



Aalborg Universitet

AALBORG UNIVERSITY
DENMARK

The presence of bacteria varies between colorectal adenocarcinomas, precursor lesions and non-malignant tissue

Bundgaard-Nielsen, Caspar; Baandrup, Ulrik T; Nielsen, Lars P; Sørensen, Suzette

Published in:
BMC Cancer

DOI (link to publication from Publisher):
[10.1186/s12885-019-5571-y](https://doi.org/10.1186/s12885-019-5571-y)

Creative Commons License
CC BY 4.0

Publication date:
2019

Document Version
Publisher's PDF, also known as Version of record

[Link to publication from Aalborg University](#)

Citation for published version (APA):

Bundgaard-Nielsen, C., Baandrup, U. T., Nielsen, L. P., & Sørensen, S. (2019). The presence of bacteria varies between colorectal adenocarcinomas, precursor lesions and non-malignant tissue. *BMC Cancer*, 19(1), [399]. <https://doi.org/10.1186/s12885-019-5571-y>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- ? Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- ? You may not further distribute the material or use it for any profit-making activity or commercial gain
- ? You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us at vbn@aub.aau.dk providing details, and we will remove access to the work immediately and investigate your claim.

RESEARCH ARTICLE

Open Access



The presence of bacteria varies between colorectal adenocarcinomas, precursor lesions and non-malignant tissue

Caspar Bundgaard-Nielsen^{1,2}, Ulrik T. Baandrup^{1,2,3}, Lars P. Nielsen⁴ and Suzette Sørensen^{1,2*} 

Abstract

Background: A causal association has been suggested between certain bacteria and colorectal cancer (CRC). Only a few studies have, however, investigated the presence of these bacteria directly in colon tissue with conflicting results. It is thus uncertain which role they may have in prognosis and carcinogenesis of CRC.

Methods: Formalin-fixed and paraffin-embedded (FFPE) colorectal tissue samples from patients diagnosed with colorectal cancer (CRC) (tumor and paired normal tissue, $n = 99$), adenomas ($n = 96$), or diverticular disease ($n = 104$) were tested for the presence and bacterial load of *Streptococcus gallolyticus* (*S. gallolyticus*), *Fusobacterium nucleatum* (*F. nucleatum*), and *Bacteroides fragilis* (*B. fragilis*) using quantitative PCR. A subsequent broader search was conducted on a subset of samples using 16S ribosomal RNA gene sequencing. Finally, to evaluate the prognostic value, the bacterial status was compared to patient outcome.

Results: *S. gallolyticus* was not detected by qPCR in any of the investigated tissue samples and *F. nucleatum* and *B. fragilis* were found to be equally distributed in tumors, paired normal tissue, and diverticula, but significantly less present in adenomas compared to both tumors and diverticula. Neither, *F. nucleatum* nor *B. fragilis* status affected the five-year prognosis of the patients. The 16S rRNA gene sequencing data revealed that tumors were associated with the *Prevotella* genus while conversely adenomas and diverticula were associated with *Acinetobacter* genus.

Conclusion: These findings do not support a role of *F. nucleatum* or *B. fragilis* during colorectal beginning, while *S. gallolyticus* was not implicated in the colorectal tissue of a Danish population. A potential role of the bacterial genera *Prevotella* and *Acinetobacter* was indicated, and requires further investigations.

Keywords: Colorectal cancer, Cancer microbiota, Colorectal adenomas, Fusobacterium, Bacteroides, Prevotella, Streptococcus, Acinetobacter

Background

Colorectal cancer (CRC) is one of the most common cancers, with approximately 1.4 million cases diagnosed and 700,000 deaths reported annually worldwide [1]. CRC originates from mutations causing abnormal proliferation in the colorectal epithelium and subsequent formation of an adenomatous growth (adenoma) [2]. Through accumulation of mutations, such adenomas may lead to CRC [3, 4]. Several risk factors are associated with development of CRC, including diet, smoking

and high alcohol consumption [5–9]. Early detection allows efficient treatment of CRC, but only 40% of cases are detected in early stage [10]. To improve diagnostics, screening systems for CRC have been implemented in many countries, where stool samples are analyzed for the presence of occult blood [11]. This, unfortunately, leads to a high number of false positive cases resulting in negative psychosocial consequences, increased costs, discomfort and complications related to follow-up diagnostic investigations [12]. Therefore, more research is needed, in order to find sensitive biomarkers for early non-invasive CRC detection.

A possible role for oncogenic bacteria in CRC was first noted in 1951 [13] and again in 1974 when it was shown

* Correspondence: Suzette.soerensen@rn.dk

¹Centre for Clinical Research, North Denmark Regional Hospital, Hjørring, Denmark

²Department of Clinical Medicine, Aalborg University, Aalborg, Denmark
Full list of author information is available at the end of the article



that 64% of patients suffering from *Streptococcus bovis*-related endocarditis, also had colonic adenomas or CRC [14]. It was later revealed that the *Streptococcus bovis* subtype, *Streptococcus gallolyticus subsp. gallolyticus* (*S. gallolyticus*) had a uniquely strong correlation with CRC. Despite clinical associations [14–18], investigations of the prevalence of *S. gallolyticus* infection directly in CRC tumors have shown conflicting results [19, 20]. Recent studies have demonstrated enrichment with the bacteria *Fusobacterium nucleatum* (*F. nucleatum*) [20–26] and *Bacteroides fragilis* (*B. fragilis*) [27–30] in tumor tissue and fecal material of CRC patients, while a subsequent investigation indicated that high-level colonization with *F. nucleatum* or *B. fragilis* were indicators of poor prognosis in CRC patients [31]. To understand the role of bacteria in colorectal carcinogenesis, we investigated the bacterial involvement in the healthy tissue-adenoma-carcinoma sequence of CRC development. Previous studies investigating precancerous adenomas, have found diverging bacterial compositions. Enrichment of *F. nucleatum* has been documented in both fecal samples from patients with adenomas [32–34] and directly in biopsies from the adenomas [32, 35, 36]. Conversely, Pagnini et al. [37] found a marked reduction of mucosal adherent bacteria in adenomas, while Shen et al. [38] did not detect *F. nucleatum* in adenomas but only in biopsies from healthy volunteers. A recent study by Rezasoltani et al. [34] demonstrated enrichment of *F. nucleatum*, *B. fragilis* and *S. bovis* in tubular, villous and tubulovillous adenomas but not in hyperplastic or serrated polyps, while in contrast, Yu et al. [39] found serrated polyps to be more frequently enriched with *F. nucleatum* compared to tubular adenomas. While a gradual increase in enrichment of *F. nucleatum* from healthy colorectal tissue to adenomas and finally to CRC has been demonstrated [32, 33, 36, 40, 41], less is known for *B. fragilis* [42] or *S. gallolyticus*.

The majority of studies investigating bacterial involvement in the adenoma-carcinoma sequence were based on fecal samples [32, 33, 40, 41]. Fecal samples are plentiful and are thus often used as a non-invasive method for investigating gut microbiota. Some variations can however, be observed between fecal microbiota and the microbiota of the mucosal lesion [43]. As a result, more information is needed concerning enrichment of *S. gallolyticus*, *F. nucleatum* and *B. fragilis* in mucosal samples during the colorectal adenoma-carcinoma sequence. Formalin-fixed and paraffin embedded (FFPE) tissue blocks may serve as an abundant source of tissue, enabling studies on bacterial involvement directly in the colorectal tissue. In this study, we compared bacterial colonization of archival colorectal tissue from non-cancerous tissue, adenomas and tumors. Furthermore, we investigated the effects of bacterial status on patient outcome.

Methods

Sample selection

Using the National Pathology Data Bank, we identified all patients diagnosed with colorectal adenocarcinoma, colorectal adenomas, and diverticular disease at the Department of Pathology, North Denmark Regional Hospital in the period 2002–2010. Following surgical removal, tissue samples were stored as FFPE tissue using standard procedures for the Department of Pathology. Number of samples included was based on sample size calculations for two proportions [44], using a power of 80%, level of confidence of 95% and published prevalences of *S. gallolyticus* [19], *F. nucleatum* [25], and *B. fragilis* [20] in tumor tissue compared to non-neoplastic surrounding tissue. Patients diagnosed with more than one of the investigated lesions were excluded. Samples containing too low DNA concentrations or non-amplifiable DNA were excluded. We collected FFPE tissue from 99 patients diagnosed with colorectal adenocarcinoma (tumors and non-neoplastic paired normal tissue), 96 patients diagnosed with colorectal adenomas, and 104 patients diagnosed with diverticular disease of the colon. An overview of samples can be seen in Additional file 1. Paired normal tissue was only routinely collected from tumors, and thus no paired normal samples were available from diverticula or adenomas. All samples were stored using standard procedures at the Department of Pathology.

Sample preparation and DNA extraction

FFPE samples were collected, with each sample, including tumors and paired normal tissue, occupying separate paraffin blocks. Consecutive tissue sections were prepared from all tissue blocks in the following order: 1 × 4 μm sections for HE (Hematoxylin and Eosin) staining, 4 × 10 μm for DNA purification, and finally 1 × 4 μm sections for comparative HE staining to ensure uniformity and for evaluation by a trained pathologist. This microscope based evaluation revealed neoplastic cells in 23 samples of paired normal tissue and these were therefore excluded, resulting in a total of 99 tumor tissue but only 76 paired normal tissue samples being included. To minimize the risk of cross-contaminations between samples, section knives were changed after each tissue block, and the microtome surface wiped clean with alcohol and RNase Away (Molecular Bioproducts). To monitor potential carry-over of bacterial DNA between samples, an empty paraffin block was included for every 11th patient tissue sample. This paraffin block was freshly prepared but otherwise handled similar to blocks containing tissue.

DNA was isolated from FFPE tissue sections using the AllPrep® DNA/RNA FFPE kit (Qiagen), according to manufacturer's instruction.

Primer design and qPCR amplification and quantification

Quantitative real time polymerase chain reaction (qPCR) was used to investigate presence and quantity of bacterial

species previously associated with CRC in the different histological tissue types. Primers targeting *S. gallolyticus* species, *S. gallolyticus* subspecies *gallolyticus*, *F. nucleatum*, and *B. fragilis* were designed in-house using Primer3 software, and tested for specificity using primer-BLAST (NCBI) [45]. The qPCR sought to determine how the relative abundance of *S. gallolyticus*, *F. nucleatum* and *B. fragilis* differed between different histological tissue types, and thus a reference gene was designed targeting the human β -actin gene. Since DNA extracted from FFPE tissue tends to be fragmented [46], we aimed for amplicon sizes shorter than 200 bp. The sequences, targets, and parameters of the individual primers are summarized in Table 1.

Bacterial DNA was purchased from DSMZ (Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures), including DNA from *S. gallolyticus* subspecies *gallolyticus* (DSM 16831), *S. gallolyticus* subsp. *macedonicus* (DSM 15879), *S. gallolyticus* subspecies *pasteurianus* (DSM 15351), *F. nucleatum* (DSM 15643), and *B. fragilis* (DSM 2151). The bacterial DNA was used for determining limit of detection (LOD) of the individual primers using a dilution series. This was found to be approximately 109 DNA copies for *S. gallolyticus* spp., 10 DNA copies for *S. gallolyticus* subsp. *gallolyticus*, 12 DNA copies for *F. nucleatum*, and 10 DNA copies for *B. fragilis*. The bacterial DNA was further used as positive control for qPCR analyses by spiking bacterial DNA into human DNA samples extracted from FFPE colorectal tumors to mimic the sample types used in this study. The ratio of bacterial DNA to total human DNA was 1:40.

qPCR was performed using the Brilliant III Ultra Fast SYBR[®] Green QPCR Master Mix (Agilent Technologies) according to manufacturer's recommendations, and analyzed on the Mx3005P qPCR System (Agilent Technologies). All experiments were performed in triplicates using 40 ng of input DNA with the following cycling

conditions: 95 °C for 10 min, 40 cycles of 95 °C for 1 min, 55 °C for 30 s and 72 °C for 30 s. In a few cases, several products were apparent on the melting curve analysis, and the PCR was then repeated using a more stringent annealing temperature of 59 °C.

For relative quantification of bacterial DNA in samples the $\Delta\Delta$ Ct method [47] was applied, utilizing the primers summarized in Table 1, with β -actin serving as reference gene.

Five year follow-up

The patient's histological history was followed over a 5 year period using the National Pathology Data Bank. Time of death or occurrence of new cases of adenomas or cancer in the colorectum were noted for each patient. Survival and disease-free survival were analyzed using the Kaplan-Meier method based on detection of bacteria. Social security numbers were not available for two patients, and their clinical data were therefore not recorded.

16S rRNA gene sequencing

To detect other potential bacterial biomarkers, the composition of bacterial genera were analyzed using 16S ribosomal RNA (rRNA) gene sequencing in a subset of the FFPE samples already investigated in this study. A total of 40 tissue samples were chosen using the Research Randomizer software [48] to randomly select 10 samples belonging to each separate histological tissue group (Additional file 2). Bacterial 16S rRNA amplicon sequencing targeting the V4 variable region, was performed by DNAsense (Denmark), and followed a modified version of an Illumina protocol [49]. Briefly, an initial PCR and clean-up was performed as described by Albertsen et al. 2015 [50] using primers targeting the V4 hypervariable region (Table 1) [51], and 35 cycles of amplification. Next, indexing primers were attached to all sequences using a second PCR, followed by clean-up [49]. Finally, all samples were pooled and sequenced

Table 1 Primers used for 16S rRNA gene sequencing and qPCR analysis

Target	Target gene (NCBI Accession number)	Primer sequence	Tm (°C)	Product (bp)
Bacteria and Archaea (sequencing)	V4 variable region of the 16S rRNA (515F and 806R [51])	F: 5'-GTGCCAGCMGCCGCGGTAA-3' R: 5'-GGACTACHVGGGTWTCTAAT-3'	65.4 49.0	~ 250–390 bp ^a
<i>S. gallolyticus</i> spp.	SodA (AP012053) (HE613569) (AP012054)	F: 5'-GCTTGGCTTGTGGTGAATGA-3' R: 5'-GCGAACGTTGCGATACTTGA-3'	59.0 59.3	144
<i>S. gallolyticus</i> subsp. <i>gallolyticus</i>	SodA (AP012053)	F: 5'- AAGCTGCGACAACCTCGCTTT – 3' R: 5'- AAGCGTGTCCCAAACGTCA – 3'	61.1 60.8	150
<i>F. nucleatum</i>	16 s ribosomal RNA (CP012717)	F: 5'-CCCAAGCAAACGCGATAAGT-3' R: 5'-GCGTTGCGTCGAATTAACC-3'	59.2 58.9	117
<i>B. fragilis</i>	16 s ribosomal RNA (M11656)	F: 5'- AGTAGAGGTGGGCGGAATTC – 3' R: 5'- GTGTCAAGTTCAGTCCAGTG – 3'	59.2 59.1	97
β -actin	β -actin (NG_007992)	F: 5'-ACTCGTCATACTCCTGCTTGC-3' R: 5'-CCTCCTCAGATCATTGCTCCTC-3'	60.1 60.0	118

^aAmplicon length varies depending on target bacteria [51]

using a MiSeq (Illumina, USA) as previously described [52]. 20% PhiX control library (Illumina) was added to estimate error rate during sequencing, a negative control (nuclease-free water) was added to eliminate background while a positive control (complex sample obtained from an anaerobic digester system) were used to monitor sequencing efficiency and batch effects.

Bioinformatics

Quality of reads were analyzed using FastQC (Babraham Bioinformatics, UK). Forward reads were trimmed using Trimmomatic v0.32 [53] to remove poor reads and reads shorter than 250 bp using the settings SLIDINGWINDOW:5:3 and MINLEN:250. The reads were next dereplicated and processed using the UPARSE workflow [54]. The initial 250 bp of all sequencing reads were clustered using the Usearch v. 7.0.1090 -cluster_otus command with default settings. Operational taxonomic units (OTUs) were formed based on 97% identity and chimeras removed using the Usearch v. 7.0.1090 -usearch_global command with -id 0.97. Finally, taxonomy was assigned using the RDP classifier [55] as implemented in the parallel_assign_taxonomy_RDP.py script in QIIME [56] using the MiDAS database v. 1.20 [57].

Statistics

Data analysis was performed using R version 3.5.2 [58] through the Rstudio IDE (<http://www.rstudio.com/>), and Microsoft Office Excel 2013. For continuous data, distributions were tested using Shapiro-Wilk test. 16S rRNA gene sequencing data was analyzed using the ampvis2 package v.2.3.11 [59] for Rstudio. α diversity was determined using OTU richness and Shannon diversity index as implemented in the amp_alphadiv command of the ampvis2 packet in R. β diversity was visualized using heat maps depicting the 20 most commonly found OTUs and explored using Principal component analysis (PCA) and redundancy analysis (RDA) clustering of Hellinger Distance transformed OTU abundances. Bacterial genera with statistical significant different distributions amongst differing tissue types, were identified using the DESeq2 package in Rstudio [60] to generate multiple hypothesis corrected p -values using the Benjamini-Hochberg procedure [61]. For a bacterial genus to be considered for further analysis, it needed to be significantly different between tissue groups, and the difference was required to be universal for the majority of samples in the tissue group. That is, for a bacteria to be considered associated with tumor tissue, it should constitute a statistically significant higher proportion of bacteria in the majority of tumor samples.

Categorical data, like presence or absence of bacteria, were analyzed using χ^2 test. For continuous data like OTU richness and Shannon diversity index, distribution

was tested using Shapiro-Wilks test while variance was tested using Bartlett's test. Normal distributed data with equal variance were compared using ANOVA followed by Tukeys post-hoc test while non-parametric data were tested using Kruskal-Wallis test followed by Dunn's post-hoc test. Finally, 5-year follow-up data were analyzed using the Kaplan-Meier method, and a log-rank test were used to compare outcome between patients positive and negative for bacterial infection.

A p value of < 0.05 was considered significant for all statistical tests, with the exception of multiple hypothesis corrected p values where a limit of < 0.01 was utilized.

Results

Demographic and histopathological description of patient samples

In this study, colon samples from four different histological tissue groups were analyzed. The demographic and histopathological characterization of these groups are presented in Tables 2 and 3 respectively. CRC patients (71 ± 10.1 years) were significantly older ($p < 0.05$) than adenoma (66 ± 11.7 years) and diverticulum patients (63 ± 14.0 years). While the location of tumor samples was more widely distributed, the majority of adenoma and especially diverticulum samples were localized in the left colon, constituting 48.5% of CRC tumor cases, 76.0% of adenoma cases and 89.4% of diverticulum cases. Information concerning age group, gender, histologic tissue group and follow up data can be found in Additional file 1.

S. gallolyticus was not detected in any of the investigated tissue groups

To establish whether CRC was associated with *S. gallolyticus*, we utilized qPCR to compare the prevalence and quantity of the bacteria in colorectal tumors, paired normal tissue, adenomas and diverticula. Surprisingly, *S. gallolyticus* was below the LOD for both primers targeting all *S. gallolyticus* spp. as well as the more sensitive primers specifically targeting *S. gallolyticus* subsp. *gallolyticus*, in all tissue types (see Additional file 1).

F. nucleatum and *B. fragilis* were enriched in tumors compared to adenomas, but not paired normal tissue or diverticula

To establish the degree of colonization with *F. nucleatum* and *B. fragilis* at different stages in the colorectal adenoma-carcinoma sequence in CRC, we utilized qPCR to compare the prevalence and quantity (Fig. 1) of the bacteria in colorectal diverticula, adenomas, tumors and paired normal tissue. For all empty paraffin blocks, the quantity of the tested bacteria were below the LOD of the primer, indicating that no cross-contamination occurred.

Table 2 Demographic characterization of the individual patient groups

Feature	CRC (Tumors)		CRC (paired normal tissue) ^a		Adenomas		Diverticula	
Number of samples	99		76 ^b		96		104	
Age (median)	71 ^c		70.5 ^c		66		63	
Standard deviation	± 10.1		± 9.9		± 11.7		± 14.0	
Age groups (%)	No.	(%)	No.	(%)	No.	(%)	No.	(%)
< 40	0	(0.0)	0	(0.0)	1	(1.0)	3	(2.9)
40–49	2	(2.0)	1	(1.3)	6	(6.3)	16	(15.4)
50–59	8	(8.1)	5	(6.6)	18	(18.8)	24	(23.1)
60–69	35	(35.4)	28	(36.8)	30	(31.3)	25	(24.0)
70–79	28	(28.3)	22	(28.9)	25	(26.0)	17	(16.3)
80–89	22	(22.2)	16	(21.1)	15	(15.6)	19	(18.3)
> 90	4	(4.0)	4	(5.3)	1	(1.0)	0	(0.0)
Gender (%)	No.	(%)	No.	(%)	No.	(%)	No.	(%)
Male	44	(44.4)	41	(53.9)	47	(48.0)	55	(52.9)
Female	55	(55.6)	35	(46.1)	51	(52.0)	49	(47.1)

^aHealthy colon mucosa surrounding the tumor tissue. ^b 23 samples were removed following examination of the HE stained sections by a trained pathologist, due to presence of cancer cells. ^cpatients donating tumors and paired normal tissue are significantly older than those with adenomas or diverticula

Table 3 Histopathological description of the four tissue groups

Tissue group	Location ^a			Disease State		
CRC (tumor) (n = 99)	Location	n	(%)	Tumor stage ^b	n	(%)
	Right colon	49	(49.5%)	<i>I</i>	7	(7.1%)
	Left colon	48	(48.5%)	<i>II</i>	41	(41.4%)
	Rectum	2	(2.0%)	<i>III</i>	43	(43.4%)
	<i>Not reported</i>	0	(0.0%)	<i>IV</i>	8	(8.1%)
CRC (paired normal tissue) (n = 76)	Location	n	(%)	Tumor stage ^b	n	(%)
	Right colon	41	(53.9%)	<i>I</i>	7	(9.2%)
	Left colon	34	(44.7%)	<i>II</i>	29	(38.2%)
	Rectum	1	(1.3%)	<i>III</i>	32	(42.1%)
	<i>Not reported</i>	0	(0.0%)	<i>IV</i>	8	(10.5%)
Adenomas (n = 96)	Location	n	(%)	Subtype	n	(%)
	Right colon	15	(15.6%)	<i>Villous</i>	1	(1.0%)
	Left colon	73	(76.0%)	<i>Tubulovillous</i>	6	(6.3%)
	Rectum	3	(3.1%)	<i>Tubular</i>	89	(92.7%)
	<i>Not reported</i>	5	(5.2%)	Dysplasia	n	(%)
				<i>Mild</i>	11	(11.5%)
<i>Moderate</i>				55	(57.3%)	
<i>Severe</i>	30	(31.3%)				
Diverticula (n = 104)	Location	n	(%)	Inflammation	n	(%)
	Right colon	4	(3.8%)	<i>Diverticulosis</i>	23	(22.1%)
	Left colon	93	(89.4%)	<i>Diverticulitis</i>	81	(77.9%)
	Rectum	0	(0.0%)			
	<i>Not reported</i>	7	(6.7%)			

^aLocation of tissue samples in the colorectal tract, based on the proposed distinction described by Bufill et al. [73]. ^b Tumor classification based on the recommendations of International Union Against Cancer [74]

F. nucleatum could be detected in 29.3% of tumor samples and *B. fragilis* in 36.4% of cases (Fig. 1a). These distributions were comparable to those found in paired normal tissue and non-malignant diverticula, except for *B. fragilis* which was detected in a higher proportion of paired normal tissue samples (52.6%, $p < 0.05$). The presence of *F. nucleatum* and *B. fragilis* were furthermore comparable when stratifying tumor samples based on cancer stages (Fig. 1b). For tumors, paired normal tissue and diverticula, the bacterial loads of *F. nucleatum* and *B. fragilis* were comparable (Fig. 1c). Intriguingly, we detected *F. nucleatum* and *B. fragilis* significantly less common in adenoma tissue (3.0 and 5.9% respectively) compared to both tumor tissue (29.3 and 36.4%, $p < 0.001$) and diverticula (31.7 and 44.2%, $p < 0.001$) (Fig. 1a). In addition, the adenomas contained significantly less *B. fragilis* DNA compared to diverticula ($p < 0.05$) (Fig. 1c).

Overall, neither *F. nucleatum* nor *B. fragilis* were found to be specifically associated with tumors of CRC patients, but both bacteria were noted by their low presence in adenomas.

***F. nucleatum* and *B. fragilis* status do not affect survival or disease-free survival of patients over a five year period**

To assess the clinical significance of *F. nucleatum* and *B. fragilis*, information on disease progression and survival were collected for all patients for a 5 year period following initial diagnosis. A Kaplan-Meier analysis was performed to examine the relationship between bacterial status in the investigated patients, with survival and risk of developing new cases of CRC or adenomas (Fig. 2). Detection of *F. nucleatum* or *B. fragilis* did not result in significant ($p > 0.05$) changes in survival or disease-free survival rates of patients within a 5 year period.

Bacterial composition of tumor tissue overlaps with that of paired normal tissue, but differs from adenoma and diverticula

To determine if CRC tissue from the four groups (CRC tumors, paired normal tissue, diverticula, and adenomas) differed in overall bacterial composition, we applied a more global approach using 16S rRNA gene sequencing. A subsection of 10 samples from each group were randomly selected and analyzed. Following quality filtering and chimera removal, 566,527 16S rRNA sequence reads (mean number per sample: $16,186.5 \pm 4814$ reads) were obtained. A total of 696 unique Operational Taxonomic Units (OTUs) were identified, with 97.99% being identified on the phylum taxonomic level and 63.51% on the genus level. One sample from the adenomas and four paired normal tissue yielded less reads than the negative controls (3045 and 3123 reads) and were thus excluded, resulting in a total of 35 samples (10 tumors, 6 paired normal tissue, 9 adenomas and 10 diverticula) being

analyzed. A rarefaction curve was produced, showing good sequencing coverage (data not shown).

We first investigated bacterial richness and diversity of the four tissue types (Fig. 3). No significant differences were observed between either tissue types for either OTU richness (Fig. 3a) or Shannon diversity index (Fig. 3b), although diverticula had a slightly higher OTU richness compared to all other tissue types. β diversity was established to determine differences and similarities in bacterial composition between the tissue types. Interestingly, the differences between bacterial compositions were minor as indicated by clustering on the PCA plot (Fig. 4a). These minor changes were elucidated through a subsequent RDA plot that reveal limited tissue specific clustering (Fig. 4b). Tumor tissue clustered separately from diverticula and adenoma tissue, but was highly similar to paired normal tissue.

Differences in proportion of *Prevotella* and *Acinetobacter* defines tissue from CRC samples versus adenoma and diverticula

We next sought to identify the bacteria that differed between the investigated subsection of colorectal tissue samples (Fig. 5a). Of especial note are the genera *Prevotella* and *Acinetobacter*. *Prevotella* is a dominant bacteria in several samples from tumors and paired normal tissue, but markedly absent from especially adenoma but also diverticula ($p < 0.01$). *Acinetobacter* are conversely not represented in tumor or paired normal tissue, but are dominant in the majority of samples originating from adenomas or diverticula ($p < 0.01$). Despite the *Streptococcus* species *S. gallolyticus* being below the LOD for all samples investigated in this study, the genus *Streptococcus* was significantly more common in tumor tissue compared to paired normal tissue, diverticula and adenomas ($p < 0.01$). The genus *Fusobacterium* was significantly more common in tumor tissue compared to adenomas ($p < 0.01$), but not paired normal tissue nor diverticula ($p > 0.05$), while no differences were observed in composition of the *Bacteroides* genus between any tissue types investigated ($p > 0.05$). We observed that a high composition of the genera *Fusobacterium* and *Bacteroides* using sequencing (Fig. 5a) did not clearly correlate with detection of the bacterial species *F. nucleatum* or *B. fragilis* using qPCR (Fig. 5b).

Discussion

In recent years, there has been a growing number of reports concerning a possible link between different bacterial species and the development of CRC. Several bacteria have been implicated, including *S. gallolyticus* [15, 17–19], *F. nucleatum* [20–22, 62] and *B. fragilis* [27–29, 63]. To investigate changes in the bacterial composition along the colorectal healthy tissue-adenoma-carcinoma sequence,

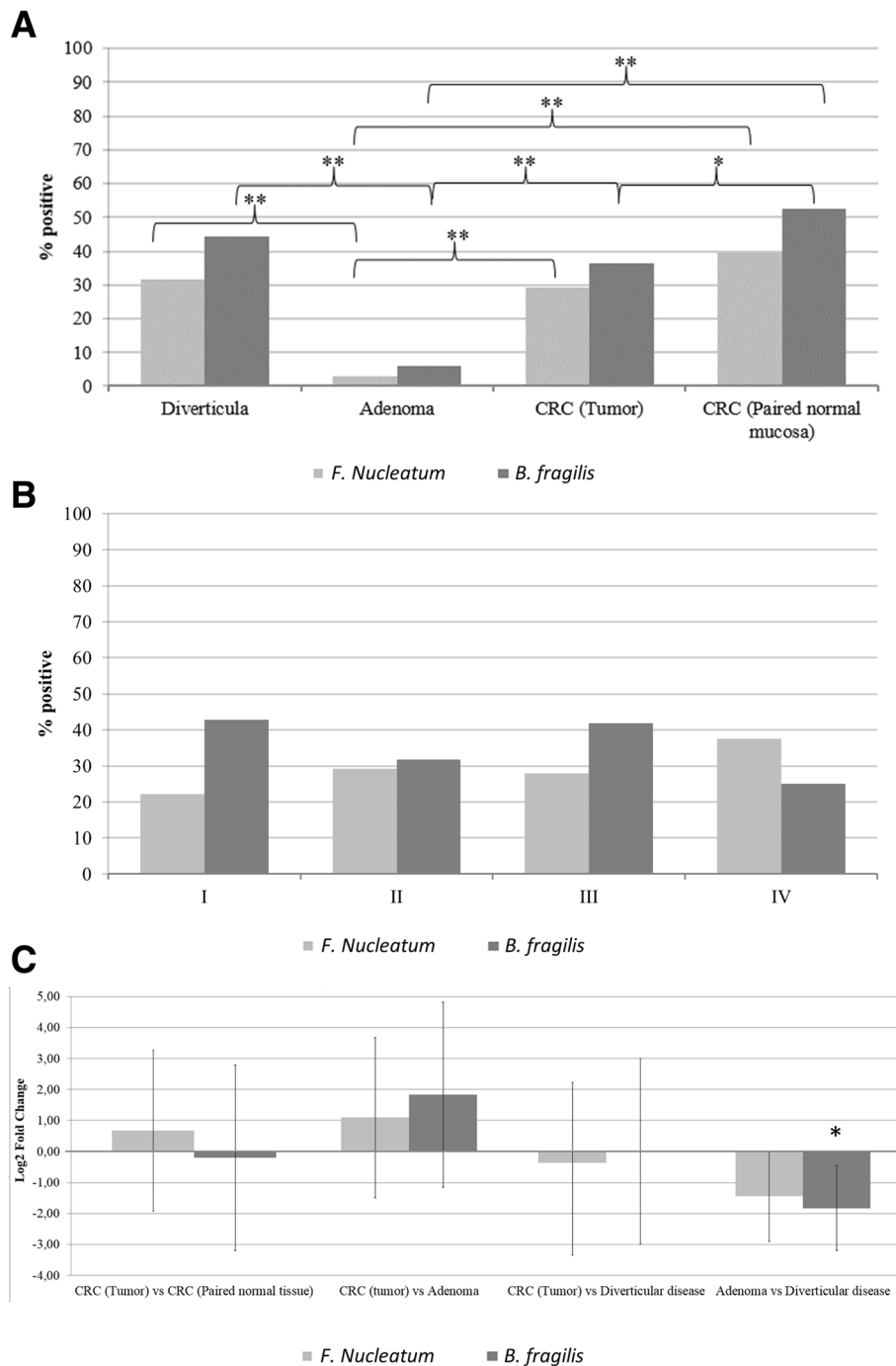


Fig. 1 Presence and quantity of *F. nucleatum* and *B. fragilis* in colorectal tissue. qPCR determination of presence and quantity of bacterial DNA in 99 colorectal tumor tissue, 76 paired normal tissue, 96 adenomas and 104 diverticula. **a** Prevalence of *F. nucleatum* and *B. fragilis* in colorectal tissue. Positivity was determined as bacterial species with a DNA quantity above the LOD of the primers. **b** Prevalence of *F. nucleatum* and *B. fragilis* in different stages of CRC. No statistical significant differences were observed. **c** Difference in quantity of *F. nucleatum* and *B. fragilis* DNA in colorectal tumor tissue compared to paired normal tissue, adenomas and diverticula as well as in adenomas compared to diverticula. Brackets denote standard deviation. * $P < 0.05$, ** $P < 0.001$

we performed qPCR and 16S rRNA gene sequencing on FFPE tissue from colorectal diverticula, adenomas, tumors and paired normal tissue.

Adenomas harbored a distinct bacterial community compared to non-malignant controls, which has been supported by others [35, 37, 38]. While the genus

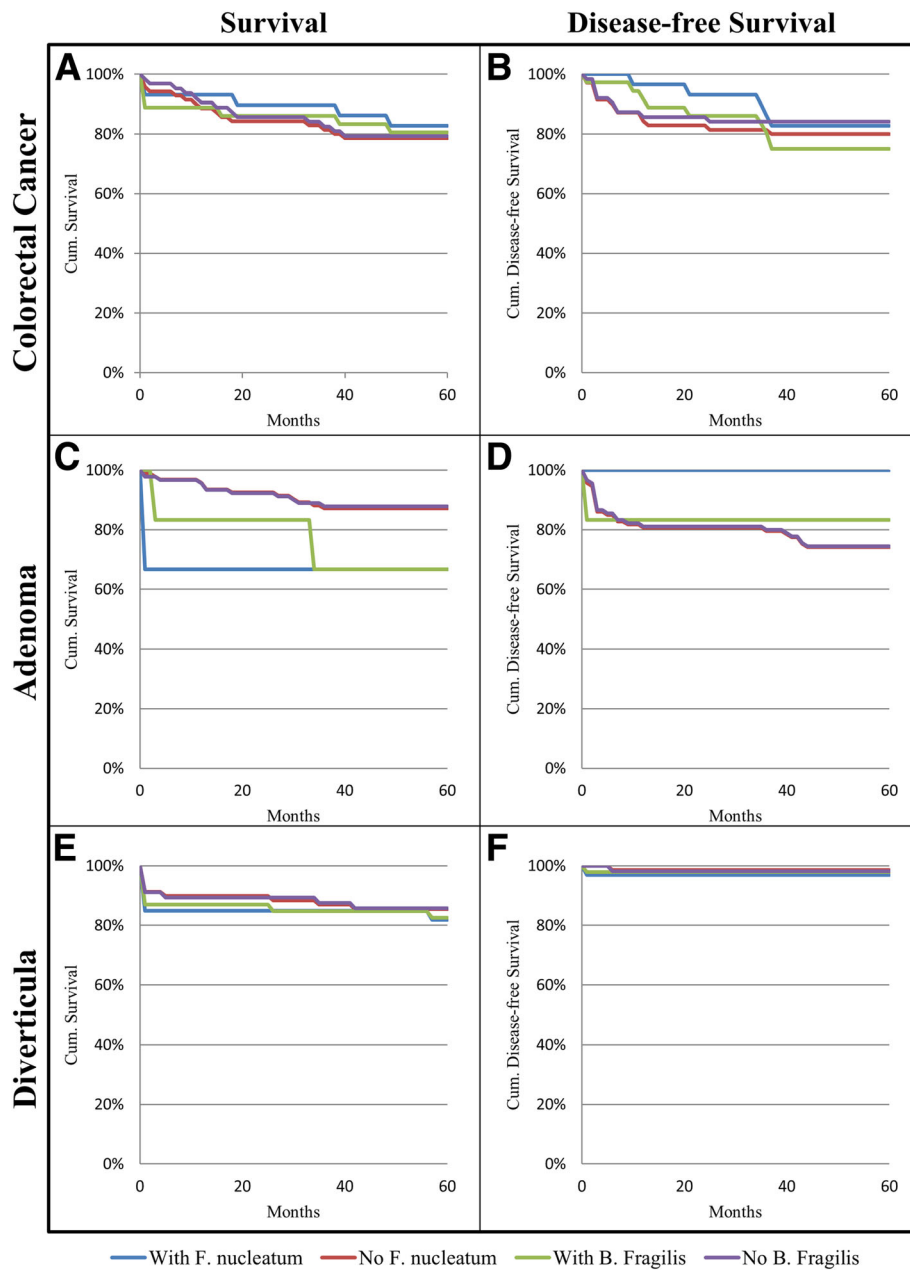


Fig. 2 Five-year follow-up based on presence of *F. nucleatum* or *B. fragilis*. Survival (a, c, e) and disease-free survival (b, d, f) of patients presenting with CRC (a and b, n = 99), adenomas (c and d, n = 96) or diverticula (e and f, n = 104) depending on presence or absence of *F. nucleatum* or *B. fragilis*. Five-year follow-up data was not available for two study participants belonging to the diverticula group. These patients were excluded from the follow-up analysis

Acinetobacter constitutes a large percentage of total bacteria in both diverticula and adenomas, the relative abundance of *Bacteroides*, as well as the percentage of samples positive for the species *F. nucleatum* and *B. fragilis*, were reduced in adenomas compared to both diverticula and paired normal tissue. The cause for this different microbial composition is unknown, but may result from increased local inflammation during adenoma formation, as previously indicated [37]. This increased

inflammation may result in development of a microbial community with oncogenic potential [42, 64]. Notably, not all adenomas transition into CRC [65], and it will therefore be interesting, to establish whether there exists different subtypes of adenomas with various bacterial compositions and potential of carcinogenic progression. During the colorectal adenoma-tumor sequence, we observed a marked increase in the relative abundance of the bacterial genus *Prevotella* as well as the species *F.*

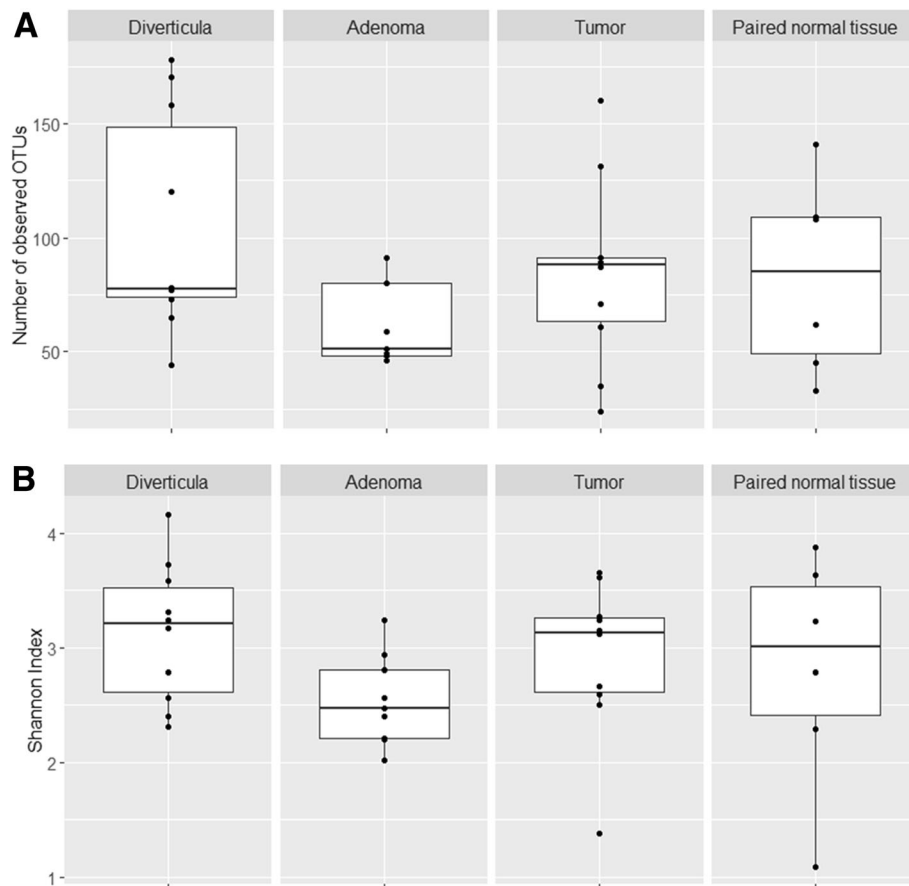


Fig. 3 Gut microbiome richness and diversity between tissue types in a subsection of samples. **a** OTU richness and **b** Shannon diversity index was compared between a subsection of the tumors, paired normal tissue, adenomas and diverticula included in this study. A total of 35 tissue samples were investigated, with 10 tumors samples, 6 paired normal tissue, 9 adenomas and 10 diverticula

nucleatum and *B. fragilis*, all of which have previously been shown to be associated with colorectal tumors [24, 32, 33, 36, 40–42, 66, 67]. These bacteria are known to promote a pro-inflammatory environment [27, 32, 63, 68, 69], and may thus drive the adenoma-tumor transition by inducing local chronic inflammation. Conversely,

we observed that bacteria belonging to the genus *Acinetobacter* were absent from all samples originating from patients diagnosed with CRC (both tumors and paired normal tissue), while being highly abundant in both diverticula and adenomas. Similar observations have been made in rectal cancer [70], and further suggests that a

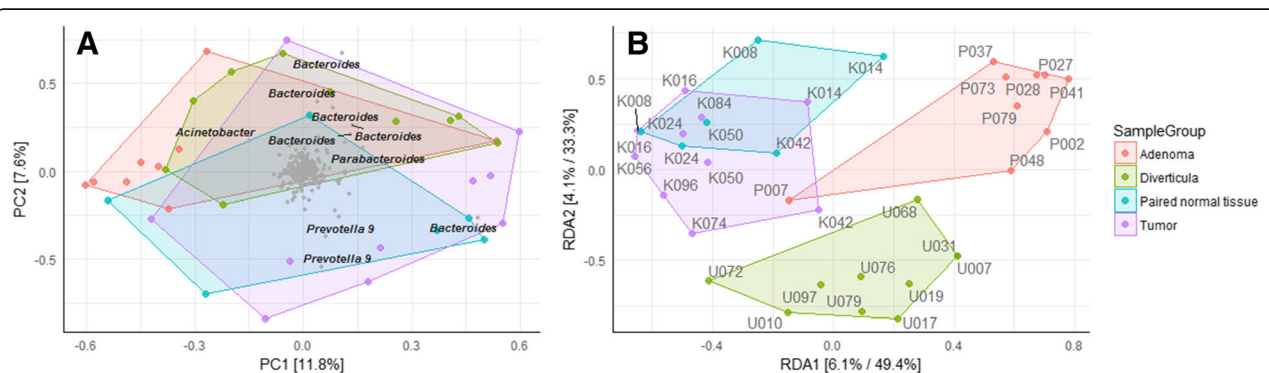
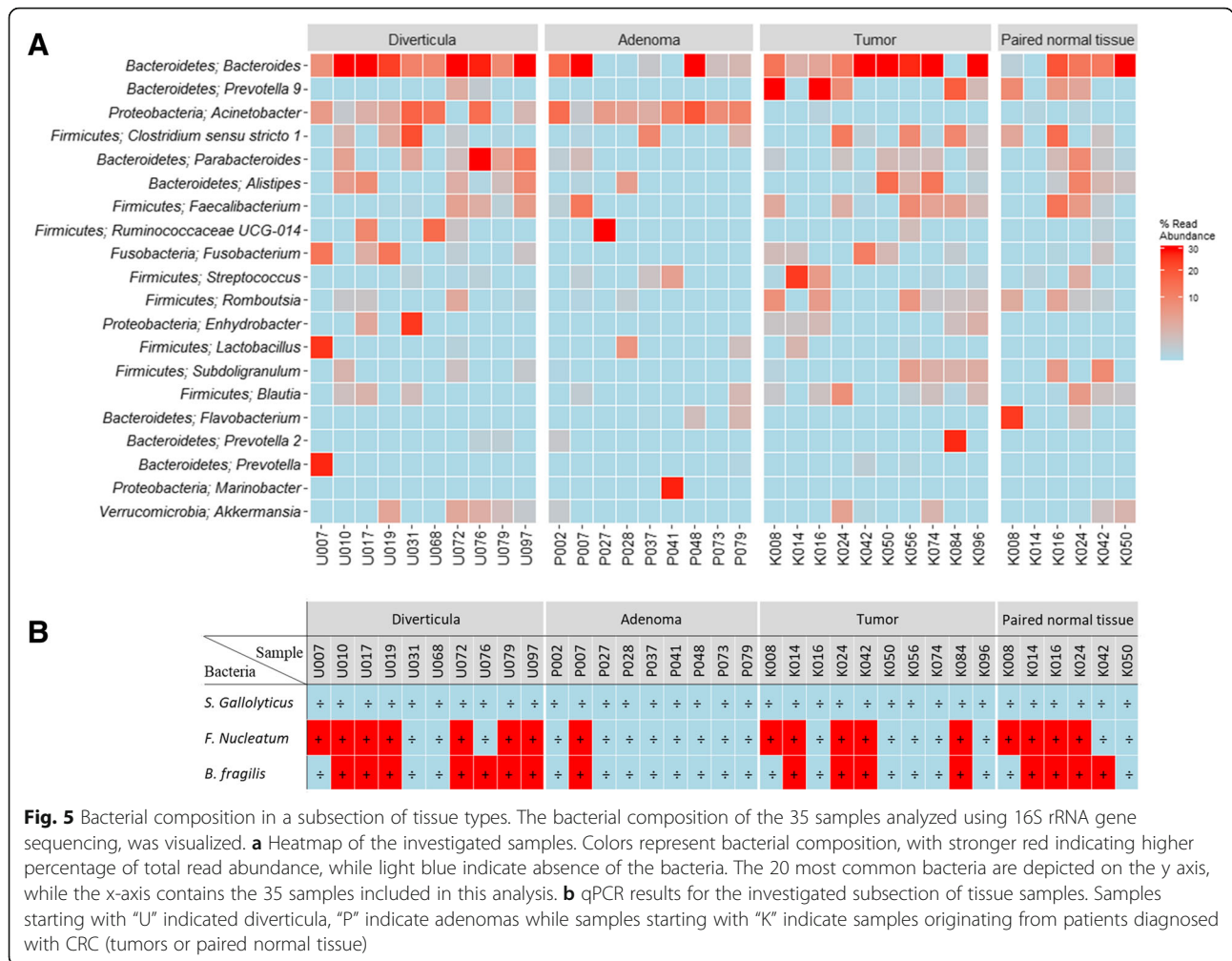


Fig. 4 Variation in bacterial composition between individual samples and tissue types. β -diversity was investigated in 35 tissue samples using **a**) PCA and subsequent **b**) RDA plots with Hellinger Distance of OTU abundances. Colored boxes represent different tissue types



distinct bacterial niche develops during the adenoma-tumor transition. In contrast to previous studies [41, 42], we did not observe a difference in the percentage of early and late stage CRC tumor samples positive for *F. nucleatum* or *B. fragilis*, indicating that these bacteria do not drive tumor progression. Finally, to elucidate the role of *F. nucleatum* and *B. fragilis* in initiation and progression of CRC, we investigated the 5 year risk of new cases of adenomas, CRC or death depending on bacterial status. In our study neither *F. nucleatum* nor *B. fragilis* affected the risk of death or the risk of developing new adenomas or CRC in either CRC, adenoma or diverticular disease patients. Overall our results suggest that the bacterial genus *Prevotella* and the species *F. nucleatum* and *B. fragilis* may play a role in the transition of adenomas to CRC, but not in initiation of adenomas nor in the progression from early to late stage colorectal tumors.

Two surprising observations were noted during this study. First, despite the noted association with CRC [14, 19, 34, 71], we did not detect *S. gallolyticus* in any of the

investigated tissue samples. The conflicting results could potentially be explained through ethnic differences in susceptibility to *S. gallolyticus* colonization of colorectal mucosa or geographical differences in *S. gallolyticus* distribution. This is supported by similar findings by Viljoen et al. [20] in a South African CRC population. Secondly, while several studies [21, 22], including the current study, utilize paired normal tissue obtained from CRC patients as a matched "healthy" control, we observed that the bacterial composition of tumor tissue and paired normal tissue overlapped considerably. While more samples are needed to validate this observation, it does question the validity of using paired normal tissue as healthy controls when investigating bacteria of CRC.

This study has a number of limitations. First, all samples used were fixed with formalin. Since formalin is known to affect DNA quality [72], this may have limited our ability to detect bacteria. Since all tissue samples were handled similarly, we do not expect the formalin fixation to affect the observed differences in bacterial load and prevalence between diagnoses. A second limitation involves the previously

reported difficulties in extracting DNA from gram-positive bacteria like *S. gallolyticus* [50]. The sequencing data revealed a high proportion of gram positive bacteria including other members of the *Streptococcus* genus. Thus, this limitation does not explain the lack of *S. gallolyticus* reported in this study. Finally, while the primers used in this study have low LODs compared to bacterial DNA, the LODs were established on purified DNA from bacteria, which would have a higher quality compared to FFPE bacterial DNA stored for up to 10 years. The true LOD of the primers in the examined tissue samples, could therefore be higher, as reported by Viljoen et al. [20]. This could prevent detection of low abundance bacteria, causing us to underestimate the bacterial colonization across all samples. This study had a specific focus on the bacterial species *F. nucleatum*, *B. fragilis* and *S. gallolyticus*. However, other studies have revealed other bacteria with an unique correlation with CRC, including *Escherichia coli* [63]. Future studies would need to include this bacteria as well.

Strengths of this study include the large number of samples included, the inclusion of precursor lesions and non-malignant tissue in addition to tumor and paired normal tissue as well as a follow-up investigation investigating the clinical relevance of the bacteria in addition to the bacterial status.

Conclusion

Our results do not support a role of *S. gallolyticus* in CRC in the Danish population. For *F. nucleatum* and *B. fragilis*, this study does not support a role in development of adenomas, although the bacteria may play a role in the adenoma-carcinoma transition. A potential role of the genera *Prevotella* and *Acinetobacter* in colorectal carcinogenesis was indicated, but warrants further studies.

Additional files

Additional file 1: qPCR results. Quantification cycles obtained using qPCR and clinical information for each clinical sample, investigated in this study. (XLSX 103 kb)

Additional file 2: Tissue samples used for 16S rRNA gene sequencing. Quantification cycles obtained using qPCR and clinical information for each clinical sample investigated using Illumina sequencing of the V4 region of the 16S rRNA gene. (XLSX 31 kb)

Additional file 3: 16S rRNA gene sequencing metadata. Information describing the samples included for bioinformatics analysis of the 16S rRNA gene sequencing data. (XLSX 11 kb)

Additional file 4: 16S rRNA gene sequences. FASTA file containing the DNA sequences generated using 16S rRNA gene sequencing. (FA 243 kb)

Additional file 5: 16S rRNA gene sequencing OTUtable. OTU table documenting numbers of different OTUs generated during the 16S rRNA gene sequencing. (TXT 163 kb)

Abbreviations

B. fragilis: *Bacteroides fragilis*; CRC: Colorectal cancer; DNA: Deoxyribonucleic acid; *F. nucleatum*: *Fusobacterium nucleatum*; FFPE: Formalin-fixed, paraffin-embedded; HE stain: Hematoxylin and eosin stain; OTU: Operational

Taxonomic Units; PCA: Principal component analysis; PCR: Polymerase chain reaction; qPCR: Quantitative real-time PCR; RDA: Redundancy analysis; RNA: Ribonucleic acid; rRNA: Ribosomal RNA; *S. gallolyticus*: *Streptococcus gallolyticus* subspecies *gallolyticus*

Acknowledgements

We would like to thank laboratory technicians Ann-Maria Jensen, Bente Wormstrup, Mette Skov Mikkelsen and Katrine Bech Hjort Lauritzen for excellent technical assistance. Furthermore, we would like to acknowledge the biotechnology students Celine Boure, Anna Hustedova, Monika Jonikaite, Zivile Kondrotaite and Patricia Riedlova for running the 16S rRNA gene sequencing.

Funding

This study was funded by a grant from the Innovation Fund Denmark. The Innovation Fund Denmark was neither involved in study design, sample collection, analysis and interpretation of data, nor in preparation of the manuscript.

Availability of data and materials

All data generated or analyzed during the current study are included in this published article and its Additional file (see Additional files 1 and 2). For data generated using 16S rRNA gene sequencing, the Metadata, OTU table and .fa file are included as Additional files (See Additional files 3, 4 and 5).

Authors' contributions

LPN, UTB, SES, and CBN designed the study; CBN performed experiments; UTB and CBN performed histological validations; CBN and SES analyzed the data; SES supervised the project; CBN and SES prepared the manuscript; LPN and UTB gave technical support and conceptual advice. All authors contributed to finalizing the manuscript, and have read and approved the manuscript.

Ethics approval and consent to participate

The study was approved by the ethical committee of the North Denmark Region (N-20150059) and reported for the Data Protection Agency. All samples were non-traceable, anonymized samples for which the Regional Ethical Committee waived the requirement for informed consent. The investigation has been conducted according to principles expressed in the Declaration of Helsinki.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Centre for Clinical Research, North Denmark Regional Hospital, Hjørring, Denmark. ²Department of Clinical Medicine, Aalborg University, Aalborg, Denmark. ³Department of Pathology, North Denmark Regional Hospital, Hjørring, Denmark. ⁴Biobank and Biomarkers, Statens Serum Institut, Copenhagen, Denmark.

Received: 31 October 2018 Accepted: 3 April 2019

Published online: 29 April 2019

References

1. L a T, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. *CA Cancer J Clin*. 2015;00:1–22.
2. Muto T, Bussey H, Morson B. The evolution of cancer of the colon and rectum. *Cancer*. 1975;36:2251–70.
3. Volgestein B, Fearon E. A genetic model for colorectal tumorigenesis. *Cell*. 1990;61:759–67. [https://doi.org/10.1016/0092-8674\(90\)90186-l](https://doi.org/10.1016/0092-8674(90)90186-l).
4. Hadac JN, Leystra AA, Paul Olson TJ, Maher ME, Payne SN, Yueh A, et al. Colon tumors with the simultaneous induction of APC, KRAS and PIK3CA mutations still progress through the adenoma-to-carcinoma sequence.

- Cancer Prev Res (Phila). 2015. <https://doi.org/10.1158/1940-6207.CAPR-15-0003>.
5. S a B, Day NE, Luben R, Ferrari P, Slimani N, Norat T, et al. Dietary fibre in food and protection against colorectal cancer in the European prospective investigation into Cancer and nutrition (EPIC): an observational study. *Lancet*. 2003;361:1496–501.
 6. Norat T, Bingham S, Ferrari P, Slimani N, Jenab M, Mazuir M, et al. Meat, fish, and colorectal cancer risk: the European prospective investigation into cancer and nutrition. *J Natl Cancer Inst*. 2005;97:906–16.
 7. Willett WC, Stampfer MJ, Colditz GA, Rosner BA, Speizer FE. Relation of meat, fat, and fiber intake to the risk of colon cancer in a prospective study among women. *N Engl J Med*. 1990;323:1664–72.
 8. Nagata C, Shimizu H, Kametani M, Takeyama N, Ohnuma T, Matsushita S. Cigarette smoking, alcohol use, and colorectal adenoma in Japanese men and women. *Dis Colon Rectum*. 1999;42:337–42.
 9. Pedersen A, Johansen C, Grønbaek M. Relations between amount and type of alcohol and colon and rectal cancer in a Danish population based cohort study. *Gut*. 2003;52:861–7.
 10. Etzioni R, Urban N, Ramsey S, McIntosh M, Schwartz S, Reid B, et al. The case for early detection. *Nat Rev Cancer*. 2003;3:243–52.
 11. Sundhedsstyrelsen. Anbefalinger vedrørende screening for tyk- & endetarmskræft. 2010.
 12. Hewitson P, Glasziou P, Watson E, Towler B, Irwig L. Cochrane systematic review of colorectal cancer screening using the fecal occult blood test (hemocult): an update. *Am J Gastroenterol*. 2008;103:1541–9.
 13. WC MC, Mason JM. Enterococcal endocarditis associated with carcinoma of the sigmoid; report of a case. *J Med Assoc State Ala*. 1951;21:162–6.
 14. Hoppes WL, Lerner PI. Nonenterococcal group-D streptococcal endocarditis caused by *Streptococcus bovis*. *Ann Intern Med*. 1974;81:588–93.
 15. Shanani S, Gumaa SA, Sandström G, Abd H. Significant association of *Streptococcus bovis* with malignant gastrointestinal diseases. *Int J Microbiol*. 2011;2011:1–5.
 16. Gupta a, Madani R, Mukhtar H. *Streptococcus bovis* endocarditis, a silent sign for colonic tumour. *Color Dis*. 2010;12:164–71.
 17. Corredoira-Sánchez J, García-Garrote F, Rabunal 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.
 18. Sharara AI, Abou Hamdan T, Malli A, El-Halabi MM, Hashash JG, Ghaith OA, et al. Association of *Streptococcus bovis* endocarditis and advanced colorectal neoplasia: a case-control study. *J Dig Dis*. 2013;14:382–7. <https://doi.org/10.1111/1751-2980.12059>.
 19. Abdulmir AS, Hafidh RR, Bakar FA. Molecular detection, quantification, and isolation of *Streptococcus gallolyticus* bacteria colonizing colorectal tumors: inflammation-driven potential of carcinogenesis via IL-1, COX-2, and IL-8. *Mol Cancer*. 2010;9:249.
 20. Viljoen KS, Dakshinamurthy A, Goldberg P, Blackburn JM. Quantitative profiling of colorectal cancer-associated bacteria reveals associations between fusobacterium spp., enterotoxigenic *Bacteroides fragilis* (ETBF) and clinicopathological features of colorectal cancer. *PLoS One*. 2015;10:e0119462. <https://doi.org/10.1371/journal.pone.0119462>.
 21. Castellarin M, Warren RL, Freeman JD, Dreolini L, Krzywinski M, Strauss J, et al. Fusobacterium nucleatum infection is prevalent in human colorectal carcinoma. *Genome Res*. 2012;22:299–306.
 22. 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;22:292–8. <https://doi.org/10.1101/gr.126573.111>.
 23. Marchesi JR, Dutilh BE, Hall N, Peters WHM, Roelofs R, Boleij A, et al. Towards the human colorectal cancer microbiome. *PLoS One*. 2011;6:e20447.
 24. Bullman S, Pedamallu CS, Sicinska E, Clancy TE, Zhang X, Cai D, et al. Analysis of fusobacterium persistence and antibiotic response in colorectal cancer. *Science*. 2017;358:1443–8. <https://doi.org/10.1126/science.aal5240>.
 25. Mima K, Sukawa Y, Nishihara R, Qian ZR, Yamauchi M, Inamura K, et al. Fusobacterium nucleatum and T cells in colorectal carcinoma. *JAMA Oncol*. 2015;1:653. <https://doi.org/10.1001/jamaoncol.2015.1377>.
 26. Kosumi K, Hamada T, Koh H, Borowsky J, Bullman S, Twombly TS, et al. The amount of Bifidobacterium genus in colorectal carcinoma tissue in relation to tumor characteristics and clinical outcome. *Am J Pathol*. 2018;48(Suppl 1):S67–9. <https://doi.org/10.1016/j.ajpath.2018.08.015>.
 27. Goodwin AC, Destefano Shields CE, Wu S, Huso DL, Wu X, Murray-Stewart TR, et al. Polyamine catabolism contributes to enterotoxigenic *Bacteroides fragilis*-induced colon tumorigenesis. *Proc Natl Acad Sci U S A*. 2011;108:15354–9. <https://doi.org/10.1073/pnas.1010203108>.
 28. Ulger Toprak N, 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. <https://doi.org/10.1111/j.1469-0691.2006.01494.x>.
 29. Wu S, Rhee K-J, Zhang M, Franco A, Sears CL. *Bacteroides fragilis* toxin stimulates intestinal epithelial cell shedding and gamma-secretase-dependent E-cadherin cleavage. *J Cell Sci*. 2007;120(Pt 11):1944–52. <https://doi.org/10.1242/jcs.03455>.
 30. Wu S, Morin PJ, Maouyo D, Sears CL. *Bacteroides fragilis* enterotoxin induces c-Myc expression and cellular proliferation. *Gastroenterology*. 2003;124:392–400.
 31. Wei Z, Cao S, Liu S, Yao Z, Sun T, Li Y, et al. Could gut microbiota serve as prognostic biomarker associated with colorectal cancer patients' survival? A pilot study on relevant mechanism. *Oncotarget*. 2016;7(29):46158–72. <https://doi.org/10.18632/oncotarget.10064>.
 32. Kostic AD, Chun E, Robertson L, Glickman JN, Gallini CA, Michaud M, et al. Fusobacterium nucleatum potentiates intestinal tumorigenesis and modulates the tumor-immune microenvironment. *Cell Host Microbe*. 2013;14:207–15. <https://doi.org/10.1016/j.chom.2013.07.007>.
 33. Zeller G, Tap J, Voigt AY, Sunagawa S, Kultima JR, Costea PI, et al. Potential of fecal microbiota for early-stage detection of colorectal cancer. *Mol Syst Biol*. 2014;10:766. <https://doi.org/10.15252/msb.20145645>.
 34. Rezasoltani S, Asadzadeh Aghdaei H, Dabiri H, Akhavan Sepahi A, Modarresi MH, Nazemalhosseini Mojarad E. The association between fecal microbiota and different types of colorectal polyp as precursors of colorectal cancer. *Microb Pathog*. 2018;124:244–9. <https://doi.org/10.1016/j.micpath.2018.08.035>.
 35. 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. <https://doi.org/10.1371/journal.pone.0053653>.
 36. Park CH, Han DS, Oh YH, Lee AR, Lee YR, Eun CS. Role of fusobacteria in the serrated pathway of colorectal carcinogenesis. *Sci Rep*. 2016;6:1–8. <https://doi.org/10.1038/srep25271>.
 37. Pagnini C, Corleto VD, Mangoni ML, Pillozzi E, Torre MS, Marchese R, et al. Alteration of local microflora and α -defensins hyper-production in colonic adenoma mucosa. *J Clin Gastroenterol*. 2011;45:602–10.
 38. Shen XJ, Rawls JF, Randall T, Burcal L, Mpande CN, Jenkins N, et al. Molecular characterization of mucosal adherent bacteria and associations with colorectal adenomas. *Gut Microbes*. 2010;1:138–47. <https://doi.org/10.4161/gmic.1.3.12360>.
 39. Yu J, Chen Y, Fu X, Zhou X, Peng Y, Shi L, et al. Invasive fusobacterium nucleatum may play a role in the carcinogenesis of proximal colon cancer through the serrated neoplasia pathway. *Int J Cancer*. 2016;139:1318–26.
 40. Feng Q, Liang S, Jia H, Stadlmayr A, Tang L, Lan Z, et al. Gut microbiome development along the colorectal adenoma-carcinoma sequence. *Nat Commun*. 2015;6:6528.
 41. Yu J, Feng Q, Wong SH, Zhang D, Liang QY, Qin Y, et al. Metagenomic analysis of faecal microbiome as a tool towards targeted non-invasive biomarkers for colorectal cancer. *Gut*. 2017;66:70–8. <https://doi.org/10.1136/gutjnl-2015-309800>.
 42. Nakatsu G, Li X, Zhou H, Sheng J, Wong SH, Wu WKK, et al. Gut mucosal microbiome across stages of colorectal carcinogenesis. *Nat Commun*. 2015;6:8727. <https://doi.org/10.1038/ncomms9727>.
 43. Gevers D, Kugathasan S, Denson LA, Vázquez-Baeza Y, Van Treuren W, Ren B, et al. The treatment-naive microbiome in new-onset Crohn's disease. *Cell Host Microbe*. 2014;15:382–92. <https://doi.org/10.1016/j.chom.2014.02.005>.
 44. Promeza. Compare Two Prevalences: Sample Size. http://www.promesa.co.nz/help/EP_ss_compare_prev.htm. Accessed 1 Sept 2015.
 45. 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.
 46. Karlsen F, Kalantari M, Chitemerere M, Johansson B, Hagmar B. Modifications of human and viral deoxyribonucleic acid by formaldehyde fixation. *Lab Invest*. 1994;71:604–11 <http://www.ncbi.nlm.nih.gov/pubmed/7967515>.
 47. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods*. 2001;25:402–8.

48. Saghaei M. An overview of randomization and minimization programs for randomized clinical trials. *J Med Signals Sens.* 2011;1:55–61 <http://www.ncbi.nlm.nih.gov/pubmed/22606659>. Accessed 13 Oct 2017.
49. Illumina. 16S Metagenomic Sequencing Library Preparation, Part # 15044223 Rev. B. 2015;1–28. http://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf. Accessed 10 Feb 2018.
50. Albertsen M, Karst SM, Ziegler AS, Kirkegaard RH, Nielsen PH. Back to basics - the influence of DNA extraction and primer choice on phylogenetic analysis of activated sludge communities. *PLoS One.* 2015;10:1–15.
51. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, et al. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc Natl Acad Sci.* 2011; 108(Supplement_1):4516–22. <https://doi.org/10.1073/pnas.1000080107>.
52. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, et al. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J.* 2012;6:1621–4. <https://doi.org/10.1038/ismej.2012.8>.
53. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 2014;30:2114–20. <https://doi.org/10.1093/bioinformatics/btu170>.
54. Edgar RC. UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat Methods.* 2013;10:996–8. <https://doi.org/10.1038/nmeth.2604>.
55. Wang Q, Garrity GM, Tiedje JM, Cole JR. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol.* 2007;73:5261–7. <https://doi.org/10.1128/AEM.00062-07>.
56. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods.* 2010;7:335–6. <https://doi.org/10.1038/nmeth.f303>.
57. McLlroy SJ, Saunders AM, Albertsen M, Nierychlo M, McLlroy B, Hansen AA, et al. MiDAS: the field guide to the microbes of activated sludge. *Database.* 2015;2015:1–8.
58. R Development Core Team. R: A Language and Environment for Statistical Computing. the R Foundation for Statistical Computing. <https://www.r-project.org/>. Accessed 1 July 2018.
59. Andersen KSS, Kirkegaard RH, Karst SM, Albertsen M. ampvis2: an R package to analyse and visualise 16S rRNA amplicon data. *bioRxiv.* 2018:299537. <https://doi.org/10.1101/299537>.
60. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15:1–21.
61. Benjamini Y, Yosef H. Determination of water content in meat pâté by karl fischer titration and its moisture sorption characteristics. *J R Stat Soc Ser B.* 1995;57:289–300.
62. 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.
63. Wu S, Rhee K-J, Albesiano E, Rabizadeh S, Wu X, Yen H-R, et al. A human colonic commensal promotes colon tumorigenesis via activation of T helper type 17 T cell responses. *Nat Med.* 2009;15:1016–22. <https://doi.org/10.1038/nm.2015>.
64. Arthur JC, Perez-Chanona E, Mühlbauer M, Tomkovich S, Uronis JM, Fan T, et al. Intestinal inflammation targets cancer-inducing activity of the microbiota. *Science.* 2012;338:120–3. <https://doi.org/10.1126/science.1224820>.
65. Brenner H, Hoffmeister M, Stegmaier C, Brenner G, Altenhofen L, Haug U. Risk of progression of advanced adenomas to colorectal cancer by age and sex: estimates based on 840 149 screening colonoscopies. *Gut.* 2007;56:1585–9.
66. Sobhani I, Tap J, Roudot-Thoraval F, Roperch JP, Letulle S, Langella P, et al. Microbial dysbiosis in colorectal cancer (CRC) patients. *PLoS One.* 2011;6: e16393.
67. Flemer B, Lynch DB, Brown JMR, Jeffery IB, Ryan FJ, Claesson MJ, et al. Tumour-associated and non-tumour-associated microbiota in colorectal cancer. *Gut.* 2017;66:633–43.
68. Rubinstein MR, Wang X, Liu W, Hao Y, Cai G, Han YW. Fusobacterium nucleatum promotes colorectal carcinogenesis by modulating E-cadherin/ β -catenin signaling via its FadA Adhesin. *Cell Host Microbe.* 2013;14:195–206. <https://doi.org/10.1016/j.chom.2013.07.012>.
69. Larsen JM. The immune response to Prevotella bacteria in chronic inflammatory disease. *Immunology.* 2017;151:363–74. <https://doi.org/10.1111/imm.12760>.
70. Thomas AM, Jesus EC, Lopes A, Aguiar S, Begnami MD, Rocha RM, et al. Tissue-associated bacterial alterations in rectal carcinoma patients revealed by 16S rRNA community profiling. *Front Cell Infect Microbiol.* 2016:1–13. <https://doi.org/10.3389/fcimb.2016.00179>.
71. 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.
72. Imrit K, Goldfischer M, Wang J, Green J, Levine J, Lombardo J, et al. Identification of bacteria in formalin-fixed, paraffin-embedded heart valve tissue via 16S rRNA gene nucleotide sequencing. *J Clin Microbiol.* 2006;44:2609–11.
73. Bufill JA. Colorectal cancer: evidence for distinct genetic categories based on proximal or distal tumor location. *Ann Intern Med.* 1990;113:779–88 <http://www.ncbi.nlm.nih.gov/pubmed/2240880>.
74. International Union against Cancer (UICC). Sobin LH, Gospodarowicz MK, Wittekind C, editors. TNM : classification of malignant tumours. 7th ed. Chichester: Wiley-Blackwell; 2010.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

