. PMID: [10940275](#)
50. Fitzhenry R J, Pickard D J, Hartland E L, Reece S, Dougan G, Phillips GF A D. Intimin type influences the site of human intestinal mucosal colonisation by enterohaemorrhagic *Escherichia coli* O157:H7. *Gut*. 2002; 50: 180–185. PMID: [11788556](#)
51. Mundy R, Schüller S, Girard F, Fairbrother JM, Phillips AD, Frankel G. Functional studies of intimin in vivo and ex vivo: implications for host specificity and tissue tropism. *Microbiology*. 2007; 153: 959–67. doi: [10.1099/mic.0.2006/003467-0](#) PMID: [17379706](#)

52. Bonnet M, Buc E, Sauvanet P, Darcha C, Dubois D, Pereira B, et al. Colonization of the human gut by *E. coli* and colorectal cancer risk. *Clin Cancer Res*. 2014; 20: 859–67. doi: [10.1158/1078-0432.CCR-13-1343](https://doi.org/10.1158/1078-0432.CCR-13-1343) PMID: [24334760](https://pubmed.ncbi.nlm.nih.gov/24334760/)
53. Cummins J, Tangney M. Bacteria and tumours: causative agents or opportunistic inhabitants? *Infect Agent Cancer*. 2013; 8: 11. doi: [10.1186/1750-9378-8-11](https://doi.org/10.1186/1750-9378-8-11) PMID: [23537317](https://pubmed.ncbi.nlm.nih.gov/23537317/)
54. Hartnett L, Egan LJ. Inflammation, DNA methylation and colitis-associated cancer. *Carcinogenesis*. 2012; 33: 723–31. doi: [10.1093/carcin/bgs006](https://doi.org/10.1093/carcin/bgs006) PMID: [22235026](https://pubmed.ncbi.nlm.nih.gov/22235026/)
55. 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; 17: 1971–8. doi: [10.1002/ibd.21606](https://doi.org/10.1002/ibd.21606) PMID: [21830275](https://pubmed.ncbi.nlm.nih.gov/21830275/)
56. Cronjé L, Paterson A, Becker PJ. Colorectal cancer in South Africa: a heritable cause suspected in many young black patients. *S Afr Med J*. 2009; 99: 103–6. PMID: [19418671](https://pubmed.ncbi.nlm.nih.gov/19418671/)
57. Bremner C, Ackerman V. Polyps and carcinoma of the large bowel in the South African Bantu. *Cancer*. 1970; 26: 991–999. PMID: [5476799](https://pubmed.ncbi.nlm.nih.gov/5476799/)
58. Degiannis E, Sliwa K, Levy R, Hale M, Saadia R. Clinicopathological trends in colorectal carcinoma in a Black South African population. *Trop Gastroenterol*. 1995; 16: 55–61. PMID: [8854959](https://pubmed.ncbi.nlm.nih.gov/8854959/)
59. Cronjé L, Becker PJ, Paterson AC, Ramsay M. Hereditary non-polyposis colorectal cancer is predicted to contribute towards colorectal cancer in young South African blacks. *S Afr J Sci*. 2009; 105. doi: [10.1590/S0038-23532009000100023](https://doi.org/10.1590/S0038-23532009000100023)
60. Kaz A, Kim YH, Dzieciatkowski S, Lynch H, Watson P, Kay Washington M, et al. Evidence for the role of aberrant DNA methylation in the pathogenesis of Lynch syndrome adenomas. *Int J Cancer*. 2007; 120: 1922–9. doi: [10.1002/ijc.22544](https://doi.org/10.1002/ijc.22544) PMID: [17278092](https://pubmed.ncbi.nlm.nih.gov/17278092/)
61. Speake D, O'Sullivan J, Evans DG, Lalloo F, Hill J, McMahon RFT. Hyperplastic polyps are innocuous lesions in hereditary nonpolyposis colorectal cancers. *Int J Surg Oncol*. 2011; 2011: 653163. doi: [10.1155/2011/653163](https://doi.org/10.1155/2011/653163) PMID: [22312515](https://pubmed.ncbi.nlm.nih.gov/22312515/)
62. Parsons MT, Buchanan DD, Thompson B, Young JP, Spurdle AB. Correlation of tumour BRAF mutations and MLH1 methylation with germline mismatch repair (MMR) gene mutation status: a literature review assessing utility of tumour features for MMR variant classification. *J Med Genet*. 2012; 49: 151–7. doi: [10.1136/jmedgenet-2011-100714](https://doi.org/10.1136/jmedgenet-2011-100714) PMID: [22368298](https://pubmed.ncbi.nlm.nih.gov/22368298/)
63. Bettstetter M, Dechant S, Ruemmele P, Grabowski M, Keller G, Holinski-Feder E, et al. Distinction of hereditary nonpolyposis colorectal cancer and sporadic microsatellite-unstable colorectal cancer through quantification of MLH1 methylation by real-time PCR. *Clin Cancer Res*. 2007; 13: 3221–8. doi: [10.1158/1078-0432.CCR-06-3064](https://doi.org/10.1158/1078-0432.CCR-06-3064) PMID: [17545526](https://pubmed.ncbi.nlm.nih.gov/17545526/)
64. Ollikainen M, Hannelius U, Lindgren CM, Abdel-Rahman WM, Kere J, Peltomäki P. Mechanisms of inactivation of MLH1 in hereditary nonpolyposis colorectal carcinoma: a novel approach. *Oncogene*. 2007; 26: 4541–9. doi: [10.1038/sj.onc.1210236](https://doi.org/10.1038/sj.onc.1210236) PMID: [17260015](https://pubmed.ncbi.nlm.nih.gov/17260015/)
65. Zhang J, Lindroos A, Ollila S, Russell A, Marra G, Mueller H, et al. Gene conversion is a frequent mechanism of inactivation of the wild-type allele in cancers from MLH1/MSH2 deletion carriers. *Cancer Res*. 2006; 66: 659–64. doi: [10.1158/0008-5472.CAN-05-4043](https://doi.org/10.1158/0008-5472.CAN-05-4043) PMID: [16423994](https://pubmed.ncbi.nlm.nih.gov/16423994/)
66. Wheeler JM, Loukola A, Aaltonen LA, Mortensen NJ, Bodmer WF. The role of hypermethylation of the hMLH1 promoter region in HNPCC versus MSI+ sporadic colorectal cancers. *J Med Genet*. 2000; 37: 588–92. PMID: [10922385](https://pubmed.ncbi.nlm.nih.gov/10922385/)
67. Jass JR. HNPCC and sporadic MSI-H colorectal cancer: a review of the morphological similarities and differences. *Am J Pathol*. 2004; 3: 93–100.
68. Patil DT, Shadrach BL, Rybicki LA, Leach BH, Pai RK. Proximal colon cancers and the serrated pathway: a systematic analysis of precursor histology and BRAF mutation status. *Mod Pathol*. Nature Publishing Group; 2012; 25: 1423–31. doi: [10.1038/modpathol.2012.98](https://doi.org/10.1038/modpathol.2012.98) PMID: [22684223](https://pubmed.ncbi.nlm.nih.gov/22684223/)
69. Gebert J, Kloor M, Lee J, Lohr M, André S, Wagner R, et al. Colonic carcinogenesis along different genetic routes: glycophenotyping of tumor cases separated by microsatellite instability/stability. *Histochem Cell Biol*. 2012; 138: 339–50. doi: [10.1007/s00418-012-0957-9](https://doi.org/10.1007/s00418-012-0957-9) PMID: [22565205](https://pubmed.ncbi.nlm.nih.gov/22565205/)
70. Campbell BJ, Yu LG, Rhodes JM. Altered glycosylation in inflammatory bowel disease: a possible role in cancer development. *Glycoconj J*. 2003; 18: 851–8.
71. Kostic AD, Chun E, Meyerson M, Garrett WS. Microbes and inflammation in colorectal cancer. *Cancer Immunol Res*. 2013; 1: 150–7. doi: [10.1158/2326-6066.CIR-13-0101](https://doi.org/10.1158/2326-6066.CIR-13-0101) PMID: [24777677](https://pubmed.ncbi.nlm.nih.gov/24777677/)
72. Dutilh BE, Backus L, van Hijum SAFT, Tjalsma H. Screening metatranscriptomes for toxin genes as functional drivers of human colorectal cancer. *Best Pract Res Clin Gastroenterol*. 2013; 27: 85–99. doi: [10.1016/j.bpg.2013.03.008](https://doi.org/10.1016/j.bpg.2013.03.008) PMID: [23768555](https://pubmed.ncbi.nlm.nih.gov/23768555/)