Shifts in the Fecal Microbiota Associated with Adenomatous Polyps

Vanessa L. Hale1,2, Jun Chen3, Stephen Johnson3, Sean C. Harrington2, Tracy C. Yab4, Thomas C. Smyrk5, Heidi Nelson1, Lisa A. Boardman4, Brooke R. Druliner4, Theodore R. Levin5, Douglas K. Rex7, Dennis J. Ahnen8, Peter Lance9, David A. Ahlquist4, and Nicholas Chia1,2,10,11

Abstract

Background: Adenomatous polyps are the most common precursor to colorectal cancer, the second leading cause of cancer-related death in the United States. We sought to learn more about early events of carcinogenesis by investigating shifts in the gut microbiota of patients with adenomas.

Methods: We analyzed 16S rRNA gene sequences from the fecal microbiota of patients with adenomas (n = 233) and without (n = 547).

Results: Multiple taxa were significantly more abundant in patients with adenomas, including Bilophila, Desulfovibrio, proinflammatory bacteria in the genus Megabacterium, and multiple Bacteroidetes species. Patients without adenomas had greater abundances of Veillonella, Firmicutes (Order Clostridia), and Actinobacteria (family Bifidobacteriaceae). Our findings were consistent with previously reported shifts in the gut microbiota of colorectal cancer patients. Importantly, the altered adenoma profile is predicted to increase primary and secondary bile acid production, as well as starch, sucrose, lipid, and phenylpropyroid metabolism.

Conclusions: These data hint that increased sugar, protein, and lipid metabolism along with increased bile acid production could promote a colonic environment that supports the growth of bile-tolerant microbes such as Bilophila and Desulfovibrio. In turn, these microbes may produce genotoxic or inflammatory metabolites such as H2S and secondary bile acids, which could play a role in catalyzing adenoma development and eventually colorectal cancer.

Impact: This study suggests a plausible biological mechanism to explain the links between shifts in the microbiota and colorectal cancer. This represents a first step toward resolving the complex interactions that shape the adenoma–carcinoma sequence of colorectal cancer and may facilitate personalized therapeutics focused on the microbiota. Cancer Epidemiology, Biomarkers & Prevention 26(1); 85–94. ©2016 AACR.

Introduction

Adenomatous polyps, or adenomas, have long been recognized as a critical precursor to colorectal cancer (1, 2), the second leading cause of cancer-related deaths in the United States (3). Although screening (4–6) and lifestyle (7–10) play important roles in colorectal cancer prevention, identifying a causal mechanism of mutagenesis is essential to understand the adenoma–carcinoma sequence and to develop new and personalized prevention strategies. The gut microbiota has recently been implicated in adenoma and colorectal cancer pathogenesis (11, 12) and offers a promising avenue for personalized prevention (13). Importantly, many of the risk factors for colorectal cancer, including diet (high red meat/high fat/low fiber; refs. 8, 14), obesity (15), physical activity (10), smoking (7), and alcohol use (9), also have significant effects on the gut microbial community (16). Because the gut microbiota alters the metabolic environment of the host, it may directly or indirectly influence mutagenesis rates (11, 17), and thus carcinogenesis.

Previous studies on the microbiome of individuals with adenomas have identified many microbes associated with these particular polyps (Table 1). However, most of these studies lack functional analyses necessary to suggest a mechanistic link between microbiota, adenoma development, and carcinogenesis. Microbial functionality, which can be predicted on the basis of microbial genomes, provides greater insight into the microbial ecology of the colon by not only indicating what taxa are differentially abundant, but also the putative function of these taxa (18). Without functional analyses, it is difficult to elucidate the role of microbes in the adenoma–carcinoma...
sequence because microbial taxa associated with adenomas and colorectal cancer vary widely by study (11, 12). In addition, many subject cohorts are relatively underpowered, ranging in size from 6 to 67 individuals with adenomas (see Table 1), making it even more difficult to identify subtle microbial or functional changes that may be underlying adenoma/colorectal cancer pathogenesis. Moreover, meta-analysis on these data is particularly challenging due to multiple biases attributed to extraction methods (19), PCR regions (20), and collection protocols (21). As such, a well-powered study with a uniform collection/ extraction protocols and functional analyses is needed to more definitively probe the link between the microbial community and adenoma development.

In this study, we compared the fecal microbiota of patients with adenomas and without adenomas. Our aim was 2-fold: To determine whether gut microbial communities can be used to predict the presence of adenomas and to elucidate the microbial ecology underlying the adenoma–carcinoma sequence. Here, we report significant shifts in the gut microbiota composition of patients with adenomas and use these changes and their predicted functional consequences to propose a model linking diet, gut microbes, and the development of adenomas, the precursors to colorectal cancer.

### Materials and Methods

#### Subject enrollment

Fecal samples \((n = 780)\) were selected from a freezer archive of stools collected without preservative buffer. All stool samples came from patients presenting for standard screening colonoscopy between 2001 and 2005 at multiple medical centers, including the Mayo Clinic, Rochester, MN; Kaiser Permanente in Sacramento and Oakland, CA; Oregon Health & Science University, Portland, OR; University of Colorado Health Sciences Center, Denver, CO; Roswell Park Cancer Institute, Buffalo, NY; Indiana University Medical Center, Indianapolis, IN; and other North Central Cancer Treatment Group institutions (22). All patients were 50 to 80 years old and were voluntarily enrolled before presenting for colonoscopy (Fig. 1). Exclusion criteria for the original study comprised premenopausal women, hematochezia or melena within the month before enrollment, prior colorectal resection, coagulopathy or anticoagulant use, chemotherapy

#### Table 1. Microbial taxa associated with adenomas in previous studies

<table>
<thead>
<tr>
<th>Study</th>
<th>Number of individuals with adenomas</th>
<th>Number of individuals without adenomas</th>
<th>Microbial taxa enriched in individuals with adenomas</th>
<th>Microbial taxa enriched in individuals without adenomas</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>20</td>
<td>20</td>
<td>Increased microbial diversity within the Clostridium leptum and C. cocoides subgroups</td>
<td>Bacteroidetes Bacteroides spp. Coprococcus spp</td>
<td>Scanlan et al. 2008 (76)</td>
</tr>
<tr>
<td>2</td>
<td>21</td>
<td>23</td>
<td>Proteobacteria Dorea spp. Faecalibacterium spp.</td>
<td>Bacteroidetes Streptococcus</td>
<td>Shen et al. 2010 (40)</td>
</tr>
<tr>
<td>3</td>
<td>33</td>
<td>38</td>
<td>TM7 Cyanobacteria Verrucomicrobia Acidovorax Aquabacterium Cloacibacterium Helicobacter Lactococcus Lactobacillus Pseudomonas</td>
<td>Streptococcus</td>
<td>Sanapareddy et al. 2012 (77)</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>6</td>
<td>Firmicutes Proteobacteria</td>
<td>Bacteroides</td>
<td>Brim et al. 2013 (36)</td>
</tr>
<tr>
<td>5</td>
<td>47</td>
<td>47</td>
<td>Enterococcus Streptococcus Bacteroidetes</td>
<td>Clostridium Roseburia Eubacterium</td>
<td>Chen et al. 2013 (37)</td>
</tr>
<tr>
<td>6</td>
<td>67</td>
<td>48</td>
<td>Fusobacterium</td>
<td>Methanobacteriales Methanobrevibacterium Faecalibacterium</td>
<td>McCoy et al. 2013 (48)</td>
</tr>
<tr>
<td>7</td>
<td>11</td>
<td>10</td>
<td>Bifidobacterium Fusobacterium Enterobacteriaceae Akkermansia Blautia</td>
<td>Methanobacteriales Methanobrevibacterium Faecalibacterium</td>
<td>Mira-Pascual et al. 2014 (49)</td>
</tr>
<tr>
<td>8</td>
<td>15</td>
<td>15</td>
<td>Bifidobacterium Eubacteria</td>
<td>Bacteroides</td>
<td>Nugent et al. 2014 (78)</td>
</tr>
<tr>
<td>9</td>
<td>30</td>
<td>30</td>
<td>Ruminococcaceae Clostridium Pseudomonas Porphyromonadaceae</td>
<td>Bacteroides Lachnospiraceae Clostridiales Clostridium</td>
<td>Zackular et al. 2014 (71)</td>
</tr>
<tr>
<td>10</td>
<td>20</td>
<td>15</td>
<td>Proteobacteria Gammaproteobacteria</td>
<td></td>
<td>Goedert 2015 (72)</td>
</tr>
</tbody>
</table>
within 3 months of enrollment, contraindications to colonoscopy, inability to desist from therapeutic doses of nonsteroidal anti-inflammatory drugs (NSAID), aerodigestive cancer within 5 years of enrollment, a fecal occult blood test within the year before enrollment, and colorectal evaluation (e.g., sigmoidoscopy or colonoscopy) within 10 years of enrollment. Patients at high risk for colorectal cancer, including patients with familial adenomatous polyposis, cancer syndromes, inflammatory bowel disease, prior colorectal cancer or adenomas, or ≥2 first-degree relatives with colorectal cancer, were also excluded.

Standard diagnostic colonoscopies were performed on all patients and included intravenous sedation (unless otherwise requested); inspection of the colonic mucosal surfaces up to the point of the cecum; and lesion assessment, including recording the location, size, number, and architecture all polypoid lesions. All polyps/lesions removed from the colon were submitted for histologic classification and reviewed by the same pathologist.

Fecal samples from patients in which at least one adenoma >1 cm was identified were included in the “adenoma” group. Fecal samples from patients with no polyps were included in the “non-adenoma” group. Fecal samples from patients who were diagnosed with colorectal cancer were excluded from analysis.

Approval for this study was granted by the Mayo Clinic’s Institutional Review Board. Fecal samples were collected under protocol #15-004021, from patients who had previously enrolled under protocol #532-00, undergone standard screening colonoscopies, and given consent for the use of their samples in future research studies.

Sample collection and processing
Fecal samples were self-collected by patients after enrollment and up to 3 months before bowel preparation and colonoscopy. Samples were collected in a bucket container mounted to a toilet seat. Promptly after defecation, whole stools were express shipped on ice in insulated containers to a central laboratory where they were immediately archived at -80°C until further processing. Samples received >48 hours after defecation were disqualified. In preparation for DNA extraction, a 4-mm biopsy punch (Miltex) was used to collect a core sample from the still-frozen feces. The frozen fecal core was immediately transferred into Chemagic lysis buffer (PerkinElmer). DNA extraction was performed on a Che-magic MSM I (PerkinElmer), using the Chemagic DNA Blood Special Kit (PerkinElmer). DNA quantification and amplification was performed as described previously (23). The 16S rRNA sequencing library was constructed at the University of Minnesota Genomics Center, and sequencing was performed at the Mayo Clinic Medical Genomics Facility, on a MiSeq using a MiSeq Reagent Kit v3 (2 × 300, 600 cycles; Illumina Inc.).

Sequence processing
After sequencing, adapter–primer sequences were removed from reads as described previously (23). Sequences were then processed via the IM-TORNADO bioinformatics pipeline, using a 97% identity threshold to assign operational taxonomic units (OTU; ref. 24). Paired R1 and R2 reads were analyzed. In total, 17,579,026 reads passed quality control. Singleton OTUs as well as samples with less than 2,000 reads were removed. Sequencing data are available at dbGaP Study Accession: phs001204.v1.p1.

Statistical analyses
\( \alpha \)-diversity and \( \beta \)-diversity. To compare the microbial communities of the adenoma and non-adenoma groups, we summarized microbiota data using both \( \alpha \)-diversity and \( \beta \)-diversity measures. Two \( \alpha \)-diversity metrics were used, the observed OTU number and the Shannon index. The observed OTU number reflects species richness, whereas the Shannon index places more weight on species evenness. \( \beta \)-diversity, in contrast, indicates the shared diversity between bacterial populations in terms of ecological distance; different distance metrics provide distinctive views of community structure. Two \( \beta \)-diversity measures, unweighted and weighted UniFrac distances, were calculated using the OTU table and a phylogenetic tree (with the “GUniFrac” function in the R package GUniFrac; ref. 16). The unweighted UniFrac reflects differences in community membership (i.e., the presence or absence of an OTU), whereas the weighted UniFrac mainly captures differences in abundance. Rarefaction was performed on the OTU table before calculating the distances.

To assess the association between adenoma status and \( \alpha \)-diversity, we fitted a linear regression model to the \( \alpha \)-diversity metrics after rarefaction, adjusting for technical covariates such as sequencing batch. A Wald test was used to determine significance.
To assess the association between adenoma status and β-diversity measures, we used the recently proposed MiRKAT, which is a kernel-based association test based on ecological distance matrices (25). We also used MiRKAT to assess the relationship between polyp characteristics (size, number, location, architecture, and histology) and β-diversity measures. In individuals with multiple polyps, a single polyp location was chosen at random, and the most severe architecture and histology per patient were selected for analysis. MiRKAT produces analytic P values for individual distance metrics, as well as a permutation-based omnibus P value that combines multiple distance metrics, for a more robust and powerful assessment of significance. For the omnibus test, significance was assessed using 1,000 permutations, and the covariate, sequencing batch, was adjusted if necessary. Ordination plots were generated using principal coordinate analysis as implemented in R (“cmdscale” function in the R “vegan” package).

Differential abundance analysis. We conducted differential abundance analysis at the phylum, class, order, family, and genus levels, and we filtered out taxa with prevalence less than 10%. We normalized the count data into relative abundances (proportions) by dividing by the total read count; taxa with a maximum proportion less than 0.2% were excluded from testing to reduce the number of the tests. To identify differentially abundant taxa while accommodating covariates (e.g., sequencing batch) and the non-normality of the count data, we used a permutation test in which a regular linear model was fitted, with taxa proportion data as the outcome variable. To reduce the effects of outliers, taxa proportion data were square-root transformed. Statistical significance was assessed using 1,000 permutations with the F-stat as the test statistic. False discovery rate (FDR) control (B-H procedure, “P<quantile”) in standard R packages) was used to correct for multiple testing, and FDR-adjusted P values or q values less than 0.2 were considered significant. This q value cutoff was chosen to avoid missing important taxa with small effect sizes and is a significance threshold frequently used in human microbiome studies (26, 27). To quantify the effect size of the differential taxa, we used the fold change of the mean relative abundance between the normal and adenoma groups.

Predictive modeling based on random forests. The machine learning algorithm random forests (RF) was used to predict adenoma status based on the microbiota profile (genus-level proportion data) using default parameters of the R implementation of the algorithm (28). The RF algorithm, due to its nonparametric assumptions, can detect both linear and nonlinear effects and potential taxon–taxon interactions, thereby identifying the taxa that best discriminate between groups. Boruta variable selection was applied to select the most discriminatory taxa based on importance values produced by RF (29). The Boruta method spikes abundance data with “shadow” taxa, which are shuffled versions of real taxa. This enables us to assess whether the importance of a given taxon is significant, that is, whether the importance is discernible from the effects that arise from random fluctuations (shadow taxa). We then assessed the ability of the Boruta-selected taxa to predict adenoma status using the receiver operating characteristic (ROC) curve, which was estimated using the 0.632+ bootstrap method to more accurately assess error rates (30).

Functional data analysis. PICRUSt was used to infer the abundance of functional categories (KEGG metabolic pathways and COG functional groups) based on the 16S rRNA data, and differential abundance analysis was performed using the same permutation test that was used for the taxon analysis (18). No prevalence-based filtering was applied before differential abundance testing, because most of the functional categories are shared across subjects. All statistical analyses were performed in R 3.0.2 (R Development Core Team).

Results

Cases (“adenoma” group) comprised 233 patients with at least one large adenoma (≥1 cm); controls included 547 patients with no polyps on colonoscopy (“non-adenoma” group). The groups did not differ with regard to the potential confounders of age, sex, race, history of smoking, history of cancer, or diagnosis of colorectal cancer or polyps in first-degree relatives (Table 2).

The overall composition of the groups’ gut microbial communities appeared similar at the levels of phylum, family, and genus (Supplementary Fig. S1A). The groups did not differ significantly in terms of microbial species richness (P = 0.21) or diversity (Shannon Index: P = 0.23; Supplementary Fig. S1B and S1C). Neither did they cluster in PCoA plots using unweighted or weighted UniFrac distance metrics (Supplementary Fig. S1D and S1E). However, our large sample size allowed us to detect small yet statistically significant differences in microbial composition between the adenoma and non-adenoma groups (MiRKAT omnibus P = 0.032). No differences in microbial composition were detected on the basis of polyp size, architecture, or location, but polyp number was significant (MiRKAT omnibus P = 0.035) and histology (hyperplastic, low-grade dysplasia, or high-grade dysplasia) was marginally significant (MiRKAT omnibus P = 0.091; Supplementary Table S1).

Next, we identified 31 specific taxa that differed in abundance between patients with and without adenomas (Fig. 2, q < 0.2). Taxa that were more abundant in the adenoma group included multiple OTUs in the Bacteroidetes phyla and Deltaaproteobacteria class, including OTUs in the Bilophila, Desulfovibrio, Sutterella, and Mogibacterium genera. Taxa more common in the non-adenoma group included Firmicutes,

Table 2. Demographics of the adenoma and non-adenoma groups

<table>
<thead>
<tr>
<th>Variable</th>
<th>Adenoma (n = 233)</th>
<th>Non-adenoma (n = 547)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (mean, SD)</td>
<td>66.5±6.9</td>
<td>66.5±6.9</td>
<td>0.60</td>
</tr>
<tr>
<td>Sex (n, %)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>100 (42.9)</td>
<td>237 (43.3)</td>
<td>0.98</td>
</tr>
<tr>
<td>Male</td>
<td>133 (57.1)</td>
<td>310 (56.7)</td>
<td></td>
</tr>
<tr>
<td>Race (n, %)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>223 (95.7)</td>
<td>503 (92)</td>
<td>0.30</td>
</tr>
<tr>
<td>Black</td>
<td>1 (0.4)</td>
<td>14 (2.6)</td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td>4 (1.7)</td>
<td>14 (2.6)</td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td>0 (0)</td>
<td>8 (1.5)</td>
<td></td>
</tr>
<tr>
<td>Native American</td>
<td>0 (0)</td>
<td>2 (0.4)</td>
<td></td>
</tr>
<tr>
<td>Other/Unknown</td>
<td>3 (1.3)</td>
<td>5 (0.9)</td>
<td></td>
</tr>
<tr>
<td>Ever smoker (n, %)</td>
<td>138 (59.2)</td>
<td>310 (56.7)</td>
<td>0.56</td>
</tr>
<tr>
<td>History of cancer, any type (n, %)</td>
<td>44 (19.5)</td>
<td>116 (21.2)</td>
<td>0.52</td>
</tr>
<tr>
<td>First-degree relative with colorectal cancer (n, %)</td>
<td>35 (15)</td>
<td>89 (16.3)</td>
<td>0.74</td>
</tr>
<tr>
<td>First-degree relative with polyps (n, %)</td>
<td>25 (10.7)</td>
<td>47 (8.6)</td>
<td>0.82</td>
</tr>
</tbody>
</table>
such as OTUs in the Clostridia class and Veillonella genus, as well as OTUs in the Bifidobacteriales order and Haemophilus genus. Despite moderate effect sizes (fold change range, 1.06–2.77), these significant results indicate that the microbiota in the adenoma group systematically differs from the non-adenoma group.

Figure 2.
Thirty-one taxa differ in abundance between patients with and without adenomas. A, Relative abundance of OTUs in each group, across taxonomic levels. B, −\log(P value) of these taxa’s differential abundance. C, Cladogram of the taxa that differed between groups.
We next assessed the utility of the gut microbiota as a clinical biomarker for adenomas using RF-based prediction. Boruta feature selection was used to select the most predictive taxa to improve prediction. Of the 31 taxa identified by differential abundance testing, the Boruta algorithm identified four genera that significantly predicted adenoma status: Streptococcus and Veillonella, which were enriched in the non-adenoma group, and Mogibacterium and Sutterella, which were enriched in the adenoma group (Fig. 3; for heatmap see Supplementary Fig. S2). The Bilophila genus was also more predictive than most other genera included in this analysis; however, these genera did not exceed the threshold for significance. An ROC curve generated with the four significantly predictive taxa resulted in an AUC of 0.6599 (Supplementary Fig. S3; DeLong test, \( P = 0.001 \)). Although significant, this level of sensitivity/specifity is too low for consideration as a clinical biomarker for adenomas. Thus, this analysis indicates that although the abundance of Streptococcus, Veillonella, Mogibacterium, and Sutterella is not sufficient to reliably identify samples from patients with adenomas, the levels of these genera are consistently altered in their respective groups.

To determine whether the taxonomic differences between the groups’ microbiota corresponded to functional changes, we performed a predictive functional analysis of the 16S rRNA sequences present (Fig. 4, \( q < 0.2 \)). PICRUSt analyses predicted that the adenoma group’s microbiota exhibits increased primary and secondary bile acid synthesis; increased galactose, starch and sucrose, and sphingolipid metabolism; and increased phenylpropanoid biosynthesis. In contrast, the non-adenoma group’s microbiota is predicted to exhibit increased biosynthesis of unsaturated fatty acids and increased purine, pyrimidine, D-Alanine, nicotinate, and nicotinamide metabolism.

Discussion

In this study, we report significant differences in the microbial composition of individuals with adenomas. We also observe differences based on polyp number and histology but not size, architecture, or polyp location, suggesting that microbial communities associated with polyps change (or are detectable) with some but not all aspects of polyp severity. We identified 31 taxa that were differentially abundant among patients with and without adenomas, and four of these taxa were significantly predictive of adenoma status, although they could not be used to reliably classify samples. On the basis of the 16S sequences present in each group, we also identified putative metabolic shifts between the microbiota of the adenoma and non-adenoma groups.

Links with colorectal cancer have already been reported in many of the taxa we identified as differentially abundant in individuals with adenomas. This suggests that changes in the microbial community associated with adenomas may represent early events in the pathway leading to colorectal cancer. For example, we identified increased levels of Bilophila, Desulfovibrio, Bacteroidetes, and Mogibacterium in individuals with adenomas. Both Bilophila and Desulfovibrio produce genotoxic hydrogen sulfide (H₂S) as an endproduct of anaerobic respiration (31–33) and have been
associated with colorectal cancer in other studies (34, 35). In addition, multiple studies have reported elevated proportions of *Bacteroidetes* in patients with adenomas (36, 37) or colorectal cancer (refs. 38, 39; but not all; see ref. 40). *Bacteroides fragilis*, in particular, causes colitis-associated carcinogenesis (41). Finally, *Mogibacterium* is an oral bacterium associated with periodontal disease and root canal infections, and it, too, has been linked to colorectal cancer (42–44).

Other taxa differentially abundant in individuals with adenomas are also plausible contributors to carcinogenesis. For example, *Sutterella*, a genus highly predictive of adenoma status (Fig. 3), may play a role in inflammation, as it has been linked to active colitis in a mouse model of inflammatory bowel disease (45). Gastrointestinal inflammation has been strongly linked to colorectal cancer pathogenesis (11). In contrast, *Veillonella*, also highly predictive of adenoma status but enriched in patients without adenomas, may exert a protective role in the colon (46) along with other taxa enriched in this group, including *Firmicutes* and *Actinobacteria* (family *Bifidobacteriales*; ref. 47). Notably, we did not identify an enrichment of *Fusobacterium* or *Porphyromonadaceae* in individuals with adenomas, as reported in other studies (39, 48, 49). This may have been due to differences in study populations, fecal collection, and preservation techniques (21, 50), library preparation (51), or primers and sequencing platforms (52, 53).

We evaluated microbial alterations in tandem with predictive functional differences identified by PICRUST. In an analysis of Human Microbiome Project data, PICRUST produced an average correlation of 0.8 between predicted functions and actual functions identified through deep metagenomic sequencing (18). In addition, PICRUST produced more accurate and reliable functional predictions than shallow metagenomic sequencing (18). Despite these strengths of PICRUST, predictive functions should be examined with care, as genomes and functions of the microbes present in a given sample may differ from the genomes and functions upon which PICRUST builds its predictions. The results from our predictive functional analysis suggest a link between microbial shifts observed in individuals with adenomas to metabolic pathways that have previously been associated with dietary risk factors common in a Western diet. The adenoma microbiota was characterized by putative functional groups associated with galactose, sphingolipid, and starch/sucrose metabolism, as well as phenylpropanoid biosynthesis. Importantly, diets high in dairy result in increased galactose metabolism; diets high in fat result in increased lipid/sphingolipid metabolism (34); diets high in refined starches and sugars lead to increased starch/sucrose metabolism (55); and diets high in protein result in increased phenylpropanoid biosynthesis (56). Diets high in animal fat and protein also lead to increased bile acid (BA) production (57). Interestingly, the adenoma microbiota is predicted to display increased levels of primary and secondary BA synthesis. These functional predictions suggest that individuals with adenomas are consuming diets higher in fat, sugar, starch, protein, and dairy than non-adenoma individuals. These findings are consistent with multiple epidemiologic studies, which have drawn links between a Western diet (high in fat, dairy, meat, and sugars) and the incidence of adenomas (58–60). This suggests a potential link between diet and the molecular mechanisms involved in adenoma pathogenesis.

We propose the following mechanism linking diet, the microbiota, and the adenoma–cancer sequence: diets high in fat and protein increase production of primary BAs, which help digest and absorb lipids in the small intestine (61, 62). This promotes the growth of bile-tolerant bacteria such as *Bifidobacteria* and some species of *Desulfovibrio*. Blooms of these species may increase the production of genotoxic metabolites such as H$_2$S (61, 62). In addition, the colon microbiota can deconjugate primary bile acids to form secondary BAs (62, 63), and some of these secondary BAs, such as lithocholic and deoxycholic acid, have cytotoxic and genotoxic effects (62–66). Elevated levels of secondary BAs and proinflammatory bacteria such as *Mogibacterium* and *Sutterella* may result in the perfect storm of DNA damage and inflammation, leading to adenoma development and eventually malignant transformation.

Several limitations of our study warrant mention. Three include a lack of information on participants’ diet, body mass index (BMI), and recent antibiotic use. Without dietary information, we cannot confirm that the adenoma group consumed a diet higher in sugar, animal fat, and protein; although, previous studies have indicated a link between Western diet and adenomas.
In addition, we are also unable to determine whether BMI acts as a confounder; it is certainly possible, as the gut microbiome of obese individuals differs significantly from the microbiome of lean individuals (67), and higher BMI has been associated with adenoma development (68). However, obese/ high BMI phenotypes are commonly associated with increased relative abundances of microbes in the phylum Firmicutes whereas lean phenotypes are associated with increased abundances of Bacteroidetes phylum microbes (67, 69). In our study, individuals with adenomas had increased abundances of Bacteroidetes microbes whereas individuals without adenomas had increased abundances of Firmicutes phylum microbes. This is opposite to what we would have expected if BMI was the main driver of adenoma development; thus, we suggest BMI was not a strong confounder in our dataset. Lack of antibiotic data prevents us from excluding or analyzing data based on antibiotic use, which can dramatically alter the gut microbiota (70); although, we have no a priori reason to believe that either group would exhibit increased antibiotic use in relation to the other. Finally, the cross-sectional nature of our data does not allow us to parse correlation versus causation between microbial alterations and adenoma status. Although our results show that observed microbial changes lack the specificity and sensitivity to serve as a clinical biomarker for adenomas, these findings provide important insights into mechanisms that may be driving adenoma development.

This study represents the largest study on microbial communities associated with adenomas to date. This robust dataset allowed us to detect subtle microbial changes that may be key to understanding how a healthy colon develops adenomas, which can then transform into carcinomas. We also adjusted our analyses for multiple comparisons, which not all studies on adenoma microbiota opted to do (37, 71, 72). Sample collection is another strength of this study. All fecal samples from individuals in the adenoma and non-adenoma groups were shipped on ice and received and frozen at −80°C within 48 hours of defecation. Previous studies have demonstrated that fecal microbial communities stored at ambient temperatures for up to 24 hours, are relatively unaffected (21), and no significant changes in microbial diversity or composition are detected in fecal samples stored at 4°C for up to 72 hours (73). In addition, long-term storage of fecal samples at −80°C seems to have little effect on overall microbial composition (50, 74); although, no study, to our knowledge, has examined fecal preservation in samples over 10 years old, as is the case with samples in this study. Notably, we only examined fecal microbiota and not the mucosal-associated microbiota, which has been reported to differ in composition and diversity (75). Every individual sampled in this study underwent a complete colonoscopy with full visualization of the colon from rectum to cecum, and colonoscopy is regarded as the most robust reference standard for presence or absence of polyps. Polyps removed during colonoscopies were all reviewed and classified by the same pathologist. Finally, our study included predictive functional analyses based on the microbial communities of the adenoma and non-adenoma groups. Functional analyses have not been performed on previous adenoma datasets, and this effort suggested key insights as to how the host and microbial community may be interacting within the context of adenoma development.

In conclusion, we have shown that the composition of the gut microbiota in individuals with adenomas differs significantly from that of healthy individuals and resembles the microbiota of individuals with colorectal cancer. Moreover, we suggest that these shifts may be consistent with the effects of the Western diet and are predicted to result in metabolic changes that could increase risk of cellular damage and mutagenesis in the gut. Collectively, our findings support a proposed model in which diet alters the microbial composition of our gastrointestinal tract, leading to an environment conducive to the development of adenomas, and potentially colorectal cancer. Future studies are needed to assess the effects of diet on the metabolic environment of the gut and the microbial community. Genotoxic metabolites such as H2S and secondary bile acids should also be examined in relation to adenoma and carcinoma development. Identifying key interactions between diet, microbial community, and metabolites that catalyze the adenoma–carcinoma sequence will give us a basis for personalized therapeutics aimed at preventing colorectal cancer.

Disclosure of Potential Conflicts of Interest
D.J. Ahnen has received speakers bureau honoraria from Ambry Genetics and is a consultant/advisory board member for Cancer Prevention Pharmaceuticals. D.A. Ahlquist reports receiving a commercial research grant, has ownership interest (including patents), and is consultant/advisory board member for Exact Sciences. No potential conflicts of interest were disclosed by the other authors.

Authors’ Contributions
Conception and design: V.L. Hale, H. Nelson, N. Chia
Development of methodology: S.C. Harrington, N. Chia
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): S.C. Harrington, T.C. Smyrk, L.A. Boardman, B.R. Druliner, T.R. Levin, D.K. Rex, D.J. Ahnen, P. Lance
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): V.L. Hale, J. Chen, S. Johnson, T.C. Smyrk
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): V.L. Hale, S. Johnson, S.C. Harrington, T.C. Yah, H. Nelson
Study supervision: T.C. Yah, N. Chia

Acknowledgments
The authors thank all members of the Chia Laboratory for their input and efforts on this project. We acknowledge Kristin Harper for her thoughtful suggestions on this article. We also thank the reviewers for their comments and advice on this article.

Grant Support
This work was supported by the Mayo Clinic Center for Individualized Medicine (to V.L. Hale, J. Chen, and N. Chia), the Gerstner Family Career Development Award (to J. Chen), the Fred C. Andersen Foundation (to H. Nelson and N. Chia), the NIH under award number R01 CA 179243 (to V.L. Hale and N. Chia), the National Cancer Institute [Colorectal Cancer Screening, Fecal Blood vs. DNA; U01 CA 89389 (to D.A. Ahlquist), and Fecal Colonoctye Screening for Colorectal Neoplasia; R01 CA 71680 (to D.A. Ahlquist); and R01 CA 170357 (to L.A. Boardman)], Exact Sciences (to T.C. Yah and D.A. Ahlquist), and Mayo Clinic (to D.A. Ahlquist).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received April 27, 2016; revised September 2, 2016; accepted September 6, 2016; published OnlineFirst September 26, 2016.

Published OnlineFirst September 26, 2016; DOI: 10.1158/1055-9965.EPI-16-0337

Downloaded from cebp.aacrjournals.org on August 31, 2017. © 2017 American Association for Cancer Research.
References

**Cancer Epidemiology, Biomarkers & Prevention**

**Shifts in the Fecal Microbiota Associated with Adenomatous Polyps**


*Cancer Epidemiol Biomarkers Prev* 2017;26:85-94. Published OnlineFirst September 26, 2016.

<table>
<thead>
<tr>
<th>Updated version</th>
<th>Access the most recent version of this article at: doi:10.1158/1055-9965.EPI-16-0337</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supplementary Material</td>
<td>Access the most recent supplemental material at: <a href="http://cebp.aacrjournals.org/content/suppl/2016/09/24/1055-9965.EPI-16-0337.DC1">http://cebp.aacrjournals.org/content/suppl/2016/09/24/1055-9965.EPI-16-0337.DC1</a></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cited articles</th>
<th>This article cites 75 articles, 11 of which you can access for free at: <a href="http://cebp.aacrjournals.org/content/26/1/85.full#ref-list-1">http://cebp.aacrjournals.org/content/26/1/85.full#ref-list-1</a></th>
</tr>
</thead>
</table>

**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.