TITLE: Fusobacterium in colonic flora and molecular features of colorectal carcinoma

Authors: Tomomitsu Tahara\textsuperscript{1,2}, Eiichiro Yamamoto\textsuperscript{3,4}, Hiromu Suzuki\textsuperscript{4}, Reo Maruyama\textsuperscript{4}, Woonbok Chung\textsuperscript{1}, Judith Garriga\textsuperscript{1}, Jaroslav Jelinek\textsuperscript{1}, Hiro-o Yamano\textsuperscript{5}, Tamotsu Sugai\textsuperscript{8}, Byonggu An\textsuperscript{7}, Imad Shureiqi\textsuperscript{8}, Minoru Toyota\textsuperscript{4}, Yutaka Kondo\textsuperscript{7}, Marcos R. H. Estécio\textsuperscript{9,10}, Jean-Pierre J. Issa\textsuperscript{1}

1 Fels Institute for Cancer Research & Molecular Biology, Temple University School of Medicine, Philadelphia PA, USA.
2 Department of Gastroenterology, Fujita Health University School of Medicine, Toyoake, Japan
3 First Department of Internal Medicine, Sapporo Medical University, Sapporo, Japan.
4 Department of Molecular Biology, Sapporo Medical University, Sapporo, Japan.
5 Department of Gastroenterology, Akita Red Cross Hospital, Akita, Japan.
6 Department of Pathology, Iwate Medical University, Morioka, Japan
7 Division of Molecular Oncology, Aichi Cancer Center Research Institute, Nagoya, Japan.
8 Department of Clinical Cancer Prevention, Division of OVP, Cancer Prevention and Population Sciences, The University of Texas MD Anderson Cancer Center, Houston, TX
9 Department of Molecular Carcinogenesis, The University of Texas MD Anderson Cancer Center, Smithville, TX, USA 10 Center for Cancer Epigenetics, The University of Texas MD Anderson Cancer Center, Houston, TX, USA.

Correspondence should be addressed to T.T. (tomomiccyu0720@yahoo.co.jp)

Running title: Fusobacterium and CIMP colon cancer

Keywords: Fusobacterium, colorectal cancer, DNA methylation, CpG island methylator phenotype, exome sequencing.
Author contributions

T.T., M.R.H.E. and J.-P.J.I conceived and designed the study.

T.T. designed and performed experiments, analyzed data. Experiments were also designed by M.R.H.E., J.-P.J.I., Y.K. and M.T.

M.R.H.E. performed bioinformatics analysis.

Data were additionally analyzed and interpreted by W.C., J.G., J.J., and J.-P.J.I.

Clinical samples were collected by E.Y., H.S., R.M., Y.K., B.A., S.I., and H.Y.

Histopathological examination was performed by T.S.

Molecular analysis was performed by E.Y., H.S., R.M., Y.K., B.A., and M.T.

T.T., M.R.H.E and J.-P.J.I. wrote the manuscript.

All authors reviewed and commented on the manuscript.

Acknowledgements

This work was supported by National Institutes of Health grants CA098006 and CA158112 to JPJI and the G.S. Hogan Gastrointestinal Research Fund at MD Anderson Cancer Center to M.R.H.E. JPI is an American Cancer Society Clinical Research professor supported by a generous gift from the F. M. Kirby Foundation.

Competing financial interests

The authors declare no competing financial interests.
Abstract

*Fusobacterium* species are part of the gut microbiome in humans. Recent studies have identified over-representation of *Fusobacterium* in colorectal cancer (CRC) tissues but it is not yet clear whether this is pathogenic or simply an epiphenomenon. In this study, we evaluated the relationship between *Fusobacterium* status and molecular features in CRCs through quantitative real-time PCR in 149 CRC tissues, 89 adjacent normal appearing mucosae and 72 colonic mucosae from cancer-free individuals. Results were correlated with CpG island methylator phenotype (CIMP) status, microsatellite instability (MSI) and mutations in *BRAF, KRAS, TP53, CHD7* and *CHD8*. Whole exome capture sequencing data were also available in 11 cases. *Fusobacterium* was detectable in 111/149 (74%) CRC tissues and heavily enriched in 9% (14/149) of the cases. As expected, *Fusobacterium* was also detected in normal appearing mucosae from both cancer and cancer-free individuals but the amount of bacteria was much lower compared to CRC tissues (a mean of 250-fold lower for Pan-*fusobacterium*). We found the *Fusobacterium*-high CRC group (FB-high) to be associated with CIMP positivity (*p*=0.001), *TP53* wild type (*p*=0.015), *hMLH1* methylation positivity (*p*=0.0028), MSI (*p*=0.018) and *CHD7/8* mutation positivity (*p*=0.002). Among the 11 cases where whole exome sequencing data was available, two that were FB-high cases also had the highest number of somatic mutations (a mean of 736 per case in FB-high vs. 225 per case in all others). Taken together, our findings show that *Fusobacterium* enrichment is associated with specific molecular subsets of CRCs, offering support for a pathogenic role in CRC for this gut microbiome component.
Introduction

The non-spore-forming, anaerobic Gram-negative bacteria, *Fusobacterium* is part of the normal flora in the human mouth and gut mucosa. *Fusobacterium* species are highly heterogeneous and some species have been recognized as opportunistic pathogens implicated in inflammatory diseases of both the mouth, such as periodontitis, and the gut, such as appendicitis and inflammatory bowel diseases\(^1\)-\(^5\). Two recent studies have linked *Fusobacterium* species with colorectal cancer (CRC). These studies demonstrated that *Fusobacterium nucleatum* (*F. nucleatum*) and whole *Fusobacterium* species (*Pan-fusobacterium*) were abundant in CRC tissues compared to adjacent normal mucosa\(^6\),\(^7\). Several infectious bacteria and viruses were previously associated with neoplasia such as human papillomavirus in cervical cancer\(^8\), Kaposi’s sarcoma-associated herpesvirus in Kaposi’s sarcoma\(^9\) and Epstein–Barr virus in lymphomas and gastric cancer\(^10\). *Fusobacterium* in CRC provided a novel concept in that a part of the normal intestinal microflora may be relevant to tumorigenesis. However, the previous studies could not exclude the possibility that the presence of *Fusobacterium* in CRC is an epiphenomenon, related to local changes triggered by the neoplastic process.

CRCs are characterized by specific genetic and epigenetic lesions. Besides common mutations in *TP53*, *KRAS* and *APC* genes\(^11\),\(^12\), epigenetic alterations in CRCs are frequent, particularly gene promoter DNA methylation. Classification of CRCs according to mutation and DNA methylation status has identified distinct subtypes based on the CpG Island Methylator Phenotype (CIMP)\(^13\). Typical high-level CIMP (CIMP-high, CIMP1) CRCs are associated with microsatellite instability (MSI) through epigenetic silencing of a mismatch repair gene *MLH1*, as well as *BRAF* mutation. Frequent mutation in chromatin regulator genes, notably, *CHD7* and *CHD8*, members of the chromodomain helicase/ ATP-dependent chromatin remodeling family were recently also discovered in CIMP1 CRCs\(^14\). Low-level CIMP (CIMP-low, CIMP2) is characterized by methylation of a limited group of genes and mutation in *KRAS*. CIMP-negative
cases have less frequent methylation changes and very frequent \textit{TP53} mutation and chromosomal instability\textsuperscript{15,16}.

Since CRCs have heterogeneous molecular and clinical features\textsuperscript{15-19}, we investigated whether \textit{Fusobacterium} status is associated with different subtypes of CRCs. We found that \textit{Fusobacterium}-high cases have a unique genetic and epigenetic profile, supporting potential links between the gut microbiome and molecular features of CRC.
Materials and Methods

Tissue samples

We used genomic DNA samples of 149 primary CRCs and 89 normal-appearing adjacent tissues from patients undergoing surgery or colonoscopy at the Johns Hopkins Hospital, MD Anderson Cancer Center, Sapporo medical University, Akita Red Cross Hospital and Aichi Cancer Center Research Institute. All CRCs used in this study were characterized previously for CIMP (all cases), MSI (n=113), BRAF mutation (n=144), KRAS mutation (n=148) and TP53 mutation status (n=143). CHD7 and CHD8 mutation were also characterized in 100 out of 149 cases. Genomic DNA was also obtained from 72 colonic biopsies in 65 cancer free subjects undergoing colonoscopy at the MD Anderson Cancer Center and Fujita Health University Hospital. 52 out of 72 these samples were from distal colon (descending and sigmoid colon, and rectum) and the remaining 20 were from the proximal colon (cecum, ascending and transverse colon). Samples were collected in accordance with institutional policies and written informed consent for tissue collection was provided by all the participants.

Quantitative PCR analysis for Fusobacterium

Quantitative real-time PCR was performed using the Universal PCR Master Mix (Bio-Rad) and StepOnePlus™ Real-Time PCR System (Applied Biosystems). F. nucleatum and pan-fusobacterium TaqMan primer/probe sets used in this study were described previously. The cycle threshold (Ct) values for F. nucleatum and pan-fusobacterium were normalized to the amount of human DNA in each reaction by using a primer/probe set for the reference gene, prostaglandin transporter (PGT), as described previously. All assays were done in duplicate and we averaged the results.
**DNA methylation analysis for cancer free subjects**

Bisulfite-treated genomic DNA from cancer free subjects was used to evaluate the methylation status of 7 CpG islands (ER, SFRP1, MYOD1, MGMT, SLC16A2, SPOCK2 and N33) using the primers listed in supplementary Table1. Bisulfite treatment of DNA was performed with an EpiTect bisulfite kit (Qiagen) according to the manufacturer’s protocol. Pyrosequencing was carried out using a Pyro Mark Q96 MD system with a Pyro-Gold reagent Kit (QIAGEN), and the results were analyzed using PyroMark Q96 ID software version 1.0 (QIAGEN).

**Whole exome capture sequencing and Gene Ontology analysis**

Genomic DNA specimens from 11 colorectal tumors and their adjacent normal tissues were submitted to Otogenetics Corporation (Norcross, GA USA) for exome capture and sequencing. Genomic DNAs were fragmented and then tested for size distribution and concentration. Illumina libraries were made using Next reagents (New England Biolabs, Ipswich, MA USA), and the resulting libraries were subjected to exome enrichment using NimbleGen SeqCap EZ Human Exome Library v2.0 (Roche NimbleGen, Inc., Madison, WI USA). The samples were then sequenced on an Illumina HiSeq2000 (Illumina, Inc., San Diego, CA USA), which generated paired-end reads of 90 or 100 nucleotides. All paired samples (tumor and normal) were sequenced on the same run, using same depth and coverage. Read results from both replicates were combined in the final analysis. Data were analyzed for quality, exome coverage, and exome-wide single nucleotide polymorphism (SNP)/Indel using the platform provided by DNAnexus (Mountain View, CA USA). We excluded all variants with a PHRED-encoded probability score lower than 35, those that were present in the DNA of the corresponding normal samples (thus excluding germline events), and those that were not in coding regions, as well as silent changes and known SNPs (except for clinically associated
SNPs). DNAnexus Genome Browser was used for visual validation of all potential somatic mutations to ensure that they were present in forward and reverse strands. The clinicopathological data for the studied cases, a detailed protocol of data analysis, summary of sequencing statistics and somatic mutations list for all samples can be found in this manuscript. Functional enrichment of mutated genes was determined by the gene ontology analysis using DAVID Bioinformatics Resources 6.7 (http://david.abcc.ncifcrf.gov/). P-values were corrected for multiple hypothesis testing using the Benjamini method.

Statistical analysis

Continuous variables among matched samples (cancer and normal tissues) were examined using the Wilcoxon signed-rank test. Continuous variables among two and three different groups were examined using the Student's t-test and One-way ANOVA, respectively. Categorical variables among two or three different groups were examined using two-sided Fisher's exact test. Two sided P value < .05 was considered statistically significant.
Results

Clinicopathologic characteristics of CRCs

We studied 104 CRCs selected based on sample availability and subsequently added 26 CIMP1, 18 CIMP2 and 1 CIMP-negative cases to expand this cohort. In total, these cases consisted of 60 CIMP-negative, 42 CIMP1 and 47 CIMP2 tumors. Clinicopathologic characteristics are shown in Table 1. As expected, CIMP1 cases presented at a higher age and were principally located in the proximal colon. CIMP1 cases were characterized by a higher incidence of mutations in \textit{BRAF} and MSI and rare mutations in \textit{KRAS} and \textit{TP53}. The CIMP2 cases were characterized by a higher incidence of mutations in \textit{KRAS} and rare MSI. The CIMP-negative cases were characterized by a higher incidence of mutations in \textit{TP53} and rare MSI.

Detection of Fusobacterium in CRC tissues and their adjacent mucosa

Among 149 CRC tumor tissues, \textit{F. nucleatum} and \textit{pan-fusobacterium} were detectable in 78 (52.3%) and 110 (73.8%) cases, respectively and 111 patients (74.4%) had either \textit{F. nucleatum} or \textit{pan-fusobacterium} detectable. Among 89 adjacent normal colonic mucosae, \textit{F. nucleatum} and \textit{pan-fusobacterium} were detectable in 27 (30.3%) and 47 (52.8%) cases, respectively (Supplementary Fig. 1). To determine the abundance of \textit{Fusobacterium} in CRC tissues, we initially compared the amount of bacteria in 89 matched tumor tissues and normal mucosae. In agreement with previous studies\textsuperscript{6,7}, we found significant enrichment of both \textit{F. nucleatum} and \textit{pan-fusobacterium} in CRC tissues compared to adjacent normal mucosae (approximate enrichment of \textit{F. nucleatum}, 3600 fold and \textit{pan-fusobacterium}, 250 fold, both \textit{p} values <0.0001 by the Wilcoxon signed-rank test, Fig.1). Over-representation of both \textit{F.
nucleatum and pan-fusobacterium in tumor versus matched normal specimens was found in more than half of the cases (51%, 45/89 for \textit{F. nucleatum} and 62%, 55/89 for \textit{pan-fusobacterium}).

\textit{Association between fusobacterium high and clinical and molecular characteristics of CRC}

The amount of \textit{F. nucleatum} and \textit{pan-fusobacterium} in detectable cases varied considerably among the samples. \textit{Pan-fusobacterium} was more commonly detected, being measurable in 74%. For both \textit{F. nucleatum} and \textit{pan-fusobacterium}, the amount of bacteria in measurable cases had an approximately Gaussian distribution, with over representation of bacteria-high cases. Based on this, we set cut-off values of 0.01 and 1 (\(2^{-\Delta \text{Ct}}\)) for \textit{F. nucleatum} and \textit{pan-fusobacterium} and identified 8 (5.4%) and 14 (9.4%) cases as having a high amount of \textit{F. nucleatum} and \textit{pan-fusobacterium}, respectively (Supplementary Fig. 2). Since \textit{F. nucleatum} and \textit{pan-Fusobacterium} status was highly correlated in both cancer and normal tissues (\(p<0.0001\), Supplementary Table 2), we defined a high amount of \textit{Fusobacterium} (FB-high) as those cases with either high \textit{F. nucleatum} or \textit{pan-fusobacterium} or both. In cancer tissues, all 8 cases with high \textit{F. nucleatum} were included in high \textit{pan-fusobacterium} cases. Therefore, all FB-high cases (n=14) corresponded to high \textit{pan-fusobacterium} cases. (Supplementary Table 2, Fig. 2). On average, these cases had 250 fold enrichment of \textit{pan-fusobacterium} when compared to the overall average of the other cancer cases. We next analyzed clinico-pathologic correlations of FB-high status.

The prevalence of FB-high was significantly elevated in CIMP-positive CRCs including CIMP1 (9/42, 21.4%) and CIMP2 CRCs (5/47, 10.6%) compared to CIMP-negative cases (0/64, 0%, \(p=0.001\)). Consistent with this, FB-high was significantly associated with molecular features that are common in CIMP CRCs, such as \textit{TP53} wild type (\(p=0.015\), \textit{hMLH1} methylation
positivity ($p=0.0028$) and MSI ($p=0.018$) (Table 2). On the other hand, prevalence of fusobacterium measurable cases were similar among CIMP1, CIMP2 and CIMP-negative cases for both F. nucleatum and pan-fusobacterium (all $p$ values >0.05, data not shown). We also found a significant association between FB-high and CHD7/8 mutation positivity (CHD7: $p=0.025$, CHD8: $p=0.035$ and CHD7/8 mutation: $p=0.002$). CHD7 and CDH8 are members of the chromodomain helicase/ATP-dependent chromatin remodeling family and both are commonly mutated in CIMP-positive CRCs in our recent study$^{14}$. Since CIMP-positive CRCs are more common in proximal colon and it is conceivable that the gut microbiome differs by site, we next assessed whether FB-high is associated with CIMP-positive CRCs in the proximal colon. Among 72 proximal CRCs, FB-high was significantly associated with CIMP ($p=0.047$). FB-high was also associated with CHD7/8 mutation ($p=0.046$) and older age ($p=0.01$), while weak associations were also found between FB-high and TP53 wild type status ($p=0.05$), hMLH1 methylation positivity ($p=0.05$) and CHD7 mutation ($p=0.06$) (Supplementary Table 4). We also investigated whether FB-high is associated with any clinical or molecular features within CIMP1 CRCs but found no significant correlations (Supplementary Table 5).

Whole exome capture and sequencing data was available for 11 CRCs and their matched normal colonic tissues$^{14}$. The 11 CRCs consisted of 8 CIMP1, 1 CIMP2 and 2 CIMP-negatives, and 2 of CIMP1 CRCs were classified as FB-high. This technology determines the sequence of ~30,000 coding genes, based on RefSeq, CCDS and miR base. There were 3495 non-silent somatic mutations in 2913 genes. The somatic mutations in the two FB-high (mean 736) was higher than that seen in CIMP1 with low/undetectable FB (mean 302, range 94 to 436) and CIMP2/CIMP-negative with low/undetectable FB presented the lowest somatic mutation rate (mean 71, range 24 to 122). These differences were statistically significant ($p=0.003$) (Fig.3). We also compared the distribution of different types of mutations (non-synonymous, stop codon and frame shift) and the context of the single base substitution mutations. Although
CIMP-1 CRCs had increased mutations in polynucleotide tracts, there was no difference in the types of mutations or the context of the single base substitution mutations across the different CIMP and *Fusobacterium* status. Non-synonymous, C to T and G to A transitions within the CpG sites were the most frequent in all the samples\(^{14}\).

To further evaluate functional differences of gene mutations among FB-high cases, we next performed Gene Ontology analysis to determine whether there was an enrichment for specific functional categories among the mutated genes in FB-high cases. This analysis showed that mutated genes in FB-high cases frequently encoded genes related to nervous system development. Interestingly, this functional category is not represented among the genes exclusively mutated in CIMP1 with low/undetectable FB and CIMP2/CIMP-negative with low/undetectable FB nor among the genes mutated in both tumor categories (Supplementary Table 6, 7). However, the number of cases available for analysis is small and these conclusions need confirmation in other datasets.

**Detection of *Fusobacterium* in non-neoplastic colonic mucosa**

Although the amount was much lower than that of cancer tissues (Fig.1), the amount of *F. nucleatum* and *pan-fusobacterium* in adjacent normal mucosae also showed a Gaussian distribution with an excess of bacteria-high cases. Based on this, we set a cut-off value of $3 \times 10^{-6}$ and 0.1 ($2^{\Delta Ct}$) for *F. nucleatum* and *pan-fusobacterium*, respectively. Among the 89 samples analyzed, 9 (10.1%) and 8 (9.9%) were classified as having a high amount of *F. nucleatum* and *pan-fusobacterium*, respectively (Supplementary Fig. 3). *F. nucleatum* and *pan-Fusobacterium* status was highly correlated in normal tissues ($p < 0.0001$, Supplementary Table 3). We then classified 13 out of 89 cases (14.6%) as FB-high, having either a high amount of *F. nucleatum* or *pan-fusobacterium* in the normal adjacent mucosa. FB-high status in normal
appearing mucosae was associated with a 15-fold increased likelihood of FB-high status in cancer tissues \( (p=0.0005, \text{Table 3}) \).

We next examined 72 non-neoplastic colonic biopsies from 65 cancer free subjects. 14 biopsies from 12 subjects (18.4\%) were classified as FB-high using the same cut off value used in cancer cases. The prevalence of FB-high was not significantly different between patients with CRCs and cancer free subjects (14.6\% vs. 18.4\%, \( p=0.66, \text{Table 4} \)). Patients with CIMP1 CRC were more likely to be FB-high in their adjacent tissues than patients with CIMP-negative CRC (29.2\% vs. 6.8\%, \( p=0.03, \text{Table 4} \)). FB-high state in the cancer free subjects was not associated with any clinical characteristics including gender, location and age (Supplementary Table 8). We also found no significant difference of FB-high state among samples from the United States (7/37, 18.9\%) or from Japan (7/35, 20\%) \( (p=0.92) \). Finally, we investigated the association between FB-high and DNA methylation status in non-neoplastic colonic mucosa using 7 different markers (\( ER, SFRP1, MYOD1, MGMT, SLC16A2, SPOCK2 \) and \( N33 \)). No significant association was found between FB-high and methylation status of any marker (Supplementary Fig.4).

Discussion

Our data show that CRC patients with a high level of \textit{Fusobacterium} in their cancer tissues have a molecularly distinct type of cancer, with a high degree of CpG island methylation and a high rate of mutations overall (though not of the \textit{TP53} gene). These data provide evidence for a pathogenic rather than passenger role for these bacteria. In favor of this argument are the facts that (1) a high level of bacteria can detected in both cancer, uninvolved adjacent mucosa and unaffected controls, (2) that the FB-high state in normal mucosa is strongly predictive of the specific molecular subtype of CRC patients and (3) that FB-high CRC have a distinct molecular profile; all these points suggest that bacteria were not simply an
epiphenomenon of the cancer state. Although the data imply a contributory role of *Fusobacterium*, they fall short of proving causation. Clearly, not all people with high levels of *Fusobacterium* have colon cancer. Thus, the interaction of this normal flora bacterium with cancer is best viewed in the light of emerging data on a pathogenic link between neoplastic cells and a permissive microenvironment. Our data are consistent with previous studies linking high-relative-abundance of *Fusobacterium* in tumor with regional lymph node metastases, which are also more likely to be CIMP positive cancers. *Fusobacterium* was also detected in a subset of resected CRC metastases, suggesting that *Fusobacterium* may be also required for the survival or maintenance of colorectal cancer cells. In fact, all FB-high CRCs were CIMP-1 or CIMP2 and none were CIMP-negative; however, only a small fraction of the total CIMP tumors are in this high FB-group.

Prevalence of *fusobacterium* measurable cases did not significantly differ across the different molecular subtypes of CRCs (data not shown). This suggests that bacteria high cases rather than simply detectable cases are important for the development of CIMP-positive CRCs. FB-high status may contribute to the development of a subset of CIMP-positive CRCs, affecting different molecular pathways. For example, we found that somatic mutations in the FB-high cases were significantly more frequent compared to CIMP1 and CIMP2/CIMP-negative with low/undetectable FB, and affected pathways seemed to be different though the small number of cases analyzed makes this conclusion tentative. Whether the different molecular pathways targeted affect patient prognosis should also be evaluated.

Although *F. nucleatum* and other *Fusobacterium* species are part of the gut microbiome in human, their invasive, adherent, and pro-inflammatory features have been noted. *Fusobacterium* have been associated with inflammatory disorders such as periodontitis, cerebral abscesses, acute appendicitis and inflammatory bowel diseases. It is interesting to note that the tumor subtype most associated with *Fusobacterium* (CIMP1 cases) have a distinct
immune response with abundant tumor infiltrating lymphocytes\textsuperscript{26, 34}. This inflammatory reaction has been thought to be a host immune response to the tumor cells and is associated with a better prognosis and longer survival\textsuperscript{26, 34}. Our data suggest that it could also be linked to an immune response to the high levels of bacteria in the peritumoral tissues. More broadly, inflammation may provide the pathogenic link between infections and cancer. Increased CpG island methylation is a noted feature of chronic inflammation, whether in the context of normal tissues (e.g. Ulcerative Colitis\textsuperscript{35, 36}) or cancer (e.g. EBV positive gastric cancer\textsuperscript{37}).

\textit{Fusobacterium} has a reported association with inflammatory bowel diseases (IBD), including both ulcerative colitis (UC) and Crohn’s diseases\textsuperscript{4, 5}, and IBD is one of the highest risk factors for CRC. Thus, the high rate of aberrant DNA methylation and somatic mutations in FB-high CRCs may reflect the fact that these cancers arise on a background of immune response triggered (or contributed to) by high levels of \textit{Fusobacterium}.

One of the interesting implications of this work is the potential of \textit{Fusobacterium} as a biomarker of cancer risk. In our studies, \textit{Fusobacterium} levels in normal colonic mucosa were higher in CIMP1 compared to CIMP-negative cases, but were also prevalent in cancer free subjects (and not associated with DNA methylation there). Thus, \textit{Fusobacterium} levels alone would not be useful as a biomarker of risk. Still, the hypothesis that \textit{Fusobacterium} contributes to neoplasia as a co-factor through tumor-microenvironment interactions suggest that it should be tested as a risk modifier, e.g. in patients with genetic and/or environmental predisposition to cancer. Also, the mean age of cancer free subjects analyzed in this study was younger than that in CRC cases, and we could not exclude the possibility that a considerable percentage of the FB-High cancer free subjects may be at increased risk of developing CRC in the future. Whether the \textit{Fusobacterium} levels in normal colonic mucosa would increase the risk of specific subtypes of CRC needs to be confirmed by prospective clinical studies. The hypothesis also deserves to be tested in animal models, where one could specifically explore the possibility of
therapeutic intervention targeting *Fusobacterium* in the prevention or treatment of colorectal cancer.
References


Figure Legends

**Fig. 1** Over-representation of *F. nucleatum* (left) and *pan-fusobacterium* (right) in CRC tissues relative to adjacent normal colonic mucosa in 89 paired cases. Statistical analysis was performed using the Wilcoxon signed-rank test.

**Fig. 2** Distribution of *Fusobacterium* in CRC patients (n=149). The cases were ranked according to the amount of *pan-fusobacterium* (right=high amount, left=low amount). Note that all *F. nucleatum* high cases (n=8) were included in *pan-fusobacterium* high cases (n=14) and there is clear separation of FB high group (n=14, 9.4%) and FB low/negative group (n=135, 90.6%). Red, CIMP1, *MLH1* methylated, *BRAF*, *KRAS* and *TP53* mutated; blue, CIMP2; grey CIMP-negative, *MLH1* unmethylated, *BRAF*, *KRAS* and *TP53* wild type; white, not determined;

**Fig. 3** Number of mutated genes determined by whole exome sequencing analysis in 11 CRCs. (2 FB-high, 6 FB-low/negative CIMP1 and 3 FB-low/negative CIMP2/CIMP-negative). Statistical analysis was performed using One-way ANOVA.
### Table 1: Clinicopathological characteristics of 149 CRCs studied

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>CIMP-negative</th>
<th>CIMP1</th>
<th>CIMP2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number</td>
<td>60</td>
<td>42</td>
<td>47</td>
</tr>
<tr>
<td>Age: mean +/- SEM$</td>
<td>64.0 +/- 1.9</td>
<td>71.8 +/- 1.3</td>
<td>66.7 +/- 1.6</td>
</tr>
<tr>
<td>Female</td>
<td>21 (35.0%)</td>
<td>21 (50.0%)</td>
<td>18 (38.3%)</td>
</tr>
<tr>
<td>Proximal location#$</td>
<td>26 (52.0%)</td>
<td>26 (86.7%)</td>
<td>22 (75.9%)</td>
</tr>
<tr>
<td>BRAF mutant§</td>
<td>2 (3.6%)</td>
<td>31 (73.8%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>KRAS mutant*</td>
<td>23 (40.0%)</td>
<td>0 (0%)</td>
<td>38 (80.9%)</td>
</tr>
<tr>
<td>TP53 mutant$$</td>
<td>37 (66.1%)</td>
<td>3 (7.1%)</td>
<td>18 (40.0%)</td>
</tr>
<tr>
<td>MSI##</td>
<td>6 (13.0%)</td>
<td>36 (97.3%)</td>
<td>0 (0%)</td>
</tr>
</tbody>
</table>

$^a$: CIMP1 vs. CIMP-negative, p=0.002, CIMP1 vs. CIMP2, p=0.01.
$^b$: CIMP1 vs. CIMP-negative, p=0.002. Data were missing in 28 cases.
$^c$: CIMP1 vs. CIMP-negative, p<0.0001, CIMP1 vs. CIMP2, p<0.0001. Data were missing in 5 cases.
$^d$: CIMP2 vs. CIMP-negative, p=0.0001, CIMP2 vs. CIMP1, p<0.0001, CIMP-negative vs. CIMP2, p<0.0001. Data was missing in one case.
$^e$: CIMP-negative vs. CIMP1, p<0.0001, CIMP-negative vs. CIMP2, p=0.02, CIMP2 vs. CIMP1, p=0.0004. Data were missing in 6 cases.
$^f$: CIMP1 vs. CIMP-negative, p<0.0001, CIMP1 vs. CIMP2, p<0.000. Data were missing in 36 cases.

Note: Proximal, cecum, and ascending and transverse colon; distal, descending and sigmoid colon, and rectum
ND, not determined.
Table 2 Association between high amount of Fusobacterium and clinical and molecular subtypes of CRCs

<table>
<thead>
<tr>
<th>Variables: n (%)</th>
<th>FB-high (%)</th>
<th>FB-low/negative (%)</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CIMP status</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CIMP-negative</td>
<td>0</td>
<td>60</td>
<td>100.0</td>
</tr>
<tr>
<td>CIMP-1</td>
<td>9</td>
<td>33</td>
<td>78.6</td>
</tr>
<tr>
<td>CIMP-2</td>
<td>5</td>
<td>42</td>
<td>89.4</td>
</tr>
<tr>
<td><strong>BRAF</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wild type</td>
<td>8</td>
<td>103</td>
<td>92.8</td>
</tr>
<tr>
<td>Mutated</td>
<td>6</td>
<td>27</td>
<td>81.8</td>
</tr>
<tr>
<td><strong>KRAS</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wild type</td>
<td>10</td>
<td>77</td>
<td>88.5</td>
</tr>
<tr>
<td>Mutated</td>
<td>4</td>
<td>57</td>
<td>93.4</td>
</tr>
<tr>
<td><strong>P53</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wild type</td>
<td>12</td>
<td>73</td>
<td>85.9</td>
</tr>
<tr>
<td>Mutated</td>
<td>1</td>
<td>57</td>
<td>98.3</td>
</tr>
<tr>
<td><strong>hMLH1</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unmethylated</td>
<td>5</td>
<td>103</td>
<td>95.4</td>
</tr>
<tr>
<td>Methylated</td>
<td>9</td>
<td>32</td>
<td>78.0</td>
</tr>
<tr>
<td><strong>MSI</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MSS</td>
<td>3</td>
<td>68</td>
<td>95.8</td>
</tr>
<tr>
<td>MSI</td>
<td>8</td>
<td>34</td>
<td>81.0</td>
</tr>
<tr>
<td><strong>CHD7</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wild type</td>
<td>7</td>
<td>81</td>
<td>92.0</td>
</tr>
<tr>
<td>Mutated</td>
<td>4</td>
<td>8</td>
<td>66.7</td>
</tr>
<tr>
<td><strong>CHD8</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wild type</td>
<td>7</td>
<td>80</td>
<td>92.0</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>mean</td>
<td>stdv</td>
</tr>
<tr>
<td>----------</td>
<td>---</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td>Mutated</td>
<td>4</td>
<td>30.8</td>
<td>9</td>
</tr>
<tr>
<td>CHD7 or 8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wild type</td>
<td>4</td>
<td>5.1</td>
<td>74</td>
</tr>
<tr>
<td>Mutated</td>
<td>7</td>
<td>31.8</td>
<td>15</td>
</tr>
<tr>
<td>Location</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distal colon</td>
<td>2</td>
<td>4.1</td>
<td>47</td>
</tr>
<tr>
<td>Proximal colon</td>
<td>9</td>
<td>12.5</td>
<td>63</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>7</td>
<td>7.9</td>
<td>82</td>
</tr>
<tr>
<td>Female</td>
<td>7</td>
<td>11.7</td>
<td>53</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>~70y</td>
<td>5</td>
<td>5.8</td>
<td>81</td>
</tr>
<tr>
<td>70y&lt;</td>
<td>9</td>
<td>14.5</td>
<td>53</td>
</tr>
<tr>
<td>Adjacent tissues</td>
<td>FB-high</td>
<td>FB low/negative</td>
<td></td>
</tr>
<tr>
<td>--------------------------</td>
<td>---------</td>
<td>-----------------</td>
<td></td>
</tr>
<tr>
<td>FB-low/negative (n=76)</td>
<td>4 (5.3%)</td>
<td>72 (94.7%)</td>
<td></td>
</tr>
<tr>
<td>FB-high (n=13)</td>
<td>6 (46.2%)</td>
<td>7 (53.8%)</td>
<td></td>
</tr>
</tbody>
</table>

Odds ratio=15.4, 95% confidence intervals=3.5-68.1, $p=0.0005$. 
Table 4 *Fusobacterium* status in non-neoplastic colonic mucosa in cancer free and CRC patients

<table>
<thead>
<tr>
<th></th>
<th>FB-high</th>
<th>FB low/negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cancer free (n=65)</td>
<td>12 (18.4%)</td>
<td>53 (80.6%)</td>
</tr>
<tr>
<td>CRC cases (n=89)</td>
<td>13 (14.6%)</td>
<td>76 (85.4%)</td>
</tr>
<tr>
<td>CIMP-negative (n=44)</td>
<td>3 (6.8%)</td>
<td>41 (93.2%)</td>
</tr>
<tr>
<td>CIMP1&lt;sup&gt;$&lt;/sup&gt; (n=24)</td>
<td>7 (29.2%)</td>
<td>17 (70.8%)</td>
</tr>
<tr>
<td>CIMP2 (n=21)</td>
<td>3 (14.3%)</td>
<td>18 (85.7%)</td>
</tr>
<tr>
<td>All CIMP (n=45)</td>
<td>10 (22.2%)</td>
<td>35 (77.8%)</td>
</tr>
</tbody>
</table>

<sup>$</sup>CIMP1 vs. CIMP-negative,  *p*=0.03
Figure 1

**F. nucleatum**

\[ p < 0.0001 \]

**Pan-fusobacterium**

\[ p < 0.0001 \]
Figure 2

![Graph showing the expression levels of different bacterial species and their association with various cancer markers.](image-url)
Figure 3

Mean mutation (n)

CIMP1 FB-high
CIMP1 FB-low/
negative
CIMP2/CIMP-negative FB low/
negative

p = 0.003
Fusobacterium in colonic flora and molecular features of colorectal carcinoma


Cancer Res  Published OnlineFirst January 2, 2014.

Updated version

Access the most recent version of this article at: doi:10.1158/0008-5472.CAN-13-1865

Supplementary Material

Access the most recent supplemental material at: http://cancerres.aacrjournals.org/content/suppl/2014/01/06/0008-5472.CAN-13-1865.DC1

Author Manuscript

Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

E-mail alerts

Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions

To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions

To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.